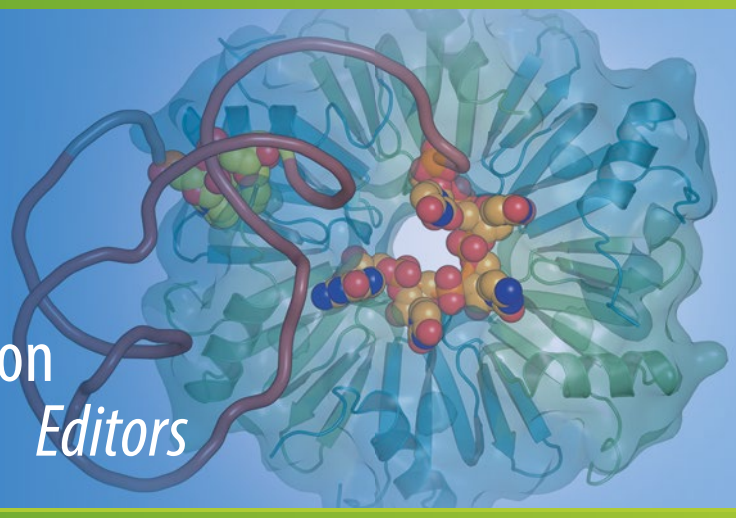
Methods in Molecular Biology 1737

Springer Protocols



Véronique Arluison
Claudio Valverde *Editors*

Bacterial Regulatory RNA

Methods and Protocols

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Bacterial Regulatory RNA

Methods and Protocols

Edited by

Véronique Arluison

Université Paris Diderot & LLB/CEA/CNRS, Gif-sur-Yvette, France

Claudio Valverde

Laboratorio de Bioquímica, Microbiología e Interacciones, Biológicas en el Suelo, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes-CONICET, Bernal, Argentina

Humana Press

Editors
Véronique Arluison
Université Paris Diderot & LLB/CEA/CNRS
Gif-sur-Yvette, France

Claudio Valverde
Laboratorio de Bioquímica
Microbiología e Interacciones
Biológicas en el Suelo
Departamento de Ciencia y Tecnología
Universidad Nacional de Quilmes-CONICET
Bernal, Argentina

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-7633-1 ISBN 978-1-4939-7634-8 (eBook)
https://doi.org/10.1007/978-1-4939-7634-8

Library of Congress Control Number: 2017964485

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media, LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

Regulatory small noncoding RNAs (sRNAs) are ubiquitous key regulators of gene expression in prokaryotes, operating at the post-transcriptional level to influence the fate of mRNA translation and/or stability, in most cases with the complicity of RNA binding proteins. Although significant progress has been made in the past 20 years to understand the function of individual sRNAs, high-throughput RNomics has recently revealed the enormous wealth of noncoding RNA species existing in a wide variety of prokaryotes, thus significantly expanding the implications of this class of regulatory molecules and boosting this new field of research of yet undimensioned relevance.

The understanding of many of the fundamental processes underlying the evolution, expression, structure, subcellular location, dynamics, and function of sRNA requires the availability of solid experimental approaches that may be applied either singly or in combinations to explore key aspects of sRNA biology. This volume collects many of the most important methods that have been recently set up for studying prokaryotic noncoding RNAs and their protein accomplices. The methods are presented in sections covering different aspects of the biology of that field: identification of ncRNAs, their differential expression, characterization of their structure and assembly, abundance, intracellular location and function, their interaction with RNA binding proteins, and plausible applications of ncRNA elements in the rapidly emerging field of synthetic biology. Each method includes a section with advice and tips from the authors. This volume aims to provide a guidebook to scientists that we hope will lead to new tools and procedures for further development in the field of sRNA biology.

Gif-sur-Yvette, France — *Véronique Arluison*
Bernal, Argentina — *Claudio Valverde*

Contents

Contributors

Frédéric H.-T. Allain • *Institute of Molecular Biology and Biophysics, ETH Zurich, Zurich, Switzerland*

Véronique Arluison • *Université Paris Diderot & LLB/CEA/CNRS, Gif-sur-Yvette, France*

Thierry Bizebard • *Expression Génétique Microbienne, UMR8261 CNRS/Université Paris 7, IBPC, Paris, France*

Thomas Bizien • *Synchrotron SOLEIL, L'Orme des Merisiers Saint Aubin, Gif-sur-Yvette, France*

Ulrich Bockelmann • *Nanobiophysics, ESPCI Paris, Paris, France*

Marc Boudvillain • *Centre de Biophysique Moléculaire (UPR 4301), CNRS, rue Charles Sadron, Orléans, France*

Magda Bujnowska • *Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, USA; The College of The University of Chicago, Chicago, IL, USA*

Florent Busi • *Université Paris Diderot-Paris 7, Sorbonne Paris Cité, Paris, France; Unité de Biologie Fonctionnelle et Adaptative, CNRS UMR 8251, Paris, France*

Marie-Claude Carrier • *Department of Biochemistry, RNA Group, Université de Sherbrooke, Sherbrooke, QC, Canada*

Juliet Cheng • *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA*

Kevin Chiem • *Center for Applied Biotechnology Studies, College of Natural Sciences and Mathematics, California State University Fullerton, Fullerton, CA, USA*

Lydia M. Contreras • *McKetta Department of Chemical Engineering, University of Texas at Austin, Austin, TX, USA*

Carol Davies-Sala • *Center for Applied Biotechnology Studies, College of Natural Sciences and Mathematics, California State University Fullerton, Fullerton, CA, USA; Fundación Instituto Leloir, IIBBA-CONICET, and FCEyN, University of Buenos Aires, Aires, Argentina*

Mildred Delaleau • *Centre de Biophysique Moléculaire (UPR 4301), CNRS, rue Charles Sadron, Orléans, France*

Louise Djapgne • *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA*

Olivier Duss • *Institute of Molecular Biology and Biophysics, ETH Zurich, Zurich, Switzerland; Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA; Department of Structural Biology, Stanford University, Stanford, CA, USA*

Carl T. Fakhry • *Department of Computer Science, University of Massachusetts Boston, Boston, MA, USA*

Pierre Fechter • *UMR7242 Biotechnologie et Signalisation Cellulaire, CNRS, Institut de Recherche de l'Ecole de Biotechnologie de Strasbourg, Université de Strasbourg, Illkirch-Graffenstaden, France*

Jingyi Fei • *Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, USA; Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, USA*

Joëlle V. Fritz • *Centre Hospitalier Luxembourg, Luxembourg, Luxembourg; Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg*

David Gally • *Division of Infection and Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, Scotland, UK*

Jonathan Gans • *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA*

Natalia I. García-Tomsig • *Grupo de Ecología Genética de la Rizosfera, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas (CSIC), Granada, Spain*

Jens Georg • *Genetics and Experimental Bioinformatics, Faculty of Biology, Institute of Biology III, University of Freiburg, Freiburg im Breisgau, Germany*

Anubrata Ghosal • *Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA*

Sander Granneman • *Institute of Structural and Molecular Biology, Centre for Synthetic and Systems Biology (SynthSys), University of Edinburgh, Edinbugh, Scotland, UK*

Janine Habier • *Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg*

Anna Heintz-Buschart • *German Centre for Integrative Biodiversity Research (iDiv) Leipzig-Halle-Jena, Leipzig, Germany; Department of Soil Ecology, Helmholtz-Centre for Environmental Research GmbH (UFZ), Halle (Saale), Germany; Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg*

Alexis Jackson • *Center for Applied Biotechnology Studies, College of Natural Sciences and Mathematics, California State University Fullerton, Fullerton, CA, USA; Fundación Instituto Leloir, IIBBA-CONICET, and FCEyN, University of Buenos Aires, Aires, Argentina*

Saumya Jani • *Center for Applied Biotechnology Studies, College of Natural Sciences and Mathematics, California State University Fullerton, Fullerton, CA, USA*

José I. Jiménez-Zurdo • *Grupo de Ecología Genética de la Rizosfera, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas (CSIC), Granada, Spain*

Birgitte Haahr Kallipolitis • *Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark*

Rahul V. Kulkarni • *Department of Physics, University of Massachusetts Boston, Boston, MA, USA*

Antonio Lagares Jr • *Laboratorio de Bioquímica, Microbiología e Interacciones Biológicas en el Suelo, Universidad Nacional de Quilmes—CONICET, Bernal, Argentina*

Ashwin Lahiry • *Department of Microbiology, The Ohio State University, Columbus, OH, USA*

Guillaume Laliberté • *Department of Biochemistry, RNA Group, Université de Sherbrooke, Sherbrooke, QC, Canada*

Richard A. Lease • *William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH, USA; Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA*

Abigail N. Leistra • *McKetta Department of Chemical Engineering, University of Texas at Austin, Austin, TX, USA*
Eva Maria Sternkopf Lillebæk • *Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark*
Antoine Malabirade • *Université Paris Diderot-Paris 7, Sorbonne Paris Cité, Paris, France*
Sergio Marco • *INSERM, U1196, Université Paris Sud, Université Paris-Saclay, Orsay, France; Institut Curie, PSL Research University, CNRS, UMR 91873348, Orsay, France*
Eric Massé • *Department of Biochemistry, RNA Group, Université de Sherbrooke, Sherbrooke, QC, Canada*
Ana M. Matia-González • *Faculty of Health and Medical Sciences, Department of Microbial and Cellular Sciences, School of Biosciences and Medicine, University of Surrey, Guildford, UK*
Patrick May • *Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg*
Eric L. McLean • *Department of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, IL, USA*
Erich Michel • *Institute of Molecular Biology and Biophysics, ETH Zurich, Zurich, Switzerland; Department of Biochemistry, University of Zurich, Zurich, Switzerland*
Mia K. Mihailovic • *McKetta Department of Chemical Engineering, University of Texas at Austin, Austin, TX, USA*
Valeria Militello • *Department of Physics and Chemistry, University of Palermo, Palermo, Italy*
Cameron Mura • *Department of Chemistry, University of Virginia, Charlottesville, VA, USA*
Cédric Nadiras • *Centre de Biophysique Moléculaire (UPR 4301), CNRS, rue Charles Sadron, Orléans, France; Ecole doctorale Santé, Sciences Biologiques et Chimie du Vivant (ED 549), Université d'Orléans, Orléans, France*
Esther N.M. Nolte-'t Hoen • *Faculty of Veterinary Medicine, Department of Biochemistry and Cell Biology, Utrecht University, Utrecht, The Netherlands*
Amanda G. Oglesby-Sherrouse • *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA*
Jonathan Osborne • *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA*
Seongjin Park • *Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, USA*
David Partouche • *Synchrotron SOLEIL, L'Orme des Merisiers Saint Aubin, Gif-sur-Yvette, France; Université Paris Diderot-Paris 7, Sorbonne Paris Cité, Paris, France*
Philippe Régnier • *Laboratoire Expression Génétique Microbienne UMR8261 CNRS/ Université Paris Diderot Institut de Biologie Physico-Chimique IBPC, Paris, France*
Marta Robledo • *Grupo de Ecología Genética de la Rizosfera, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas (CSIC), Granada, Spain*
Christophe Sandt • *Synchrotron SOLEIL, L'Orme des Merisiers Saint Aubin, Gif-sur-Yvette, France*
Annie Schwartz • *Centre de Biophysique Moléculaire (UPR 4301), CNRS, rue Charles Sadron, Orléans, France*

Alfonso Soler-Bistué • *Center for Applied Biotechnology Studies, College of Natural Sciences and Mathematics, California State University Fullerton, Fullerton, CA, USA; Fundación Instituto Leloir, IIBBA-CONICET, and FCEyN, University of Buenos Aires, Aires, Argentina*

Kimberly A. Stanek • *Department of Chemistry, University of Virginia, Charlottesville, VA, USA*

Samuel D. Stimple • *William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH, USA*

Brandon Sy • *School of Biotechnology and Biomolecular Sciences, University of New South Wales Sydney, Sydney, Australia*

Joseph E. Taris • *William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH, USA*

David Tollervey • *Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, Scotland, UK*

Marcelo E. Tolmasky • *Center for Applied Biotechnology Studies, College of Natural Sciences and Mathematics, California State University Fullerton, Fullerton, CA, USA*

Jai J. Tree • *School of Biotechnology and Biomolecular Sciences, University of New South Wales Sydney, Sydney, Australia*

Sylvain Trépout • *INSERM, U1196, Université Paris Sud, Université Paris-Saclay, Orsay, France; Institut Curie, PSL Research University, CNRS, UMR 91873348, Orsay, France*

Claudio Valverde • *Laboratorio de Bioquímica, Microbiología e Interacciones, Biológicas en el Suelo, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes-CONICET, Bernal, Argentina*

Marisela Velez • *Instituto de Catálisis y Petroleoquímica, CSIC, Madrid, Spain*

Jörg Vogel • *Institute of Molecular Infection Biology, University of Würzburg, Würzburg, Germany; Helmholtz Institute for RNA-Based Infection Research (HIRI), Würzburg, Germany*

Alexander J. Westermann • *Institute of Molecular Infection Biology, University of Würzburg, Würzburg, Germany*

Frank Wien • *Synchrotron SOLEIL, L'Orme des Merisiers Saint Aubin, Gif-sur-Yvette, France*

Anke K. Wienecke-Baldacchino • *Life Sciences Research Unit, University of Luxembourg, Belvaux, Luxembourg*

Paul Wilmes • *Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg*

Julia Wong • *School of Biotechnology and Biomolecular Sciences, University of New South Wales Sydney, Sydney, Australia*

David W. Wood • *William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH, USA; Department of Microbiology, The Ohio State University, Columbus, OH, USA*

Patrick R. Wright • *Bioinformatics Group, Department of Computer Science, University of Freiburg, Freiburg im Breisgau, Germany*

Kourosh Zarringhalam • *Department of Mathematics, University of Massachusetts Boston, Boston, MA, USA*

Angeles Zorreguieta • *Fundación Instituto Leloir, IIBBA-CONICET, and FCEyN, University of Buenos Aires, Aires, Argentina*

Part I

Identification of sRNA

Chapter 1

Workflow for a Computational Analysis of an sRNA Candidate in Bacteria

Patrick R. Wright and Jens Georg

Abstract

Computational methods can often facilitate the functional characterization of individual sRNAs and furthermore allow high-throughput analysis on large numbers of sRNA candidates. This chapter outlines a potential workflow for computational sRNA analyses and describes in detail methods for homolog detection, target prediction, and functional characterization based on enrichment analysis. The cyanobacterial sRNA IsaR1 is used as a specific example. All methods are available as webservers and easily accessible for nonexpert users.

Key words Computational methods, sRNA conservation, Target prediction, Functional characterization, Cyanobacteria, IsaR1

1 Introduction

Today, it is well established that small regulatory RNAs (sRNAs) are important players in bacterial gene regulation [1]. Some sRNAs from *Escherichia coli* and other model organisms have been studied for decades and their functions are now well known. However, nearly every new transcriptome study, especially in non-model organisms, reveals hundreds of previously unknown sRNA candidates. This poses the question how this abundance of data can be efficiently analyzed in order to identify good leads for follow-up projects. Of course, powerful wet-lab methods exist. They range from the classic genetic approaches like knockout and overexpression to the more advanced "omics" techniques [2]. While recent interactomics methods [3, 4] partly allow to simultaneously detect RNA–RNA interactions for multiple sRNAs, most of the methods require to concentrate on one single RNA. Computational sRNA analysis can aid in the selection of promising candidates for experimental testing without initially requiring a wet-lab overhead. Furthermore, aspects like the phylogenetic conservation or conservation of secondary structures are only approachable with

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_1,

biocomputational methods. Many other important questions can be at least partly addressed by bioinformatics.

1. *Is the sRNA conserved and which is its phylogenetic distribution?* If the sRNA is conserved, especially over a great evolutionary distance, it is more likely to have an important physiological function. Moreover, the existence of homologs enables comparative methods for target prediction, prediction of coding potential, and secondary structure elucidation.
2. *Does the detected transcript potentially code for a peptide?* If yes the transcript might not be a regulatory RNA, but rather code for a small-protein [5]. Alternatively, it could be a dual function RNA such as SgrS, RNAIII, or RyhB [6–8] coding for a peptide and an sRNA.
3. *Does the sRNA function by interaction with other RNAs or with proteins?* The majority of the sRNAs described so far function by base pair interactions to other RNAs; sometimes these interactions are facilitated by so-called RNA chaperones like Hfq [9] or ProQ [10, 11]. However, there are prominent examples such as the 6S RNA [12] or CsrB/CsrC [13] which work according to current literature only by interactions with proteins. The question if an sRNA works by RNA or protein interactions is not directly accessible by computational methods, but a meaningful target prediction speaks in favor of an RNA-dependent mechanism.
4. *If it is an RNA–RNA interacting sRNA, which are the targeted RNAs?*
5. *Which is the physiological function of the sRNA?* If there is a meaningful target prediction, it is likely that the sRNA acts by RNA–RNA interaction. Furthermore, a functional enrichment analysis of the predictions can hint at the physiological function, especially if combined with existing experimental data.

Clearly, these questions can only be conclusively answered after some experimental effort, but bioinformatics can help in this process. Figure 1 features a bioinformatic workflow and the respective tools (Table 1) to answer the above questions. The first step should always be a search for homologs. This can be done as explained below under application of the GLASSgo webserver [14]. If sRNA homologs exist, this opens up the road for comparative genomics algorithms. In particular, sRNA target prediction with CopraRNA [15–17] or the calculation of potentially conserved secondary structures with RNAz [18], LocARNA [19], or RNAalifold [20] is enabled. Furthermore, the prediction of small open reading frames of potential small proteins or dual function sRNAs with RNAcode [21] becomes possible. The latter two methods are not addressed in this chapter. Without homologs, the analysis is mostly restricted to single organism target prediction, e.g., by IntaRNA [16, 22, 23].

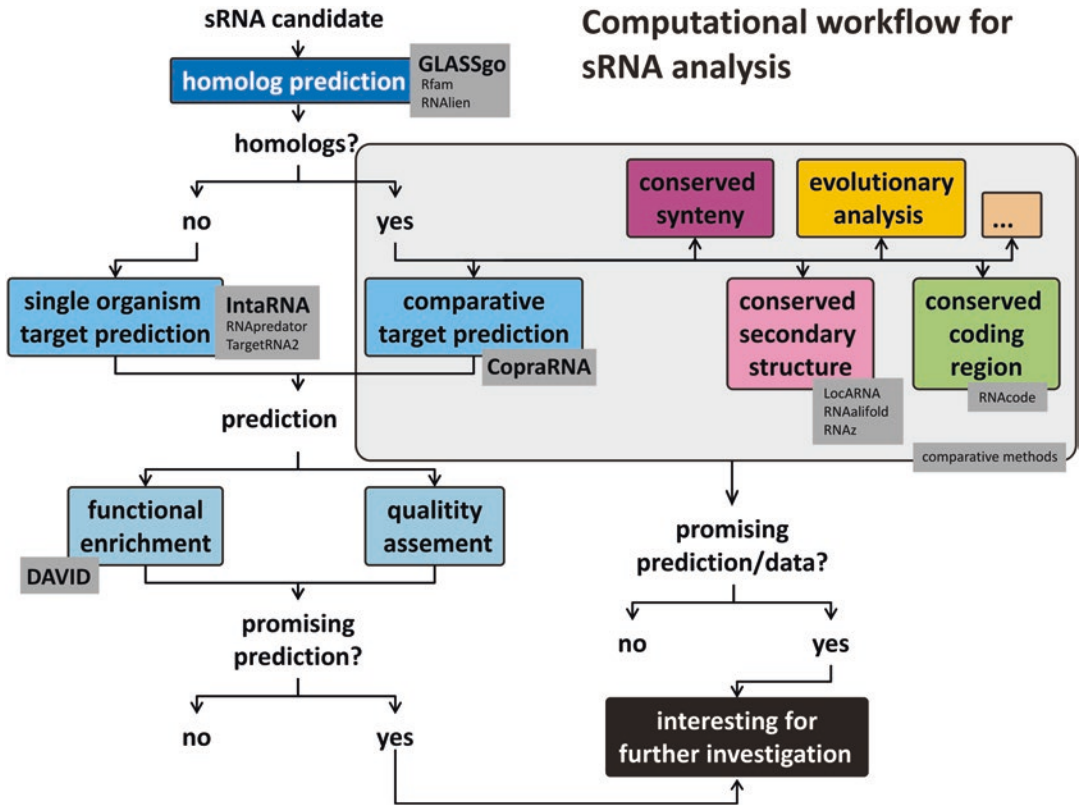


Fig. 1 Illustrative workflow for computational sRNA analysis. Names of bioinformatic tools are given in the gray boxes. Tools addressed in this chapter are in bold face and have increased font size. The resources for the tools are given in Table 1

2 Materials (Table 1)

Table 1
Computational tools and resources mentioned in this chapter

Name	Description	Resource
Homolog detection		
Rfam	Database of known sRNAs	http://rfam.xfam.org/
GLASSgo	Webserver for the de novo detection of sRNA homologs	http://rna.informatik.uni-freiburg.de/GLASSgo/Input.jsp
RNAlien	Webserver for unsupervised generation of covariance models	http://nibiru.tbi.univie.ac.at/rnalien
sRNA target prediction		
IntaRNA	Webserver for single organism sRNA target prediction	http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp

(continued)

Table 1
(continued)

Name	Description	Resource
CopraRNA	Webserver for comparative sRNA target prediction	http://rna.informatik.uni-freiburg.de/CopraRNA/Input.jsp
RNApredator	Webserver for single organism sRNA target prediction	http://rna.tbi.univie.ac.at/cgi-bin/RNApredator/target_search.cgi
TargetRNA2	Webserver for single organism sRNA target prediction	http://cs.wellesley.edu/~btjaden/TargetRNA2/
Prediction of conserved secondary structure		
RNAalifold	Prediction of consensus secondary structures	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAalifold.cgi
RNAz	Prediction of conserved secondary structures	http://rna.tbi.univie.ac.at/cgi-bin/RNAz/RNAz.cgi
Multiple sequence/structure alignment		
LocaRNA	Multiple sequence/structure alignment	http://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp
MAFFT	Multiple sequence or sequence/structure alignment tool	http://mafft.cbrc.jp/alignment/server/
Alignment visualization		
Jalview	Alignment editing and visualization	http://www.jalview.org/
Functional enrichment analysis		
DAVID-WS	Webserver for functional enrichment	https://david.ncifcrf.gov/

3 Methods

3.1 Analysis of sRNA Conservation

Homolog detection is often a cumbersome task, which requires a substantial amount of expert knowledge and manual curation [24]. The Rfam database [25] stores family models of diverse known noncoding RNA species and can be used to see if the input sRNA is already known. There are also tools for an automatic de novo search for conserved sRNAs. RNAlien [26] allows an unsupervised generation of covariance models from a single input sRNA, which can be used to scan databases for homologs. In contrast, GLASSgo [14] allows a complete automatic de novo identification of homologs for newly detected sRNAs starting from a single input query. Due to the good runtime, sensitivity, and specificity [14] only GLASSgo will be presented in detail.

3.2 GLASSgo

GLASSgo is part of the Freiburg RNA Tools and easily accessible as webserver (http://rna.informatik.uni-freiburg.de/GLASSgo/

Input.jsp). GLASSgo performs a low stringency iterative BLAST search against the NCBInt database. If the local hits are shorter than the query, the hits are extended based on the local pairwise BLAST alignments on both sides, to match the length of the query. The BLAST search is repeated with selected hits from different levels of sequence identity to the input sRNA in order to enhance sensitivity. Finally, all hits with a sequence identity above the threshold (default 52%) are returned. This procedure is followed by a tree-based auto-adaptive structural filter. In general, it is recommended to use the standard parameters on a specified taxonomic group for a first try. If the results are not satisfactory, the parameters can be adapted. The following steps describing the GLASSgo input and results page are marked by numbers in Fig. 2.

3.2.1 Input Page

1. Paste the sequence of your sRNA in FASTA format in the input window. A FASTA entry has a header row, which starts with a ">" sign and may contain a description of the sequence, followed by the sequence information in the consecutive rows. Further entries start with a header row again.
2. Optional: Restrict the BLAST search in the NCBInt database to a specific taxonomic group. It is possible to select one of the three taxonomic domains eukaryotes, archaea, or bacteria or one of the bacterial phyla. This selection can reduce false positives if an sRNA is only conserved in a specific taxonomic group. Note that the same alignment bit score (i.e., an alignment of equal quality) results in a different *E*-value in databases of different sizes. The *E*-value will be higher (worse) in bigger databases; an initial alignment that reaches the *E*-value threshold in a small database, e.g., a taxon-specific sub-database, might not be reported when searched against the full NCBI database at the same *E*-value cutoff.
3. In order to change further parameters, the "manually" button in the "Parameter setup" option has to be selected.
4. The maximum *E*-value defines how many sequences are considered for sequence extension and pairwise identity calculation. A high *E*-value potentially increases the sensitivity, but may drastically increase runtime and may reduce specificity.
5. The minimum allowed identity [%] (PI) defines the lowest possible global pairwise identity to the input sRNA. Reduction of the threshold may enhance sensitivity at potential cost of specificity.
6. The strength of the structure-based filter can be adapted. In automatic mode, the value is automatically adapted to the length of the input sRNA. The shorter the RNA, the less information is encoded in the primary sequence and the stricter the structural filter is. In manual mode, the structural filter can be

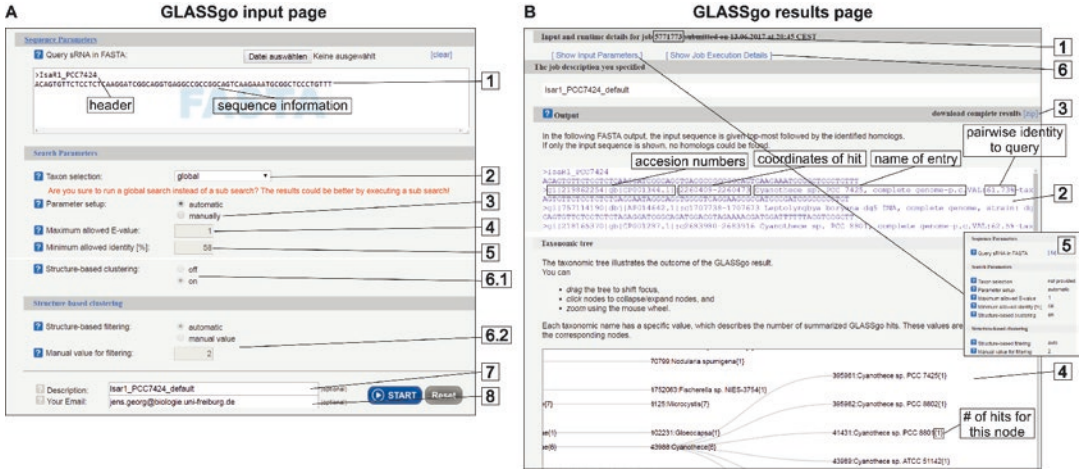


Fig. 2 (**a**) GLASSgo webserver input page and (**b**) results page. The numbers in the figure refer to the numbers in the GLASSgo section

switched off to increase sensitivity (Fig. 2a/6.1) or modulated by the user (Fig. 2a/6.2). A value of 0 allows no structural difference while increasing values result in a more relaxed filter. Meaningful values lie between 0.5 and 2.5.

7. Finally, it is possible to add a description to your job which is displayed at the top of the result page and in an optional E-Mail notification. This can be useful to distinguish runs with different parameter settings.
8. Here it is possible to leave an E-Mail address to receive a notification when and if the job has finished. Your run will be assigned to a job ID. It is possible to retrieve the results with this ID via the "results" button within a storage period of 1 month.

3.2.2 Results Page

The main outputs of GLASSgo are a FASTA file with the predicted sRNAs and an interactive taxonomic tree of the respective organisms.

1. Job ID for result retrieval.
2. FASTA sequences. The first entry of the FASTA file is the input sRNA with the user-specified FASTA header. The headers of the predictions contain information about the source NCBI entry, i.e., the accession number(s) of the respective entry, the coordinates of the prediction, the name of the entry, the pairwise identity to the query, and the NCBI TaxID of the respective organism. The TaxID is a unique identifier for each taxonomic group in the NCBI Taxonomy Browser.
3. The download option allows retrieving a zip folder which is named by the job ID. It contains the input parameters for the webserver call (xxxx.input), the results FASTA file (xxxx.

result), the json file which is used to draw the taxonomic tree (xxxx.result.json), information about the GLASSgo version (xxxx.version), and the input sRNA (input_query.fa).

4. The taxonomic tree allows a fast survey of the phylogenetic distribution of the organisms with a predicted homolog.
5. Furthermore, information about the input parameters is available by clicking "Show Input Parameters."
6. Runtime information can be retrieved with "Show Job Execution Details."
7. The "Restart" button allows to directly rerun the prediction with a changed parameter setting, or the same parameters and another input sequence.

3.2.3 Troubleshooting

Problem	Reason	Solution
No or only few homologs with very high sequence identity have been detected	The initial BLAST hits of true homologs have an *E*-value above the threshold	Rerun GLASSgo with relaxed parameters (*E*-value ≥10, structural filter off, PI threshold = 55%, select phylum of input sRNA organism) to scan if there might be any potential further homologs. Relaxed parameters may result in many false positives
	Structural filter is too strict	
	The homologs have a PI below the threshold	
	There are no further homologs	–
Only few homologs are detected. Some have a relatively low PI	Other homologs are rather distinct from the query and cannot be detected by default or even relaxed parameters	Rerun GLASSgo with a detected homolog that has a low PI
	There are no further homologs	–
The result contains many false positives	The parameters are too loose for the particular sRNA	You can restrict search to the phylum of the input sRNA, increase the PI-threshold, or tighten the structural filter (lower the value)
An increase in the *E*-value results in less predictions	Due to the auto-adaptive nature of the structural filter, more "pre-filter" sequences sometimes result in a stricter filtering and less output sequences	Switch off or relax the structural filter

3.2.4 Post-processing

GLASSgo has a high specificity, but the results may nevertheless contain false positive predictions. For most downstream applications and evolutionary analyses it may be important to use only true homologs as input. A definitive classification of a potential homolog as a biologically true homolog is a challenging task.

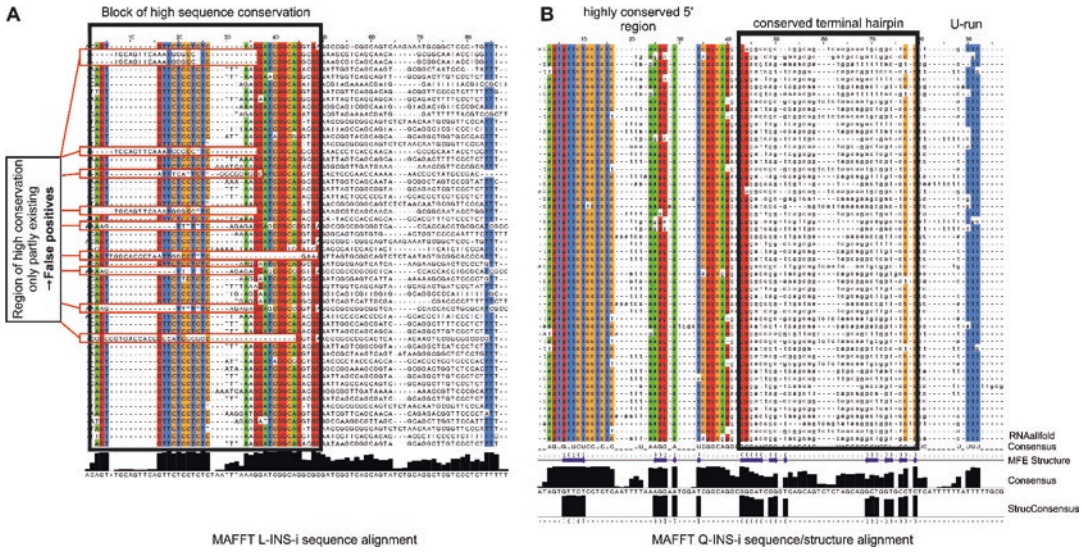


Fig. 3 (**a**) Example for the post-processing of a GLASSgo prediction for IsaR1 with strongly relaxed parameters based on a multiple sequence alignment. The MAFFT L-INS-i sequence alignment is visualized in Jalview and all alignment positions with a sequence identity ≥80% are colored. There is a highly conserved region at the 5′ end of the alignments, which likely signifies a region of functional importance. Ten sequences do not share the conservation in this region and are easily detectable by eye because they are not colored. These sequences are false positives and can be deleted from the alignment. (**b**) MAFFT Q-INS-i sequence/structure alignment of IsaR1 homologs after the deletion of false positives. RNAalifold was performed on the alignment directly from Jalview. The dot bracket representation of the MFE consensus structure shows a conserved hairpin at the 3′ end, which is followed by a U-run. This indicates a Rho-independent transcription terminator and supports that the given sequences are transcribed

Figure 3 features an example for a fast false positive classification based on a multiple sequence alignment.

1. A good indicator is high sequence conservation. As a conservative rule of thumb, a global sequence conservation ≥80% is a strong indicator of a true positive homolog. The sequence conservation with respect to the input sRNA can be easily accessed from the header in the GLASSgo output FASTA file.

   However, according to current literature most sRNAs do not code for proteins and there is often a much higher sequence variability compared to homologous protein-coding mRNAs. Nevertheless, the function of most *trans*-acting sRNAs that function via RNA–RNA interactions is also encoded in their primary sequence. Most known sRNAs interact via one or more short sequence patches with multiple RNA targets. These multiple interactions put a constraint on the evolution of the interacting sequences in the sRNA, which often results in local regions with high sequence conservation. These regions of high sequence conservation can be used for true positive classification.

2. The first step for a further analysis should be a multiple sequence alignment. False positive sequences might prevent a meaningful alignment including structural information and a sequence-only alignment is better for a first overview. A "sequence-only" alignment can be conducted, e.g., on the MAFFT webserver [27]. The tool allows showing the results in FASTA format. The plain text from the FASTA file can be copied to the clipboard to be used for the next step and/or pasted into a text file to be saved for archiving.
3. Next, the alignments need to be properly visualized in order to draw swift conclusions "by eye." An easy-to-use alignment editor is Jalview [28]. The alignment can be opened from a file (File → Input Alignment → From File) or directly from the clipboard (File → Input Alignment → from Textbox → "paste your alignment in FASTA format in the Textbox" → New Window). This alignment should be colored using the "Colour" pull down menu, e.g., by "Nucleotide." Then, only alignment positions that are conserved in most of the sequences (e.g., 80%) should be colored. For this, select "Above Identity Threshold" in the "Colour" menu. A slider to select the identity threshold appears and only respective positions are colored. After this step, blocks with high sequence conservation should appear, assuming that the majority of the sequences in the alignment are true positives and that the local sequence conservation of some regions is higher than the global conservation. Predictions that do not share conservation within these blocks of high sequence conservation are likely to be false positives and can be easily identified by eye. False positives can be selected (left mouse click on the sequence name) and deleted from the alignment (Edit → Delete). The resulting file can be saved in FASTA format (File → Save as). Jalview allows direct usage of many sequence-based alignment tools if un-aligned sequences are loaded (Web Service → Alignment → "select method of choice").
4. The refined set of sequences without false positives can be used for a sequence/structure alignment followed by a prediction of a conserved secondary structure with RNAalifold [20]. If there are less than 30 predicted homologs, this can be done with the LocARNA webserver (http://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp [19]). The LocARNA webserver additionally does a comparative secondary structure prediction based on the alignment using RNAalifold [20]. A simultaneous sequence/structure alignment with up to 200 sequences can be done with MAFFT using the Q-INS-i version (http://mafft.cbrc.jp/alignment/server/ [27]). The local versions of LocARNA or MAFFT do not have input restrictions and can be easily installed on 64-bit Linux and Mac OSX systems using

the bioconda package manager (https://bioconda.github.io/index.html). All tools allow showing the results in FASTA format. The plain text from the FASTA file can be copied to the clipboard to be used for the next step and/or pasted into a text file to be saved for archiving. Jalview allows direct usage of RNAalifold (Web Service → Secondary Structure Prediction → RNAalifold Prediction). After refreshing the alignment view (View → New View), the RNAalifold consensus sequence, the dot bracket representation of the consensus MFE-structure, and structural consensus are displayed below. In the IsaR1 example there is a conserved terminal hairpin followed by several uridine nucleotides which is an indicator of a Rho-independent terminator [29]. A conserved Rho-independent terminator is a signal that the respective sequences are indeed independent transcripts.

5. There are other more sophisticated [30], but also more labor-intensive methods to identify true homologs, such as a conserved secondary structure analysis or the search for a conserved promoter. Also, a similar synteny is a strong indicator that the sequences are indeed related and originate from a common ancestor. These points have been addressed in detail in Chapter 10.

Some predictions can result in a high number of true positives. This is often the case for phyla with many sequenced genomes, such as Proteobacteria. While each positive hit can be of interest in terms of evolutionary analyses, other downstream applications might benefit from a meaningful selection of homologs. A good strategy is to select candidates based on taxonomy, e.g., one homolog from each species or genus. An automatic homolog selection for downstream applications is planned.

3.2.5 *Use Case Example*

An example for a complicated candidate is the cyanobacterial sRNA IsaR1 [31]. IsaR1 was first detected in *Synechocystis* PCC803. Using the PCC6803 IsaR1 sequence and default parameters, GLASSgo revealed only 100% identical homologs from *Synechocystis* PCC6803 sub-strains and a very close homolog from *Synechocystis* PCC6714 with a PI >90%. Another run with relaxed parameters (*E*-value = 20, BLAST restricted to cyanobacteria) revealed one further hit from *Cyanothece* sp. PCC 7424 and a PI of 62.7%. Using this more remote homolog as a new query with default parameters, GLASSgo detected 47 true positive IsaR1 candidates excluding the PCC6803 and PCC6714 homologs. If both lists are combined there are 49 predictions, which are sufficient for most approaches. However, they still do not fully describe all IsaR1 homologs that are present in the NCBI database. A more exhaustive list of homologs can be retrieved by rerunning GLASSgo with another input and/or relaxed parameters without too much effort,

as each run takes only a few minutes. Pre-computed results of the three use case GLASSgo runs are accessible at the GLASSgo webserver.

3.3 Target Prediction

Computational target prediction can provide a highly valuable contribution to the functional characterization of an sRNA. In optimal cases it can even singlehandedly uncover the approximate function of the investigated RNA within some hours and without any experimental efforts or costs. On the downside, target prediction remains a challenging task. RNA–RNA interactions are often short and may contain gaps and bulges. This makes it hard to distinguish them from random complementarities in a genome-wide analysis. Thus, even the best methods produce a non-negligible amount of false positives. Furthermore, most given RNA sequences will produce some kind of prediction on the genomic scale, even if they do not function by RNA–RNA interaction at all. In the following we will provide guidelines on how to interpret target prediction results and when it is better not to rely on them. There is a range of easy-to-use webservers for whole genome sRNA target predictions (e.g., IntaRNA [16, 22, 23], CopraRNA [15], TargetRNA2 [32], and RNApredator [33]). This protocol will focus on the currently best performing tools IntaRNA and CopraRNA [34, 35]. CopraRNA is a comparative algorithm that employs IntaRNA. The webserver interfaces for both tools are highly similar. If an sRNA has three or more homologs, preferably from different species, it is recommended to use CopraRNA which has a superior specificity and sensitivity [15, 34]. In general, and if possible, we suggest using more homologs that are evenly distributed in the phylogenetic tree (e.g. based on16S rDNA). If no homologs are available, then IntaRNA can be used for single whole genome target predictions. The following steps describing the IntaRNA and CopraRNA input and results page are marked by numbers in Fig. 4.

3.3.1 IntaRNA (Whole Genome Target Prediction)

IntaRNA has a pairwise analysis mode (default) and a whole genome mode. In the pairwise mode (explained further below) the sRNA and the putative target sequence(s) are supplied in FASTA format by the user. In the whole genome mode the user specifies a genome and the putative target sequences, i.e., the 5′ or 3′ UTRs of annotated genes are extracted automatically.

Input Page

1. Paste your sRNA sequence in FASTA format into the input window.
2. Click on the “Get target RNA sequences from NCBI Genome” link to enable whole genome target prediction.
3. Type the RefSeq ID of the genome of interest in the “Target NCBI RefSeq ID” window. You can find the ID of your

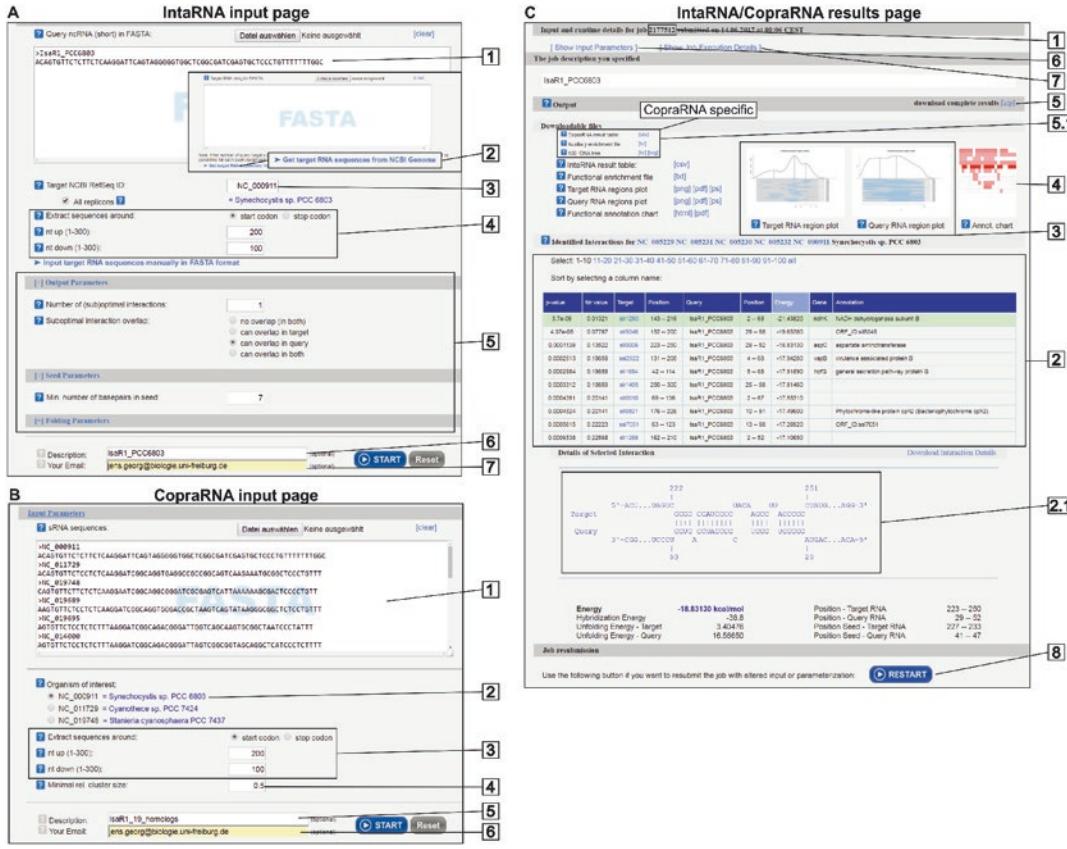


Fig. 4 (**a**) IntaRNA and (**b**) CopraRNA webserver input pages and (**c**) the IntaRNA/CopraRNA results page. The numbers in the figure refer to the numbers in the IntaRNA and CopraRNA sections

organism in this list http://rna.informatik.uni-freiburg.de/CopraRNA/CopraRNA_available_organisms.txt." The RefSeq ID should start with NC_ or NZ_. If you check the box "All replicons," all replicons belonging to this organism, e.g., additional chromosomes or plasmids, are included in the analysis.

4. Next, it is possible to specify the regions from the genome which are extracted for the prediction. Known bacterial sRNAs mostly target the 5′untranslated region (UTR) or the first nucleotides of the coding sequence. Thus, it is recommended to use the region around the "start codon." It is possible to specify the length of the UTR and the extracted coding region from 1 to 300 nucleotides. For this automatic extraction it is not possible to use potentially available sequencing data to specify the actual UTR lengths. Sequences around the stop codon can also be retrieved if interactions are expected in the 3′UTR [36] by selecting "stop codon."
5. The following sections "Output Parameters," "Seed Parameters," and "Folding Parameters" should only be

changed by expert users who are familiar with the influence of the respective parameters.

6. It is possible to add a description to your job which is displayed at the top of the result page and in the optional E-Mail notification. This can be useful to distinguish runs with different parameter settings.
7. Finally, it is possible and suggested to leave an E-Mail address to receive a notification when the job is finished.

Results Page

1. Every job has its own unique ID with which results can be retrieved at a later time within the 30-day storage period. This ID is also important when asking the webserver support questions about specific runs.
2. The central result of an IntaRNA whole genome target prediction is an energy score sorted list of duplex predictions between the input sRNA and all putative targets. The top 100 predictions are displayed in the browser. The full prediction table in *.csv format can be downloaded as part of the complete result *.zip archive (5). The respective file is called "intarna_websrv_table.csv." This file may be useful for users who want to perform more complex downstream analyses or who are interested in predictions beyond the top 100. The result table contains information on the IDs (locus tag) and annotations of putative targets. Furthermore, central aspects of the predicted duplex are supplied. By selecting an interaction of interest from the webserver list, the user can have a closer look at the predicted duplex and the detailed properties of the predicted interaction, which is displayed below the list (2.1).
3. Secondary results of the prediction are the regions plots for both the sRNA and the putative targets. The bottom part of the plots shows the interacting region for the top 25 predicted targets. For the targets, the relative position with respect to the start or stop codon is shown. On the sRNA plot, the position is shown with respect to the full length sRNA. The density plots on the top are based on all predictions with a p-value ≤ 0.01. The plots can be downloaded as *.png, *.pdf, or *.ps files.
4. The second downstream result is the functional enrichment, which is based on the DAVID-WS [37]. It is calculated based on the top 50 predictions and gives information on functional patterns within the top predictions. This result can be investigated by selecting the "Annot. chart" on the top right of the output page or by looking at the raw output, which is available as *.txt file next to "Functional enrichment file." The functional annotation chart can be downloaded as *.pdf or *.html.
5. The full back-end result directory is available as *.zip archive next to "download complete results."

6. Furthermore, information about the input parameters and files are available by clicking "Show Input Parameters."
7. Runtime information can be retrieved with "Show Job Execution Details."
8. The bottom of the result page features the "Restart" button, which enables to directly rerun the prediction with a changed parameter setting or RNA input.

3.3.2 CopraRNA

The selection of input organisms is a critical step for the CopraRNA prediction and different sets of organisms can yield more or less different results. If the organisms are too close, e.g., if the sRNA is only conserved in sub-strains of the same organism, CopraRNA will not have a benefit over a single organism prediction performed with IntaRNA. In general, some targets might only be conserved in a subset of organisms, while other mRNAs are targeted by all homologs of an sRNA (core targets). Thus, it is hard to give a definitive rule for the organism selection. However, some rules of thumb which have been helpful to us are the following:

- A complete RefSeq genome needs to be available. Compatible organisms are present in this list "http://rna.informatik.uni-freiburg.de/CopraRNA/CopraRNA_available_organisms.txt"
- The more organisms, the better (up to 20 are supported by the webserver).
- Use several organisms which are phylogenetically close to your species of interest, to be able to pick up targets that are potentially only conserved in these organisms.
- The remaining organisms should have an equal phylogenetic spread and contain also distantly related homologs to allow for a robust prediction of core targets.
- Running predictions with subsets of the original input set can be helpful [38].

Input Page

1. Paste your homologous sRNA sequences in FASTA format into the input window. The FASTA headers need to be formatted to specifically point at the RefSeq ID of the affiliated organisms. You can find the ID of your organism in this list "http://rna.informatik.uni-freiburg.de/CopraRNA/CopraRNA_available_organisms.txt." CopraRNA-compatible RefSeq IDs start with NC_ or NZ_. CopraRNA automatically includes all additional chromosomes and plasmids for each organism even though only one RefSeq ID pointing at the organism is supplied.
2. Select your organism of interest. This is important for the webserver browser display and the post-processing steps, especially for the functional enrichment and annotation, which is only performed for the organism of interest. Thus, it should be

either the best annotated organism to improve the functional enrichment or the organism you are working on.

3. Next, just like for the IntaRNA interface, it is possible to specify the regions from the genome which are extracted for the prediction. Thus, the same guidelines apply (*see* IntaRNA Input page description point 4).
4. The minimal relative cluster size parameter adjusts how many genes must at least be present in every putative target cluster with respect to the total number of participating organisms. For a cutoff of 0.5, this means that at least half of the organisms in a single prediction set need to have a homolog belonging to a specific gene cluster. If this is not the case, the cluster is not considered in the prediction. Setting higher stringency (i.e., bigger values) may reduce noise, but can also cause loss of real targets. This is currently an experimental parameter and should not be changed from 0.5 unless for specific reasons.
5. It is possible to add a description to your job which is displayed at the top of the result page and in the optional E-Mail notification. This can be useful to distinguish runs with different parameter settings.
6. It is possible and suggested to leave an E-Mail address to receive a notification when the job is finished.

Results Page

1. Every job has its own unique ID with which results can be retrieved at a later time during the 30-day storage period. This ID is also important when asking the webserver support questions about specific runs.
2. The central result of a CopraRNA whole genome target prediction is a CopraRNA *p*-value sorted list of target predictions. The top 100 predictions are displayed in the browser. The annotation and specific prediction details are shown for the organism of interest. The prediction table in *.csv format can be downloaded by clicking the link next to "CopraRNA result table." This file shows more details for the other organisms participating in the comparative prediction and thus enables more detailed downstream analyses. By selecting an interaction of interest from the webserver list, the user can have a closer look at the predicted duplex and the detailed properties of the predicted interaction (2.1).
3. Secondary results of the prediction are the regions plots for both the sRNA and the putative targets. The bottom part of the plots shows the interacting region for the top 20 predicted targets. For the targets, the relative position with respect to the start or stop codon is shown. On the sRNA plot, the position is shown with respect to the full length sRNA. The regions are depicted for each homologous sequence, which can give an

indication of how conserved the position of a predicted interaction is. The different colors are purely for resolution. The density plots on the top are based on all predictions with a CopraRNA *p*-value ≤0.01. The plots can be downloaded as *.png, *.pdf, or *.ps files.

4. The second downstream result is the functional enrichment, which is based on the DAVID-WS [37]. It is calculated based on the top 100 predictions that have homologs in the organism of interest and gives information on functional patterns within the top predictions. This result can be investigated by selecting the "Annot. chart" on the top right of the output page or by looking at the raw output, which is available as *.txt file next to "Functional enrichment file." The functional annotation chart can be furthermore downloaded as *.pdf or *.html.
5. The full back-end result directory is available as *.zip archive next to "download complete results." The full prediction result is called "CopraRNA_result_all.csv" and can be imported in spreadsheet software. Among other outputs there are also the individual IntaRNA predictions for each genome, which are named by the convention "*RefseqID* _upfromstartpos_*xxx*_down_*xxx*.final.csv" (Fig. 4c/5.1). Information on the 16S rDNA phylogeny of the organisms participating in the comparative prediction is available in *.txt or *.svg format next to "16S rDNA tree." CopraRNA features also an "auxiliary enrichment file" that contains targets from the single organism IntaRNA prediction for the organism of interest that are not in the CopraRNA top list, but fit to the enriched terms.
6. Furthermore, information about the input parameters and files are available by clicking "Show Input Parameters."
7. Runtime information can be retrieved with "Show Job Execution Details."
8. The bottom of the result page features the "Restart" button, which enables to directly rerun the prediction with changed parameters and input files.

Frequently Asked Questions and Troubleshooting

Question	Answer
Why are only organisms supported that are part of the RefSeq database?	In order to guarantee easy usability, CopraRNA requires a certain degree of consistency within the files that it accesses. RefSeq is—in most cases—a very reliable database that meets reasonable consistency terms

Question	Answer
What are additional homologs?	Sometimes, the clustering of homologous genes assigns several genes from one organism to the same cluster. In this case, the analysis is only executed on the candidate with the best IntaRNA energy score. In order to prevent losing the other putative targets, they are added at the end as additional homologs. If a known target clusters with additional homologs, they are good candidates for further investigation
My prediction list contains the same putative target gene more than once. Why?	In the process of clustering putative target genes, it sometimes happens that extremely similar clusters are generated which may only differ in one gene which is not part of your organism of interest. This may appear to be duplication when only considering the organism of interest, but it is not a real duplication. A closer look at the "cluster.tab" file from the results archive can clarify when and how this happens
Is CopraRNA deterministic? Previous results are not exactly identical to my results when I rerun the tool with the same homologs and organisms. Why?	Due to the *p*-value sampling for clusters that do not contain genes from each participating organism, CopraRNA is not a deterministic algorithm. However, usually only slight differences between distinct analyses are to be expected
The rank 1 prediction has a *p*-value of 0	Sometimes sRNAs have a conserved synteny and are encoded opposite to the 5′ UTR of a specific gene in many of the investigated organisms. This can lead to an arbitrarily good prediction and *p*-value. The corresponding interaction between sRNA and target UTR displayed by the webserver will show a long perfect complementary in these cases. When in doubt, a closer look into the sRNA synteny helps to detect these cases

Problem	Reason	Solution
The prediction failed to produce a result	Something in the back-end has not worked as expected	Read the error message on the result page and see if you can change the input accordingly. If this is not possible, get in touch with the developers and they will have a closer look
There is no functional enrichment	There are no enriched terms in your prediction or the annotation of the organism of interest might be poor	Retry enriching manually with another set of genes (e.g., top 25, 50, 150) or for another organism. Try to change DAVID parameters
	Your organism is not in the DAVID database	Try using another organism of interest
	The DAVID-WS was temporarily not accessible	Rerun the target prediction or perform enrichment analysis manually

3.3.3 IntaRNA (Pairwise Analysis)

After the careful investigation of whole genome target predictions certain RNA–RNA interaction pairs will stand out and call for follow-up verification (including mutational analyses) in the wet-lab. However, before starting these experiments, a closer look at the selected RNA–RNA pairs with the pairwise version of IntaRNA is necessary.

Input Page

1. "Input target RNA sequences manually in FASTA format" is set as standard for the input.
2. Paste your sRNA sequence in FASTA format into the "Query ncRNA" window.
3. Paste your target RNA sequence in FASTA format into the "Target RNA" window.
4. Set the "Number of (sub)optimal interactions" to at least 10 and "Suboptimal interaction overlap" to "can overlap in both."
5. Start the prediction.

Result Page

1. The output page displays a list of putative hybrids for your RNA–RNA pair of interest.
2. The different hybrids can be visually inspected by clicking the entry in the list.

Because the interaction predicted by whole genome target predictions is not always the hybrid that is important for the interaction in vivo, the list of potential hybrids can serve as a template of examples to pick from when planning experiments. Also, in some cases sRNAs may employ redundant regions to interact with their targets [39] and this will only become evident when using pairwise IntaRNA. Furthermore, mutational experiments can be designed with pairwise IntaRNA by performing the experiments in silico before going to the lab. An inserted mutation should show a pronounced difference in the IntaRNA energy score when compared to the wild-type interaction. Conversely, a compensatory mutation should compensate for this difference in energy and restore an energy score close to or lower than the original value.

3.4 Functional Enrichment DAVID

The results of predictive algorithms can be puzzling and hard to evaluate, especially if the user's background knowledge is limited, the result lists are lengthy, or true positive predictions are intermingled with false positives. However, if the prediction is generally sound, underlying physiological patterns can be extracted under application of functional enrichment methods. Here, it is of use that *trans*-acting sRNAs often have multiple targets that are involved in functionally related processes [15]. These processes can be uncovered by a functional enrichment analysis. Currently, our method of choice is DAVID [37]. DAVID is built around a highly

comprehensive knowledge base and thus interfaces many different functional classification sources including the Gene Ontology (GO) [40] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [41] databases. Related functional terms are combined to cluster in order to simplify the result and account for functional overlaps. A term is enriched if it is statistically overrepresented in a list of interest (for example, the top reported candidates of a predictive algorithm) compared to the frequency of the term in a background list (usually the whole genome). The assessment of the enrichment is centrally based on a statistical test referred to as Fisher's exact test. This test calculates *p*-values for different functional terms and the best scoring functional groups are reported back. The *p*-value reflects how likely it is that an equally strong or stronger overrepresentation of a term appears by chance. Thus, lower *p*-values indicate more statistically significant results. Importantly, the main quality measure is not directly a *p*-value but rather the enrichment score, which is calculated by computing the geometric mean of *p*-values for all functional terms within an annotation cluster and applying the negative $\log_{10}()$ function on that mean. Given this and a *p*-value significance threshold of $p \leq 0.05$, an enrichment score ≥ 1.3 is considered statistically significant because $(-1) \times \log_{10}(0.05) = 1.3$.

The CopraRNA and IntaRNA webserver versions perform a functional enrichment analysis for the top 100 and 50 predictions, respectively. In CopraRNA this is only performed for the organism of interest. However, it might be useful to investigate a higher or lower number of predictions to achieve a meaningful functional enrichment. In these cases the functional enrichment analysis can also be performed manually at the DAVID website (https://david.ncifcrf.gov/). This allows individual handling of input lists and software parameters. Currently, DAVID is running version 6.8 but we are still using version 6.7 (https://david-d.ncifcrf.gov/) alongside because it has been commonly returning better results in our use cases. Both CopraRNA and IntaRNA are still only interfacing version 6.7 in their automatic functional enrichments. When using CopraRNA, this way the functional enrichment can be also done for "non-organism of interest" predictions.

3.4.1 Input Page

1. To manually perform the functional enrichment analysis with the DAVID browser interface, the prediction result table of interest from the archive *.zip file ("CopraRNA_result_all.csv" for CopraRNA and "intarna_websrv_table.csv" for IntaRNA) should be imported into spreadsheet software. If using Microsoft Excel, this is done by DATA → From text → "select *.csv file" → "check "Delimited" in the pop up window" → "press next" → "check "Semicolon" for IntaRNA or "comma" for CopraRNA as delimiter" → Finish. In the IntaRNA file the Entrez GeneIDs are given in column 22. In the CopraRNA file

the Entrez Gene IDs (DAVID v6.7/v6.8) or locus tags (only DAVID v6.8) need to be extracted ("select the column of your organism of choice" → Home → Find&Select → Replace → Find what. = "*:", Replace with = "" → Replace all).

2. Within the DAVID site "Functional Annotation" on the top left needs to be selected. Then the gene list of interest can be pasted into the window on the left and the correct identifier (ENTREZ_GENE_ID or LOCUS_TAG depending on what you are using) needs to be selected. In the third step, the list type can be specified. If it is your gene list of interest then "Gene List" needs to be selected. In certain cases it is also advisable to submit a custom background (the standard is the whole genome). For CopraRNA, for instance, the background should be submitted as all candidates that are present in the entire result list. The according *.csv file is available from the "complete results" *.zip archive. The background can be submitted in the same way as the gene list. In the fourth step, the list can be submitted with "Submit List." Selecting specific lists and backgrounds can be performed by using the "List" and "Background" panels on the top left after lists have been submitted.

3.4.2 Results Page

1. When the correct list and background have been submitted, the "Functional Annotation Clustering" button can be clicked. A new window pops up showing clusters of functional terms and their enrichment scores. The options of the functional enrichment analysis can be adjusted by opening the "Options" section and subsequently clicking "Rerun using options." The CopraRNA webserver uses the options: Classification stringency: high; Initial Group Membership: 2; Final Group Membership: 2.

A comprehensive list also containing alternatives to DAVID is presented here [42].

3.5 Use Case Example

The IntaRNA prediction was performed with the *Synechocystis* PCC6803 IsaR1 homolog against the 5′ UTRs (200 nt upstream and 100 nt downstream of the annotated start codons) of the whole *Synechocystis* genome (RefSeq ID: NC_000911). Otherwise, no changes to the default parameters were made. The fastest way to judge the prediction is to look into the visualization of the functional enrichment. One cluster with an enrichment score of 0.95 is displayed. This is below the threshold of 1.3 and means that the statistical significance of the enrichment of these terms is relatively weak in comparison to the expected appearance of the terms based on the background. The cluster consists of 21 terms including "metalloprotein," "GO:0005506~ironionbinding," "iron," "GO:0022900~electrontransportchain," and "GO:0015979~

photosynthesis" (Fig. 5). Despite the low statistical significance, the enriched terms are highly promising because it is known that IsaR1 is induced upon iron depletion [43]. In this example, the enrichment score did not improve when the enrichment was manually repeated with the top 25, 100, or 150 hits instead of the top 50 predictions. The region plots are generally less informative in the single organism IntaRNA analysis. In the case of IsaR1 there is no clearly detectable interaction region in the sRNA (Fig. 6). The prediction contains several verified targets in the top 100 list; e.g., *petF* (#14), *petB* (#18), and *hemA* (#20). From an *a priori* perspective, only the combination of the known expression profile of IsaR1 and the functional enrichment would make this prediction interesting for further investigation.

The CopraRNA prediction was performed on 19 IsaR1 homologs retrieved from the GLASSgo webserver using default CopraRNA parameters. The prediction differs in terms of the functional enrichment and the exact result list from the CopraRNA prediction in the original publication [31]. This is due to the usage of different IsaR1 homologs. However, the overall conclusions remain fundamentally unchanged. The functional enrichment is clearly more distinctive for the comparative prediction. There are three clusters and two of them have an enrichment score above 1.3. The genes within these clusters partially overlap, which means that there are, judged by this analysis, no independent physiological functions predicted to be regulated. The enriched terms are again, e.g., "GO:0005506~ironionbinding," "GO:0009055~elect roncarrieractivity," "iron," or "iron-sulfur," and they all fit to the expression profile of IsaR1 (Fig. 5). The number in front of each term indicates the fold enrichment of the term in the predictions (i.e., to what extent they are more frequent in the predictions than expected from the frequency in the background). In the case of the CopraRNA prediction, the functional enrichment further improves when performed on the top 150 hits, indicating that there might be additional targets beyond the top 100 list. Interestingly, the rank 1 target "ycf24" is not represented in the functional enrichment. However, we now know that it is one of the most important IsaR1 targets. Thus, for a closer inspection and experimental verification, it is recommended to focus on the top 20 list plus the functionally enriched targets. The downloadable "Auxiliary enrichment file" may contain targets from the single organism IntaRNA prediction that are not in the CopraRNA top list, but fit to the enriched terms. These targets might be specific to the organism of interest. Another interesting candidate is "upp" at prediction rank 10. From a synteny analysis we know that IsaR1 is encoded opposite to the upp 5′ UTR in many organisms [31]. This genomic location leads to a 100% complementarity to the *upp* 5′ UTR in these strains and to an arbitrarily good prediction. Of course, it does not exclude that these interactions might nevertheless be

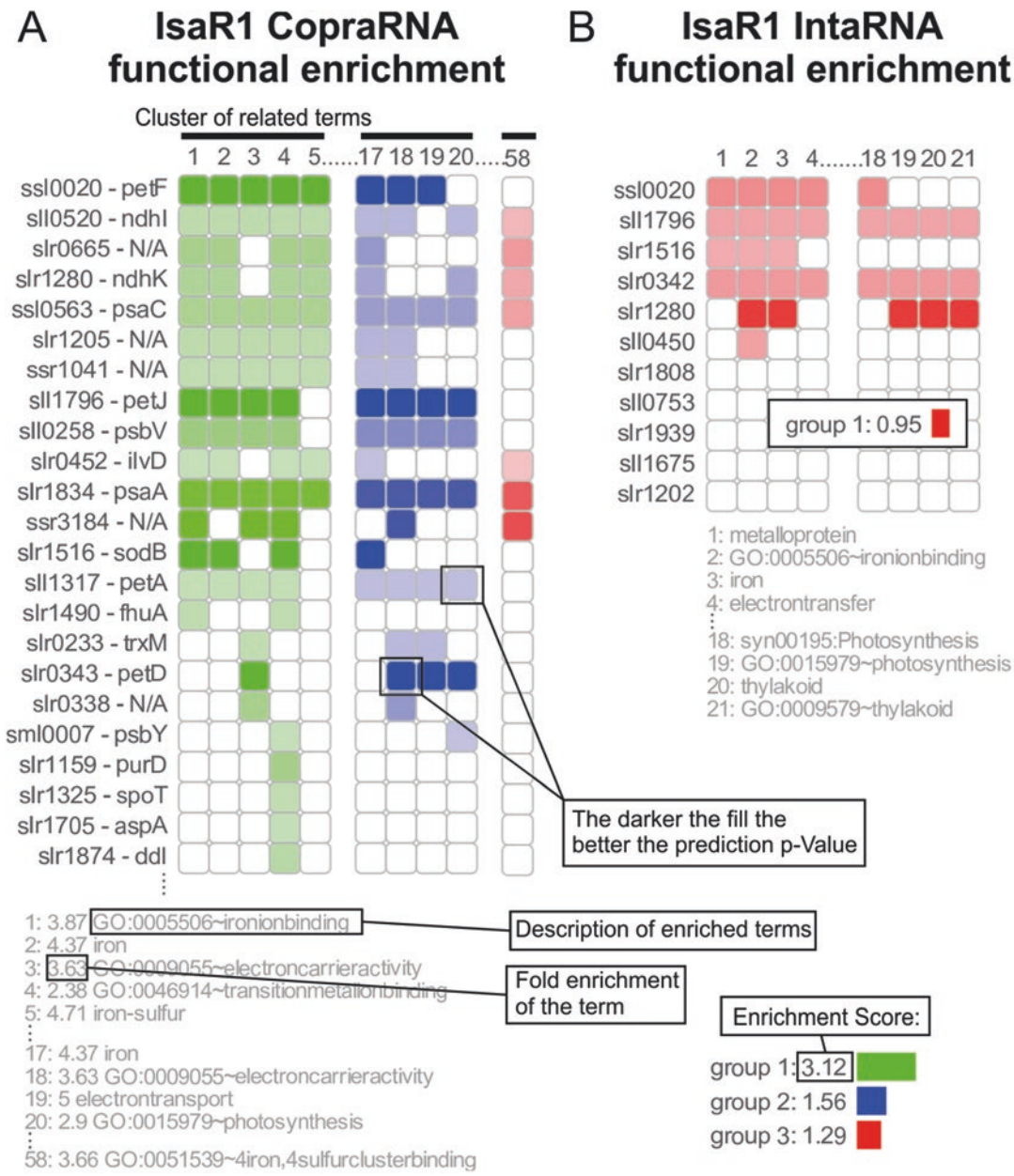


Fig. 5 Visualization of the functional enrichment of the IsaR1 CopraRNA (**a**) and IntaRNA (**b**) predictions. The displayed enrichment heat maps are modified versions of the regular webserver outputs

biologically relevant. Often, these synteny artifacts are rank 1 predictions with an unusually good *p*-value. The mRNA regions plot has a clear peak around the start codon and the predicted interaction sites within the homologous UTRs cluster spatially (Fig. 6). Both results are an indicator of a meaningful prediction. Furthermore, the sRNA regions plot shows a peak around the 5′ end of IsaR1 (Fig. 6). This coincides with the highly conserved

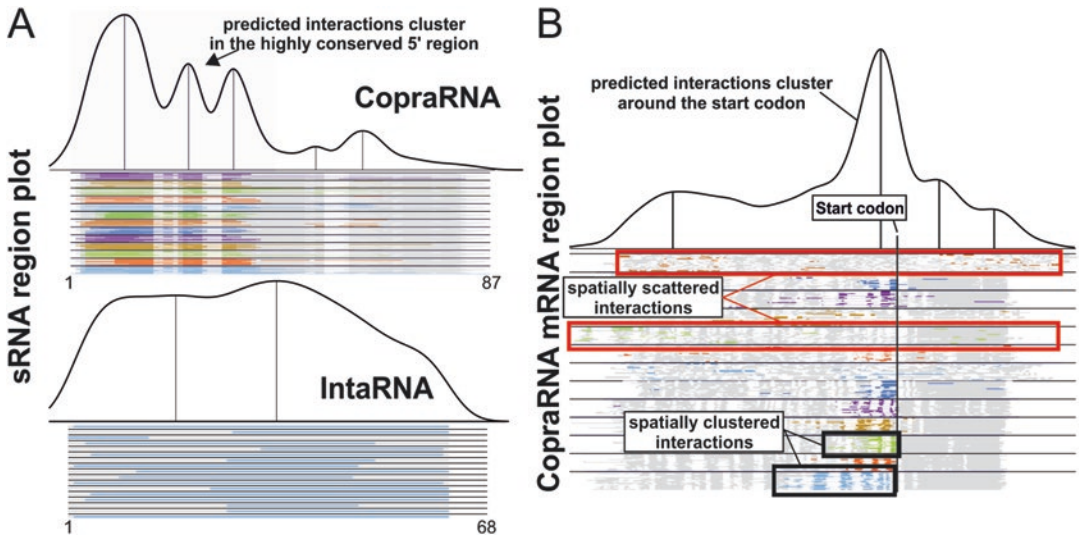


Fig. 6 (**a**) sRNA region plots for IntaRNA and CopraRNA and the CopraRNA mRNA regions plot (**b**) for the use case example. The displayed plots are modified versions of the regular webserver outputs

region that was detected in the multiple sequence alignment of IsaR1. From an a priori perspective, all indicators support that the prediction is sound and that the sRNA is a good candidate for further investigation.

These examples show that the comparative prediction is much more informative and that CopraRNA should be always used if homologs from an sRNA of interest are available. It should be noted that CopraRNA conducts a joint prediction for all input organisms. Each individually high-ranking target might not be regulated in all organisms, but they may be specific to a given subgroup. Pre-computed results of the IntaRNA and CopraRNA IsaR1 use case example runs are accessible at the respective webservers.

3.6 Assessment of Prediction Results

The initial question after each run of a predictive algorithm is whether or not the result is sound enough to merit follow-up experiments in the wet-lab. As stated above, most given RNA sequences will produce some kind of prediction on the genomic scale, even if they do not function by RNA–RNA interaction. The *p*-value statistics are not useful in this respect, so which are the guidelines that can be provided for the classification of meaningfulness?

1. For both IntaRNA and CopraRNA, the functional enrichment score is the best initial approximation for the quality of a specific target prediction. Enrichment scores ≥ 1.3 are considered statistically significant and predictions returning strongly significant terms with many candidates are often a good indicator

of correct predictions and definitely warrant subsequent wet-lab work. CopraRNA prediction together with functional enrichment can singlehandedly reveal the approximate function of an sRNA, e.g., regulation of photosynthesis by PsrR1 [44], regulation of iron homeostasis by RyhB [15] or IsaR1 [31], regulation of amino acid biosynthesis by GcvB [15], and regulation of sugar metabolism by Spot42 [15].

2. There are, however, predictions that do not report overrepresented terms. This can be due to several reasons. Firstly, there may indeed be no broad underlying functional pattern. This is possible for sRNAs that—supposedly—only have a few targets such as MicL [45]. In these cases, the functional enrichment cannot be used as a guideline.
3. Secondly, the functional annotation of the selected organism of interest may be rudimentary or not included in the DAVID database. In this case, it is advisable to change the organism of interest to the species for which the best functional annotation is expected. Alternatively, functional patterns can sometimes be inferred by browsing the "Annotation" column in the prediction results by eye and looking for keywords or phrases that frequently appear. Also, the enrichment analysis can be performed independently from DAVID by manually performing tests such as Fisher's exact test. This strategy was successfully used to uncover an enrichment of cell cycle associated targets for the sRNA EcpR1 in *Sinorhizobium meliloti* [38].
4. Finally, the lack of functionally enriched terms can also indicate poor prediction accuracy. If many sRNAs are being predictively screened for follow-up projects, it is definitely advisable to first focus on candidates with pronounced functional enrichments. If, however, a specific sRNA is centrally interesting, the combination of the predictions with other information sources can be valuable.
5. The integration of the computational data with experimental data can be helpful. Thus, expression patterns of an sRNA or knockout/overexpression phenotypes can aid in the interpretation of the target prediction. In the case of the sRNA NsiR4 [46], the prediction produced no meaningful functional enrichment. However, the sRNA was known to be induced under nitrogen depletion and the rank 1 target was a regulator of nitrogen homeostasis. Ultimately, this prediction turned out to be correct. If a potentially available strain with a genomic manipulation of the sRNA has a phenotype, there might be targets within the top predictions that are suited to explain this phenotype.
6. The mRNA and sRNA regions plots can be indicative of promising results. A pronounced peak around the translation initiation region can be an indicator of a good prediction even if no functionally enriched terms have been reported. However,

functional interactions have been reported far upstream and downstream of the start codon [47–50]. Therefore, the lack of such a peak in the mRNAs region plot does not rule out a good prediction. The interaction sites within a functional RNA–RNA interacting sRNA should also cluster in specific regions, which are likely to have higher sequence conservation than the flanking sequences. The CopraRNA regions plots give information about the respective sequences of all organisms used. The predicted interaction regions of true targets in the sRNA sequences and within the homologous mRNA sequences should cluster in roughly the same area. A widely distributed pattern without clustering argues against a true target.

7. Some functional sRNAs such as ArcZ [51] are not well accessible for target prediction [15]. Thus, the lack of a meaningful prediction does not rule out that the respective sRNA still works by RNA–RNA interaction.

4 Conclusion

A computational analysis should be the starting point for each sRNA analysis, regardless of whether it is a single sRNA candidate or a large-scale survey. The results of the different steps in a bioinformatic workflow can give valuable directions for further investigations. However, it is also important to be aware of the limitations of computational methods and to interpret the results in this respect. Physiologically important sRNAs might not show a distinctive pattern in predictions and may thus be overlooked. Even comparative target predictions often have high numbers of false positives and any given piece of RNA will produce a prediction of some kind. Furthermore, predictions are restricted to known mechanisms. Thus, interactions outside of the UTRs are not detected by the default workflow. Nevertheless, there are clear advantages: In contrast to wet-lab experiments, the presented tools are free of charge and extremely time-efficient. It should be noted that sensitivity and specificity of the CopraRNA target predictions can compete with pulse expression microarray-based methods [15] or interactomics methods [3] when benchmarked against each other. An advantage of computational predictions is that they are—unlike wet-lab experiments—not dependent on specific environmental conditions to uncover a condition-dependent target or function. In fact, predictions and wet-lab methods are often partly complementary and synergistic [15], and superior results are often achieved by the combination of various resources as shown for IsaR1 [31].

References

1. Wagner EGH, Romby P (2015) Chapter 3. Small RNAs in bacteria and Archaea: who they are, what they do, and how they do it. In: Friedmann T, Dunlap JC, Goodwin SF (eds) Advances in genetics. Academic Press, Waltham, pp 133–208
2. Barquist L, Vogel J (2015) Accelerating discovery and functional analysis of small RNAs with new technologies. Annu Rev Genet 49: 367–394. https://doi.org/10.1146/annurev-genet-112414-054804
3. Melamed S, Peer A, Faigenbaum-Romm R, Gatt YE, Reiss N, Bar A, Altuvia Y, Argaman L, Margalit H (2016) Global mapping of small RNA-target interactions in bacteria. Mol Cell 63:884–897. https://doi.org/10.1016/j.molcel.2016.07.026
4. Waters SA, McAteer SP, Kudla G, Pang I, Deshpande NP, Amos TG, Leong KW, Wilkins MR, Strugnell R, Gally DL, Tollervey D, Tree JJ (2017) Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. EMBO J 36:374–387. 10.15252/embj.201694639
5. Storz G, Wolf YI, Ramamurthi KS (2014) Small proteins can no longer be ignored. Annu Rev Biochem 83:753–777. https://doi.org/10.1146/annurev-biochem-070611-102400
6. Gimpel M, Brantl S (2017) Dual-function small regulatory RNAs in bacteria. Mol Microbiol 103:387–397. https://doi.org/10.1111/mmi.13558
7. Neuhaus K, Landstorfer R, Simon S, Schober S, Wright PR, Smith C, Backofen R, Wecko R, Keim DA, Scherer S (2017) Differentiation of ncRNAs from small mRNAs in Escherichia coli O157:H7 EDL933 (EHEC) by combined RNAseq and RIBOseq – ryhB encodes the regulatory RNA RyhB and a peptide, RyhP. BMC Genomics 18:216. https://doi.org/10.1186/s12864-017-3586-9
8. Wadler CS, Vanderpool CK (2007) A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc Natl Acad Sci 104:20454–20459. https://doi.org/10.1073/pnas.0708102104
9. Updegrove TB, Zhang A, Storz G (2016) Hfq: the flexible RNA matchmaker. Curr Opin Microbiol 30:133–138. https://doi.org/10.1016/j.mib.2016.02.003
10. Olejniczak M, Storz G (2017) ProQ/FinO-domain proteins: another ubiquitous family of RNA matchmakers? Mol Microbiol 104:905–915. https://doi.org/10.1111/mmi.13679
11. Smirnov A, Förstner KU, Holmqvist E, Otto A, Günster R, Becher D, Reinhardt R, Vogel J (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. Proc Natl Acad Sci U S A 113:11591–11596. https://doi.org/10.1073/pnas.1609981113
12. Cavanagh AT, Wassarman KM (2014) 6S RNA, a global regulator of transcription in Escherichia coli, Bacillus subtilis, and beyond. Annu Rev Microbiol 68:45–60. https://doi.org/10.1146/annurev-micro-092611-150135
13. Holmqvist E, Wright PR, Li L, Bischler T, Barquist L, Reinhardt R, Backofen R, Vogel J (2016) Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. EMBO J 35:991–1011. 10.15252/embj.201593360
14. Lott S, Schäfer R, Mann M, Hess WR, Voß B, Georg J GLASSgo - Automated and reliable detection of sRNA homologs from a single input sequence. (submitted)
15. Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R, Georg J (2013) Comparative genomics boosts target prediction for bacterial small RNAs. Proc Natl Acad Sci 110:E3487–E3496. https://doi.org/10.1073/pnas.1303248110
16. Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, Kleinkauf R, Hess WR, Backofen R (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. Nucleic Acids Res 42:W119–W123. https://doi.org/10.1093/nar/gku359
17. Patrick R. Wright (2016) Predicting small RNA targets in prokaryotes - a challenge beyond the barriers of thermodynamic models. https://freidok.uni-freiburg.de/data/11472
18. Gruber AR, Findeiß S, Washietl S, Hofacker IL, Stadler PF (2010) RNAz 2.0: improved noncoding RNA detection. Pac Symp Biocomput:69–79
19. Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. RNA 18:900–914. https://doi.org/10.1261/rna.029041.111
20. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF (2008) RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics 9:474. https://doi.org/10.1186/1471-2105-9-474

21. Washietl S, Findeiß S, Müller SA, Kalkhof S, von Bergen M, Hofacker IL, Stadler PF, Goldman N (2011) RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. RNA 17:578–594. https://doi.org/10.1261/rna.2536111
22. Busch A, Richter AS, Backofen R (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics 24:2849–2856. https://doi.org/10.1093/bioinformatics/btn544
23. Mann M, Wright PR, Backofen R (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. Nucleic Acids Res. https://doi.org/10.1093/nar/gkx279
24. Menzel P, Gorodkin J, Stadler PF (2009) The tedious task of finding homologous noncoding RNA genes. RNA 15:2075–2082. https://doi.org/10.1261/rna.1556009
25. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD (2015) Rfam 12.0: updates to the RNA families database. Nucleic Acids Res 43:D130–D137. https://doi.org/10.1093/nar/gku1063
26. Eggenhofer F, Hofacker IL, Höner zu Siederdissen C (2016) RNAlien – unsupervised RNA family model construction. Nucleic Acids Res 44:8433–8441. https://doi.org/10.1093/nar/gkw558
27. Katoh K, Toh H (2008) Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. BMC Bioinformatics 9:212. https://doi.org/10.1186/1471-2105-9-212
28. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191. https://doi.org/10.1093/bioinformatics/btp033
29. Morita T, Nishino R, Aiba H (2017) Role of terminator hairpin in biogenesis of functional Hfq-binding sRNAs. RNA 23:1419–1431. https://doi.org/10.1261/rna.060756.117
30. Lagares A, Roux I, Valverde C (2016) Phylogenetic distribution and evolutionary pattern of an α-proteobacterial small RNA gene that controls polyhydroxybutyrate accumulation in Sinorhizobium meliloti. Mol Phylogenet Evol 99:182–193. https://doi.org/10.1016/j.ympev.2016.03.026
31. Georg J, Kostova G, Vuorijoki L, Schön V, Kadowaki T, Huokko T, Baumgartner D, Müller M, Klähn S, Allahverdiyeva Y, Hihara Y, Futschik ME, Aro E-M, Hess WR (2017) Acclimation of oxygenic photosynthesis to iron starvation is controlled by the sRNA IsaR1. Curr Biol. https://doi.org/10.1016/j.cub.2017.04.010
32. Kery MB, Feldman M, Livny J, Tjaden B (2014) TargetRNA2: identifying targets of small regulatory RNAs in bacteria. Nucleic Acids Res 42:W124–W129. https://doi.org/10.1093/nar/gku317
33. Eggenhofer F, Tafer H, Stadler PF, Hofacker IL (2011) RNApredator: fast accessibility-based prediction of sRNA targets. Nucleic Acids Res 39:W149–W154. https://doi.org/10.1093/nar/gkr467
34. Pain A, Ott A, Amine H, Rochat T, Bouloc P, Gautheret D (2015) An assessment of bacterial small RNA target prediction programs. RNA Biol 12:509–513. https://doi.org/10.1080/15476286.2015.1020269
35. Umu SU, Gardner PP (2017) A comprehensive benchmark of RNA–RNA interaction prediction tools for all domains of life. Bioinformatics 33:988–996. https://doi.org/10.1093/bioinformatics/btw728
36. Babski J, Maier L-K, Heyer R, Jaschinski K, Prasse D, Jäger D, Randau L, Schmitz RA, Marchfelder A, Soppa J (2014) Small regulatory RNAs in Archaea. RNA Biol 11:484–493. https://doi.org/10.4161/rna.28452
37. Jiao X, Sherman BT, Huang DW, Stephens R, Baseler MW, Lane HC, Lempicki RA (2012) DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics 28:1805–1806. https://doi.org/10.1093/bioinformatics/bts251
38. Robledo M, Frage B, Wright PR, Becker A (2015) A stress-induced small RNA modulates alpha-rhizobial cell cycle progression. PLoS Genet 11:e1005153. https://doi.org/10.1371/journal.pgen.1005153
39. Robledo M, Peregrina A, Millán V, García-Tomsig NI, Torres-Quesada O, Mateos PF, Becker A, Jiménez-Zurdo JI (2017) A conserved α-proteobacterial small RNA contributes to osmoadaptation and symbiotic efficiency of rhizobia on legume roots. Environ Microbiol 9:2661–2680. https://doi.org/10.1111/1462-2920.13757
40. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein

D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R, Gene Ontology Consortium (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258–D261. https://doi.org/10.1093/nar/gkh036

41. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27
42. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1–13. https://doi.org/10.1093/nar/gkn923
43. Hernández-Prieto MA, Schön V, Georg J, Barreira L, Varela J, Hess WR, Futschik ME (2012) Iron deprivation in *Synechocystis*: inference of pathways, non-coding RNAs, and regulatory elements from comprehensive expression profiling. G3 2:1475–1495. https://doi.org/10.1534/g3.112.003863
44. Georg J, Dienst D, Schürgers N, Wallner T, Kopp D, Stazic D, Kuchmina E, Klähn S, Lokstein H, Hess WR, Wilde A (2014) The small regulatory RNA SyR1/PsrR1 controls photosynthetic functions in cyanobacteria. Plant Cell 26:3661–3679. https://doi.org/10.1105/tpc.114.129767
45. Guo MS, Updegrove TB, Gogol EB, Shabalina SA, Gross CA, Storz G (2014) MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. Genes Dev 28:1620–1634. https://doi.org/10.1101/gad.243485.114
46. Klähn S, Schaal C, Georg J, Baumgartner D, Knippen G, Hagemann M, Muro-Pastor AM, Hess WR (2015) The sRNA NsiR4 is involved in nitrogen assimilation control in cyanobacteria by targeting glutamine synthetase inactivating factor IF7. Proc Natl Acad Sci U S A 112:E6243–E6252. https://doi.org/10.1073/pnas.1508412112
47. Desnoyers G, Massé E (2012) Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. Genes Dev 26:726–739. https://doi.org/10.1101/gad.182493.111
48. Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JCD, Vogel J (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. Mol Microbiol 84:428–445. https://doi.org/10.1111/j.1365-2958.2012.08031.x
49. Sharma CM, Darfeuille F, Plantinga TH, Vogel J (2007) A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. Genes Dev 21:2804–2817. https://doi.org/10.1101/gad.447207
50. Pfeiffer V, Papenfort K, Lucchini S, Hinton JCD, Vogel J (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nat Struct Mol Biol 16:840–846. https://doi.org/10.1038/nsmb.1631
51. Papenfort K, Said N, Welsink T, Lucchini S, Hinton JCD, Vogel J (2009) Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. Mol Microbiol 74:139–158. https://doi.org/10.1111/j.1365-2958.2009.06857.x

Chapter 2

Guidelines for Inferring and Characterizing a Family of Bacterial *trans*-Acting Small Noncoding RNAs

Antonio Lagares Jr. and Claudio Valverde

Abstract

So far, every sequenced bacterial transcriptome encompasses hundreds of small regulatory noncoding RNAs (sRNAs). From those sRNAs that have been already characterized, we learned that their regulatory functions could span over almost every bacterial process, mostly acting at the posttranscriptional control of gene expression (Wagner and Romby, Adv Genet 90:133–208, 2015). Canonical molecular mechanisms of sRNA action have been described to rely on both sequence and/or structural traits of the RNA molecule. As for protein-coding genes, the conservation of sRNAs among species suggests conserved and adjusted functions across evolution. Knowing the phylogenetic distribution of an sRNA gene and how its functional traits have evolved may help to get a broad picture of its biological role in each single species. Here, we present a simple computational workflow to identify close and distant sRNA homologs present in sequenced bacterial genomes, which allows defining novel sRNA families. This strategy is based on the use of Covariance Models (CM) and assumes the conservation of sequence and structure of functional sRNA genes throughout evolution. Moreover, by carefully inspecting the conservation of the close genomic context of every member of the RNA family and how the patterns of microsynteny follow the path of species evolution, it is possible to define subgroups of sRNA orthologs, which in turn enables the definition of RNA subfamilies.

Key words sRNA, Homology, Phylogeny, Covariance model, RNA families

1 Introduction

Small noncoding regulatory RNAs (sRNAs) play a major role in the fine-tuning of bacterial gene expression. These RNAs range in size from 50 to 400 nucleotides and regulate in *trans* the activity of target RNAs or proteins by direct molecular interactions that take place in vivo [1]. Canonical mechanisms of small RNA-based regulation have been described to rely on sequence and/or structure traits, depending on the nature of the targets. As for protein-coding genes, sRNA genes can show very limited phylogenetic distribution—in some cases it could be possibly underestimated due to limitations in the availability of genomic information—or they can be broadly distributed (e.g., the *Escherichia coli* sRNAs

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_2,

Spot-42, GcvB, and RyhB [2]; and the alpha-proteobacterial sRNA MmgR [3]). Multiple mechanisms have been described to drive small RNA-target co-evolution [4], which shape ancient regulations in divergent species. Integrating information about the phylogenetic distribution of a conserved sRNA gene during its functional characterization clearly helps to build up a broad picture of its physiological role in each single species.

In a broad sense, RNA families are rather defined on the base of the methods used to search for homolog genes and the corresponding criteria to infer homology, which are largely dependent on sequence/structural considerations. Automated tools such as RNAlien [5] and GLASSgo became recently available for the detection of sRNA homologs (the latter described in more detail in this book by Wright and Georg, Chapter 10), although they only take into account sequence conservation. Since it is well known that sRNA function may be conserved without significant sequence conservation, the identification of sRNA homolog genes based on simple sequence homology searches could be sometimes hampered. In this regard, computational algorithms that mind both sequence and structure conservation during evolution have been developed to model RNA families and allow performing predictions of RNA homolog sequences in genomic databases. Among them, Covariance Models (CM) [6] are a special case of profile stochastic context-free grammar that score both sequence and RNA secondary structure consensus and could be used to search DNA databases for RNA structure and sequence similarities. In fact, Rfam database is a collection of RNA families that are modeled by CMs [7]. Even though more challenging and hence less commonly implemented, thermodynamic matchers [8] constitute a suitable alternative strategy to model RNA families in which almost only structural features, but little sequence, are thought to be conserved.

This chapter describes a simple workflow to identify and characterize novel small noncoding RNA families based on the use of CM, which requires basic handling of elemental computational resources and thus can be easily followed by nonexpert users.

Stated in brief, an exhaustive search for homologous genes exclusively based in sequence conservation is performed initially in order to build a starting multiple sequence alignment that supports a consensus structure (Fig. 1, Phases 1 and 2). This alignment is then used as the seed to model the initial CM profile, which after its calibration is used to search against a database of genomic DNA sequences for remote homologous genes in which RNA structural features may have constrained the evolution of the gene instead of just the sequence itself (Fig. 1, Phase 3). The steps of CM construction, database search, and homolog identification are repeated upon incorporation of the newly identified homologs into the multiple alignment after each iteration. The process stops when the

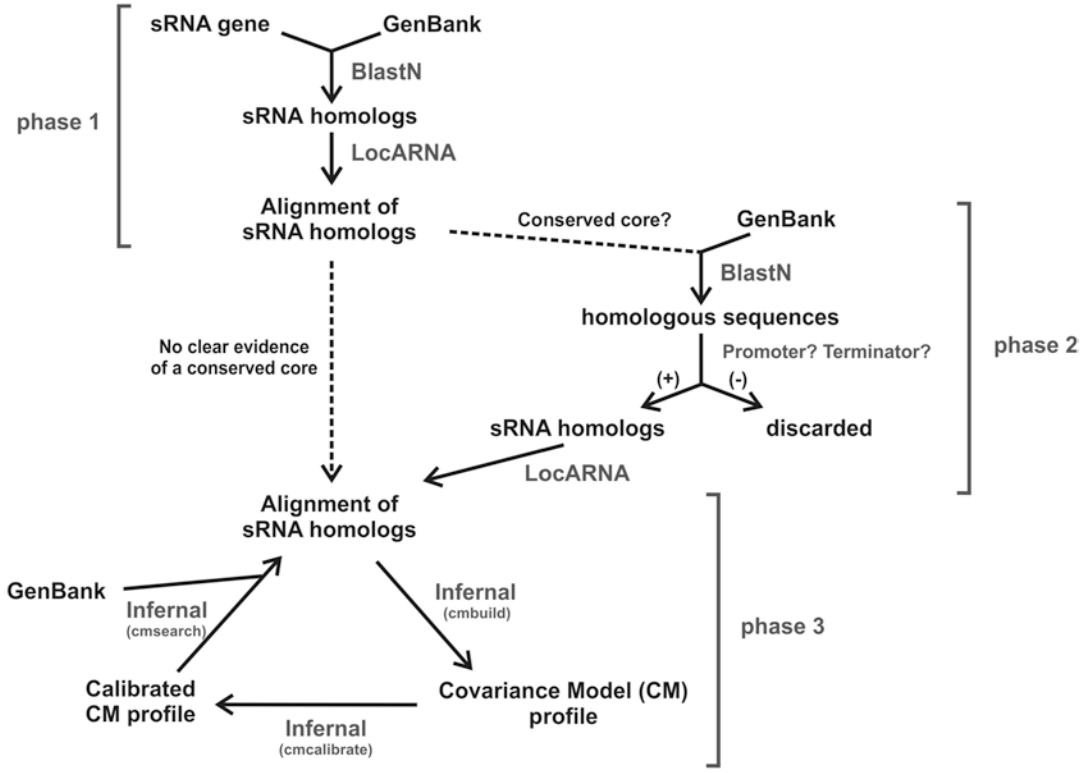


Fig. 1 Workflow for sRNA homologs search and novel RNA family construction

number of new hits reaches a plateau, and the entire set of detected homologs is defined as a novel RNA family. Additionally, the identity of the flanking genes to each homolog sequence is retrieved and a manual inspection for the presence of microsynteny is performed, which could be used as a criterion to define subfamilies of gene orthologs.

2 Materials and Installation Guidelines

- Linux and Windows operating systems.
- Infernal and Easel library packages (for Linux; *see* Subheading 2.1 for downloading and installation guidelines) [9].
- LocARNA package [10] (for Linux; *see* Subheading 2.2 for downloading and installation guidelines).
- Complete bacterial genome sequences currently available at NCBI (*see* Subheading 2.3 for downloading and unpacking guidelines).
- blastn resource (available at https://blast.ncbi.nlm.nih.gov/Blast.cgi).

- SILVA rRNA database (online at https://www.arb-silva.de/search/).
- MEGA7 software for Windows [11] (available for download at http://www.megasoftware.net/).
- Microbial genomic context viewer [12] (MGcV; available at http://mgcv.cmbi.ru.nl/).

2.1 How to Download, Compile, and Install Infernal and Easel Library Packages for Linux (Following the Developer's User Guide)

```
> wget eddylab.org/infernal/infernal-1.1.2.tar.gz
> tar xf infernal-1.1.2.tar.gz
> cd infernal-1.1.2
> ./configure
> make
> make check
> make install
> cd easel; make install
```

2.2 LocARNA Package Installation in Linux

- Download miniconda for Linux (https://conda.io/miniconda.html) and install it:

```
> bash Miniconda3-latest-Linux-x86_64.sh
```

- Add package bioconda.

```
> conda config --add channels bioconda
```

- Install LocARNA (together with the needed packages) using conda:

```
> conda install locarna
```

2.3 How to Download the Complete Bacterial Genomic Sequences from NCBI Assembly Server and Compile Them into a Single FASTA File

- Enter https://www.ncbi.nlm.nih.gov/assembly/, and search bacteria.
- Filter: Status (latest), Assembly level (complete genome).
- Download assemblies from GenBank source database as Genomic FASTA (by default, they would be downloaded as a single packed file named genome_assemblies.tar).
- To unpack the .tar file and get all the corresponding genome sequence files (.fna formatted), enter the Linux terminal, locate into the directory where the packed .tar file is stored, and run the following command:

```
> tar -xvf genome_assemblies.tar
```

- Then, locate inside the unpacked folder and run the following command to decompress all genome sequences:

```
> gzip -d *.gz
```

- Merge all genome sequence files into a single FASTA file by running (*see* **Note 1**):

```
> cat *.fna > ../allbacterialgenomes.fasta
```

3 Methods

Subheadings 3.1 and 3.2 are meant to be performed in a Linux platform. To simplify the syntax, it is requested that every file that is used as input in each of these steps is copied into the same predefined working directory. At the beginning, set this directory as the current directory in the terminal.

3.1 Initial Extensive Search of sRNA Homologs Based on Sequence Conservation (Phases 1 and 2)

1. Query the NCBI nonredundant database of bacterial organism sequences using blastn (https://blast.ncbi.nlm.nih.gov/Blast.cgi; with default parameters) with the sRNA gene sequence as input. Select all significant hits (e.g., those with an *E*-value <0.0001) and download the aligned sequences as a single FASTA file (*see* Fig. 2; in the example, the output file is called sRNA_initialblastn_homologs.fasta).
2. Create a multiple sequence alignment of the sRNA homologous sequences collected in Subheading 3.1, **step 1** using LocARNA [10] (run with default parameters). LocARNA generates a multiple sequence alignment with consensus structure information as output. RNAalifold plots the alignment in an output file (*see* results in Fig. 3). To perform this step, run the following command at the terminal (*see* **Note 2**).

```
> mlocarna --stockholm sRNA_initialblastn_homologs.
fasta
> cd sRNA_initialblastn_homologs.out/results
> RNAalifold --aln result.aln
# > mlocarna [options] <FASTA file with input se-
quences to be aligned>
# > RNAalifold [options] <sequence alignment file to
be plotted>
```

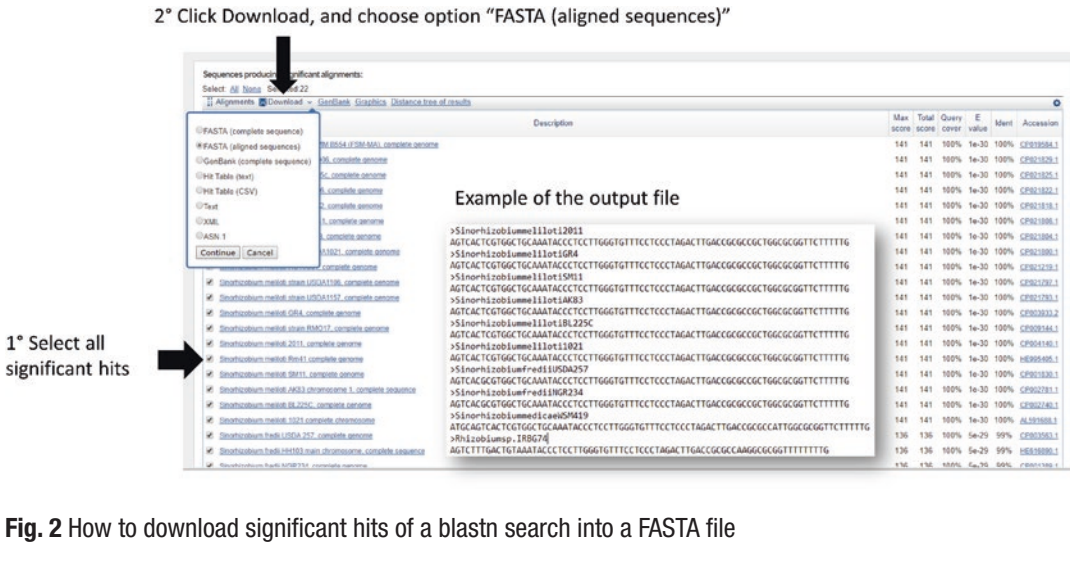


Fig. 2 How to download significant hits of a blastn search into a FASTA file

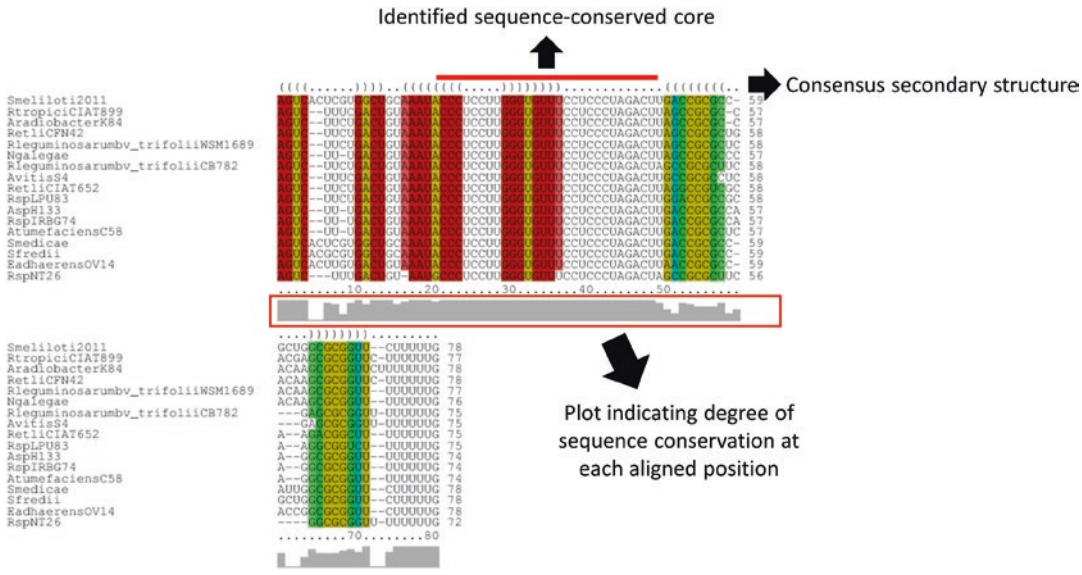


Fig. 3 LocARNA-based multiple sequence and structure alignment of blastn hits

It is recommended to include `--stockholm` in the command line in order to get an output file already formatted for the following step of Covariance Model construction. Input sequences must be manually inspected and filtered by length match. An output alignment file called result_out.sto will be created by default. For the sake of clarity, it is recommended to rename this file (e.g., as aligned_sRNA_initialblastn_homologs.sto).

3. Open the multiple alignment file created in Subheading 3.1, **step 2** and look it over to search for the presence of a conserved core of sequence among the homologs. This is relevant because carrying out a new round of blastn search using the identified highly conserved region of the sRNA sequence as the query would improve the sensitivity of the search. The rationale is that such a refined search may allow finding further homologs in which a higher degree of sequence divergence may have taken place within RNA domains not directly involved in the sRNA mechanism of action. The degree of sequence conservation at each aligned position is indicated as a gray bar chart that can be inspected in the output figure (.ps formatted) under the alignment (*see* Fig. 3). If a conserved internal sequence core is identified, follow Subheading 3.1, **step 4** (Phase 2); otherwise, continue with Subheading 3.2, **step 1** (Phase 3).
4. Run a second round of blastn search (with default parameters) with the consensus sequence of the sRNA conserved core as the query (Fig. 4) against the database of bacterial genomes.

```
>sRNA_conserved_core
CCCTCCTTGGGTGTTTCCTCCCTAGACTT
```

Fig. 4 Example of the sRNA conserved core consensus sequence extracted from the alignment shown in Fig. 3 that is used as query at Subheading 3.1, **step 4**

Manually download the homolog sequences found (e.g., those hits with *E*-value <0.0001), together with the corresponding 200 nt-long genomic sequences laying immediately upstream and downstream of each aligned region (this range is set as it is expected to cover enough genomic sequence to be able to reconstruct the complete gene at next step) (*see* **Note 3**).

5. Check for the presence of a putative promoter (e.g., run the Neural Network promoter prediction sever BDGP at http://www.fruitfly.org/seq_tools/promoter.html [13]) and a putative *Rho*-independent terminator (e.g., run the ARNold server at http://rna.igmors.u-psud.fr/toolbox/arnold/ [14]) located in the same genomic DNA strand within the regions immediately upstream and downstream of each homolog core sequence, respectively (*see* **Note 4**). This filter would impede collecting homologs whose promoters are highly divergent from those detected by the available algorithms (which in fact could potentially be detected later in the workflow by means of the Covariance Model), but would keep the stringency high enough to reduce the inclusion of false positives (i.e., homolog core sequences that diverged long enough to lose their identity as transcriptional units coding for *trans*-acting sRNAs). Retain only those sequence hits that are predicted to be flanked by both functional features, and discard those reconstructed homolog sRNA sequences that differ in length from the original input (*see* **Note 5**). Compile the filtered sRNA sequences into a new FASTA file (in our example, named close&remotesequencehomologs.fasta).
6. Run LocARNA to create a multiple sequence alignment with consensus structure annotation using the updated set of sRNA homologs. At the terminal, run the following command:

```
> mlocarna --stockholm
close&remotesequencehomologs.fasta
```

Rename the output file (by default, result_out.sto) as aligned_close&remotesequencehomologs.sto and manually inspect and edit the alignment in order to improve its quality. Dismiss the alignment created in Subheading 3.1, **step 2**.

3.2 Expanded Search of Remote sRNA Homologs with Structural Covariance, and Definition of a Novel RNA Family (Phase 3)

1. Build and calibrate a new Covariance Model (CM) profile from the input multiple sequence alignment.

   A carefully curated multiple homolog sequence and secondary structure alignment is required ahead CM construction (*see* **Note 6**). In order to build the initial CM, the following command should be run on the terminal:

   ```
   > cmbuild Cmfirstiteration.cm aligned_close&
   remotesequencehomologs.sto
   # > cmbuild <name of the new CM file> <name of the
   input alignment>
   ```

   As a result, a CM profile (called Cmfirstiteration.cm) will be created based on the alignment created and edited in Subheading 3.1, **step 6**, and a summary of the main properties of the resulting profile and its building process will be shown at the terminal.

   Since in this chapter the CM is meant to be used for searching within a genome database for homologous sequences matching the profile, it is mandatory to calibrate the CM in advance. This step is performed by the cmcalibrate program, and as a result, it enables to later estimate the statistical significance (*E*-value) of the hits found using the calibrated CM when searching against a nucleotide database. The calibration is performed with the command as follows:

   ```
   > cmcalibrate Cmfirstiteration.cm
   # > cmcalibrate <name of the CM file to be calibrat-
   ed>
   ```

2. Search the genome database with the CM profile.

   After calibration, the CM (i.e., Cmfirstiteration.cm) is ready to be used to query the genome database (i.e., allbacterialgenomes.fasta) for sequences that share conserved sequence and/or structure features. To this end, cmsearch program is executed with the following command line:

   ```
   > cmsearch --tblout resultfirstiteration.txt
   --incE 0.0001 -A resultfirstiteration.sto Cm-
   firstiteration.cm allbacterialgenomes.fasta
   # > cmsearch [options] <name of the CM file> <name
   of the database>
   ```

   It is useful to include the --tblout option, which saves the output results information in a single table at the specified file resultsfirstiteration.txt. The program could also be set up to include in the output alignment only those hits whose *E*-value is lower than a predefined significance threshold (e.g., those with *E*-values <0.0001) by adding the option --incE followed by the desired set value. The resulting aligned homolog sequences are saved into the file resultfirstiteration.sto.

3. CM maturation by iteration of CM build, calibration and database search steps (Subheading 3.2, **steps 1–3**).

   In order to gain sensitivity aiming to collect novel and remotely related homolog sequences, the initial CM profile has to be updated by incorporating all the new detected sequences. For this purpose, the output alignment generated in Subheading 3.2, **step 2** (i.e., resultfirstiteration.sto) is used as input to build and calibrate a novel CM (i.e., CMseconditeration.cm), as follows:

   ```
   > cmbuild Cmseconditeration.cm resultfirstiteration.
   sto
   ```

   And then,

   ```
   > cmcalibrate Cmseconditeration.cm
   ```

   Finally, the file Cmseconditeration.cm is used as the query to repeat the genomic database search.

   ```
   > cmsearch --tblout resultseconditeration.txt
   --incE 0.0001 -A resultseconditeration.sto Cmsecon-
   diteration.cm allbacterialgenomes.fasta
   ```

   It is suggested to stop the iterative loops when the number of newly recruited significant sequence hits at the cycle i is zero, or it is higher than the number of new significant sequences rendered by the cycle (i—1).

   The complete set of sRNA homologs that were detected using this workflow is defined as a novel RNA family.

3.3 Subfamilies of Orthologs: Microsynteny as an Indicator of Gene Orthology

1. Retrieving information of the genomic context of sRNA homologs

   To assess whether there are groups of sRNA homologs that display local synteny (microsynteny) at their genomic positions, the identity of the open reading frames flanking each homolog in its respective replicon has to be first determined. The online server *Microbial Genomic context Viewer* (MGcV), a tool for the visualization of small scale genome regions, could be used for this purpose. To this end, from the output table created after the last iteration performed with `cmsearch`, a comma separated value formatted file has to be created, including the following features in consecutive lines: genomic NC_code, sRNA start- and end- genomic coordinates for each sRNA homolog (an example is shown Fig. 5; *see* **Note 7**). This data is used as input for the MGcV (after selecting *Genomic positions* as Input type). It is recommended to color the displayed genes by the COG (Cluster of Orthologous Groups of proteins [15]) to which they belong (or alternatively the corresponding PFAM [16], if any), in order to be able to identify those regions that show microsynteny at a glance. Genomic regions located between the start and

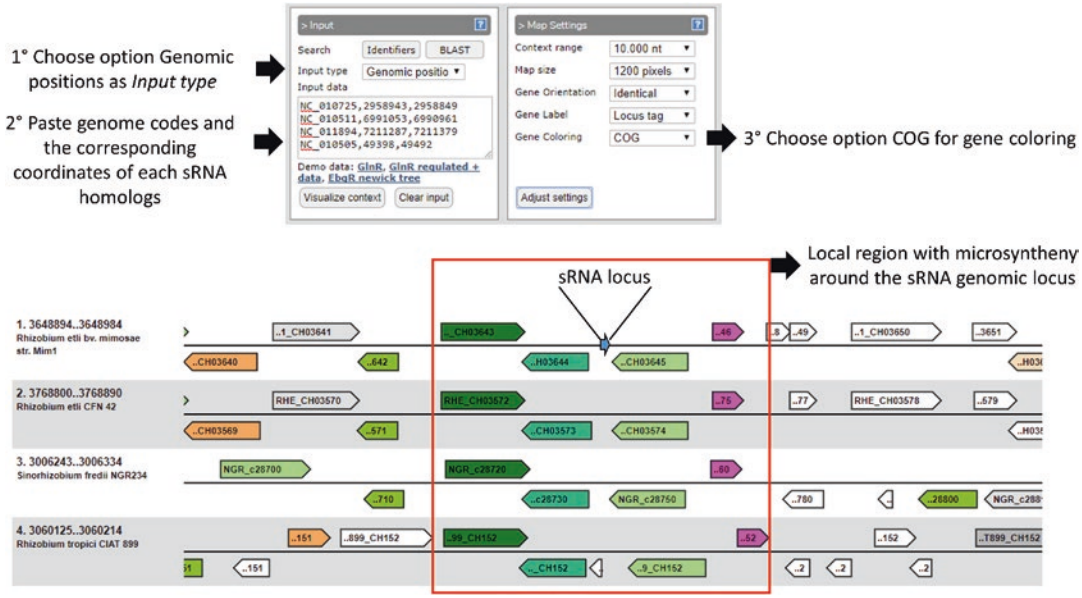


Fig. 5 Microsynteny around an sRNA locus, as revealed by using the Microbial Genomic context Viewer

end coordinates entered as input will be displayed in the center of each hit (*see* **Note 8**). Further manual inspection at the identity and homology of the flanking genes must be carried out with those initially preselected candidate regions.

2. If any specific protein-coding gene is found to be physically linked to a subfamily of orthologous sRNA genes (OLPCG, *o*rtholog-*l*inked *p*rotein-*c*oding *g*ene), it could be also informative to reverse the approach and to characterize the intergenic regions flanking the annotated homologs of the OLPCG in phylogenetically related species to understand the mechanisms of sRNA ortholog gene loss or lack of inheritance at those empty intergenic regions. This can be also performed with the MGcV (*see* Fig. 6 for a description of the workflow). At the >Input window, perform a BLAST search into the database of selected genomes with the amino acid sequence of the OLPCG of interest. Select those significant hits to visualize their genomic context.

3.4 Mapping Ortholog Distribution into a Phylogenetic Tree of Species

It would be informative to analyze how ortholog subfamilies are distributed in the phylogenetic tree of species in order to easily visualize those clades in which particular sRNA orthologs may have been lost, and to infer to what extent vertical and horizontal gene transfer (HGT) may have contributed to their spreading. The analysis of the phylogenetic incongruence is a powerful strategy for inferring HGT [17]. The simple visualization of the occurrence of orthologs in only one species or a clade (presence/absence) within

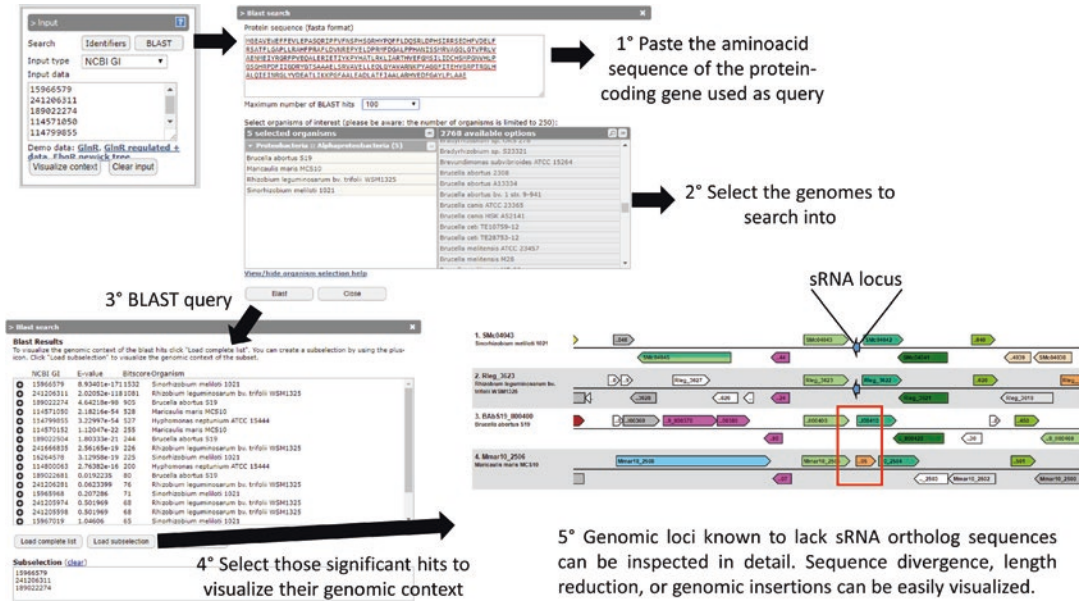


Fig. 6 Using MGcV to visualize the genomic context of sRNA ortholog-linked homolog ORFs

a phylogeny is a direct evidence of gene acquisition events by HGT [18]. This information could be relevant later on to understand how novel regulatory networks are developed at the early stages upon gene arrival.

1. Building a phylogenetic cladogram of species

   A consensus cladogram representing the species phylogeny could be inferred by using the corresponding 16 rDNA gene sequences [19]. In order to get a comprehensive picture, we recommend considering at least one representative species of related taxonomic groups that do not bear any sRNA homolog in addition to those species in which members of the RNA family have been detected. 16 rDNA gene sequences could be manually downloaded from the Search Option at the SILVA server as a single multiple sequence FASTA file (*see* **Note 9**).

   At MEGA7 software, proceed to import the sequences from the downloaded file (Align → Edit/Build Alignment → Retrieve sequences from a file). Select all the sequences, and proceed to align them using the built-in ClustalW tool at the MEGA7 Alignment Explorer (Alignment → Align by ClustalW; default parameters could be kept). Activate the aligned sequences to be used as input for phylogenetic analysis (Data → Phylogenetic analysis) and construct a phylogenetic tree from the main window of MEGA7 (Phylogeny → Construct/Test Neighbor-Joining Tree ([*see* **Note 10**]; suggested settings for tree statistical validation: Test of phylogeny: Bootstrap method, No. of Bootstrap replications: 500).

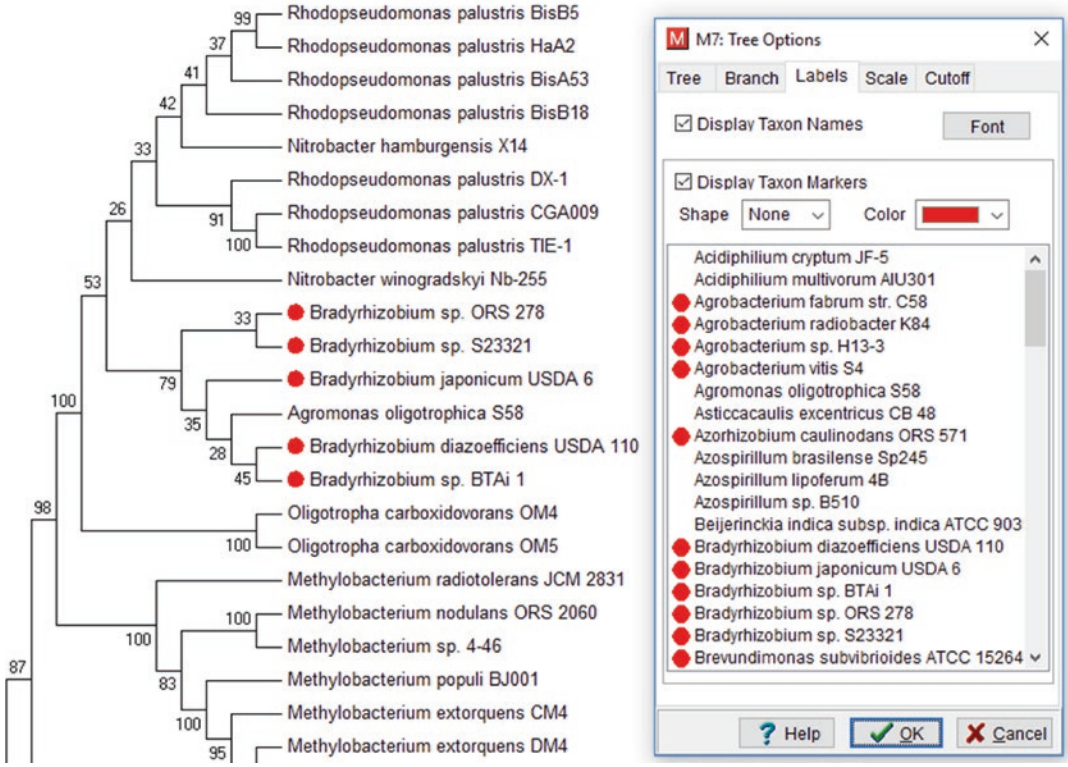


Fig. 7 Mapping orthologous groups with taxon markers in MEGA7 Tree Explorer

2. Mapping of sRNA ortholog subfamilies into the phylogenetic tree

   Distinctive taxon markers can be added to taxa in which members of a particular subfamily of orthologs were found. At Mega Tree Explorer, go to View tab and enter to Options. Select the Labels tab, and assign a corresponding marker to each group of taxa (Fig. 7).

4 Case of Study

The phylogenetic analysis of the alpha-proteobacterial *mmgR* gene has been performed following the workflow described in this chapter [3]. An initial search of homolog sequences of the *S. meliloti* small RNA MmgR retrieved a set of 33 sequence hits, all of them belonging to species of the *Rhizobiaceae* family (phase 1). Upon the recognition of an internal fully conserved 28-mer that was shared between all the initially identified sequences, the conserved motif was used as query to perform a new search against the database and 85 homolog sequences were found, from which only 76

sequences were located in intergenic regions and predicted to be flanked by both a promoter and a *Rho*-independent terminator, which led thus to be further considered as possible true sRNA homologs (phase 2). It is worth to mention that the fact that 76 homolog sequences were distributed among 67 sequenced alpha-proteobacterial genomes—no longer restricted to *Rhizobiaceae*—indirectly implied the presence of at least 9 paralogs in the set. A CM profile was built from the multiple alignment of the sRNA homolog sequences, which was calibrated and finally used to query the bacterial genome database for homolog genes with sequence and structural conserved traits. With this approach, 137 novel homologs were detected, broadening the phylogenetic distribution of the RNA family to 93 strains. The initial multiple alignment of homologs was updated upon incorporation of the new set of detected sequences, a new CM was built and calibrated, and a second round of search with CM against the genomic database was performed. This process was iterated three times, ending up with a total of 243 homolog genes distributed in 95 genomes. From these results, it becomes clear that the CM-based iterative homolog search strategy was mainly helpful for expanding the scope of detection of *mmgR* gene paralogs. The complete set of *mmgR* homologs defined a novel RNA family (alpha-r8). After an extensive inspection at the local genomic region of each sRNA homolog, a remarkable microsynteny was found between a large group of *mmgR* homologs and homologs of an N-formylglutamate amidohydrolase-coding gene. This finding allowed the definition of a subfamily of orthologs, alpha-r8s1, which has a long evolutionary existence within the alpha-proteobacteria.

5 Notes

1. On June 2017, the file comprising the concatenated bacterial genomes sizes around 30 Gigabytes. This would be the only such heavy file needed during the present computational workflow. The concatenated file will be saved in the parental directory, that is, the folder one level up to that where the individual sequence files are located (as indicated by the following command ../)
2. Be aware that upper and lower cases mind. For simplicity, avoid using spaces in file names.
3. If multiple independent conserved regions are identified within a sRNA sequence, proceed to repeat Subheading 3.1, **steps 4** and **5** with each one of them, and collect all homolog sequences to perform Subheading 3.1, **step 6**.
4. Since promoter consensus sequences vary considerably among species and prediction algorithms are not optimized for each

single species, it is recommended to try lowering the score-threshold for promoter detection and check if any predicted transcriptional start site matches the relative start position of the sRNA transcript.

5. If the length of the sequence spanning between promoter and terminator does not match the length of the original sRNA sequence, it will introduce gaps into the multiple sequence alignment and lower the power of database search. If novel sequence motifs have been acquired or lost in any found homolog, do not include it in the alignment.
6. Stockholm alignments can be manually edited with a text editor to improve their quality.
7. NC_codes directly taken from the corresponding hits that are listed in the cmsearch output table must be trimmed immediately before the dot. For example, NC_0030471 has to be shortened to NC_003047
8. When the input type is set as Genomic positions, it is not possible to reverse gene orientation; thus, data will be mirrored when sRNAs are annotated in genomic minus strands.
9. At the SILVA webpage, enter the Search tab and select the option "Search." Then, type the required organism name, strain, or accession number in the SSU 128 (for *Small Subunit*) database. Select a suitable hit by clicking on the corresponding open box located at the left and the sequence will be automatically added to the cart. Repeat these steps for every rDNA sequence of interest, and collect all the sequences in the cart. Click on the download button, select output format options (e.g., FASTA without gaps), and start the export process. A new file will be generated and be available for download as a new job at the Download tab. Proceed to select the corresponding job and download the file.
10. This is a suggestion; other Statistical Methods for tree construction may be used, but they can demand higher computing resources without significant improvements on phylogenetic inference using 16S rDNA.

References

1. Wagner EG, Romby P (2015) Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. Adv Genet 90:133–208. https://doi.org/10.1016/bs.adgen.2015.05.001
2. Peer A, Margalit H (2014) Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. RNA 20(7):994–1003. https://doi.org/10.1261/rna.043133.113
3. Lagares A Jr, Roux I, Valverde C (2016) Phylogenetic distribution and evolutionary pattern of an alpha-proteobacterial small RNA gene that controls polyhydroxybutyrate accumulation in *Sinorhizobium meliloti*. Mol Phylogenet Evol 99:182–193. https://doi.org/10.1016/j.ympev.2016.03.026
4. Updegrove TB, Shabalina SA, Storz G (2015) How do base-pairing small RNAs evolve?

FEMS Microbiol Rev 39(3):379–391. https://doi.org/10.1093/femsre/fuv014

5. Eggenhofer F, Hofacker IL, Honer Z, Siederdissen C (2016) RNAlien - unsupervised RNA family model construction. Nucleic Acids Res 44(17):8433–8441. https://doi.org/10.1093/nar/gkw558
6. Eddy SR, Durbin R (1994) RNA sequence analysis using covariance models. Nucleic Acids Res 22(11):2079–2088. https://doi.org/10.1093/nar/22.11.2079
7. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD (2015) Rfam 12.0: updates to the RNA families database. Nucleic Acids Res 43(Database issue):D130–D137. https://doi.org/10.1093/nar/gku1063
8. Hochsmann T, Hochsmann M, Giegerich R (2006) Thermodynamic matchers: strengthening the significance of RNA folding energies. Comput Syst Bioinformatics Conf:111–121
9. Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22):2933–2935. https://doi.org/10.1093/bioinformatics/btt509
10. Smith C, Heyne S, Richter AS, Will S, Backofen R (2010) Freiburg RNA tools: a web server integrating INTARNA, EXPARNA and LOCARNA. Nucleic Acids Res 38(Web Server):W373–W377. https://doi.org/10.1093/nar/gkq316
11. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874. https://doi.org/10.1093/molbev/msw054
12. Overmars L, Kerkhoven R, Siezen RJ, Francke C (2013) MGcV: the microbial genomic context viewer for comparative genome analysis. BMC Genomics 14:209. https://doi.org/10.1186/1471-2164-14-209
13. Reese MG (2001) Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Comput Chem 26(1):51–56. https://doi.org/10.1016/S0097-8485(01)00099-7
14. Naville M, Ghuillot-Gaudeffroy A, Marchais A, Gautheret D (2011) ARNold: a web tool for the prediction of Rho-independent transcription terminators. RNA Biol 8(1):11–13
15. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28(1):33–36
16. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44(D1): D279–D285. https://doi.org/10.1093/nar/gkv1344
17. Nakhleh L (2013) Computational approaches to species phylogeny inference and gene tree reconciliation. Trends Ecol Evol 28(12): 719–728. https://doi.org/10.1016/j.tree.2013.09.004
18. Ravenhall M, Škunca N, Lassalle F, Dessimoz C (2015) Inferring Horizontal Gene Transfer. PLoS Comput Biol 11(5):e1004095. https://doi.org/10.1371/journal.pcbi.1004095
19. Woese CR (1987) Bacterial evolution. Microbiol Rev 51(2):221–271

Chapter 3

Bioinformatic Approach for Prediction of CsrA/RsmA-Regulating Small RNAs in Bacteria

Carl T. Fakhry, Kourosh Zarringhalam, and Rahul V. Kulkarni

Abstract

CsrA/RsmA is a RNA-binding protein that functions as a global regulator controlling important processes such as virulence, secondary metabolism, motility, and biofilm formation in diverse bacterial species. The activity of CsrA/RsmA is regulated by small RNAs that contain multiple binding sites for the protein. The expression of these noncoding RNAs effectively sequesters the protein and reduces free cellular levels of CsrA/RsmA. While multiple bacterial small RNAs that bind to and regulate CsrA/RsmA levels have been discovered, it is anticipated that there are several such small RNAs that remain undiscovered. To assist in the discovery of these small RNAs, we have developed a bioinformatics approach that combines sequence- and structure-based features to predict small RNA regulators of CsrA/RsmA. This approach analyzes structural motifs in the ensemble of low energy secondary structures of known small RNA regulators of CsrA/RsmA and trains a binary classifier on these features. The proposed machine learning approach leads to several testable predictions for small RNA regulators of CsrA/RsmA, thereby complementing and accelerating experimental efforts aimed at discovery of noncoding RNAs in the CsrA/RsmA pathway.

Key words CsrA, RsmA, Post-transcriptional regulation, Small RNAs, Noncoding RNA, Machine learning, Computational predictions

1 Introduction

The family of RNA-binding proteins represented by CsrA (*c*arbon *s*torage *r*egulator A)/RsmA (*r*egulator of *s*econdary *m*etabolism A) contains global post-transcriptional regulators that coordinate the transition from exponential to stationary growth phases in several bacterial species [1]. In *Escherichia coli*, CsrA plays a critical role in regulating processes related to carbon metabolism, motility as well as biofilm formation [2]. CsrA/RsmA homologs (henceforth referred to as CsrA for notational simplicity) are also known to regulate the virulence factors of animal and plant pathogens. This has been documented by several studies in bacterial species such as *Salmonella enterica* serovar *typhimurium*, *Pseudomonas aeruginosa*, *Legionella pneumophila*, and *Bacillus subtilis* [3–6]. The development of tools that enable further discovery and expansion

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_3,

of CsrA pathway regulators can thus significantly advance our knowledge about an important mechanism for global gene regulation in bacteria.

The activity of CsrA is known to be regulated by small noncoding RNAs which bind to multiple copies of the protein leading to a reduction in free CsrA levels in the cell [7]. However, identification of CsrA-regulating small RNAs in bacterial genomes is a challenging enterprise, both experimentally and computationally. Many of the small RNAs that interact with CsrA have low degrees of sequence conservation. The most striking conserved feature is the presence of multiple binding sites for CsrA. For most of the CsrA-regulating small RNAs discovered so far, it has been shown that their transcription is induced by the GacS/A two-component system [8–11], which is present only in the Gram-negative gamma-proteobacteria. The promoters of these small RNAs have a short upstream activating sequence that has been identified as a GacA-binding site. This feature, combined with the presence of Rho-independent transcription terminators at the end of the small RNA genes and the presence of multiple CsrA-binding motifs, allowed the design of a sequence-based algorithm that predicts these small RNAs in gamma-proteobacteria in our previous work [12]. That work demonstrated that computational searches based on locating intergenic regions with high frequencies of the CsrA-binding motif (ANGGA/AGGA) can lead to the identification of experimentally known CsrA-regulating noncoding small RNAs. Furthermore, this approach also led to predictions for several previously undiscovered CsrA-regulating small RNAs, some of which were confirmed by subsequent studies, e.g., by experiments in *L. pneumophila* [13–15]. Subsequently, computational approaches for identification of RNA families have led to the prediction of CsrA-regulating small RNAs for several species in the gamma-proteobacteria as documented at RFAM [16].

It is noteworthy that there are numerous species that encode highly conserved RsmA homologs which do not have orthologs of the GacS/A two-component system found in the gamma-proteobacteria. Assuming that the mode of action of the CsrA pathway in such species is similar to that of the gamma-proteobacteria, this observation suggests the existence of so-far undiscovered CsrA-regulating small RNAs in these species. Furthermore, the transcription of such small RNAs is likely to be controlled by other signal transduction systems, given the absence of the GacS/A two-component system. It is therefore conceivable that even in the gamma-proteobacteria there are additional, Gac-independent small RNAs that bind to and regulate cellular levels of CsrA. Thus, there is a need for bioinformatics approaches which lead to novel predictions of CsrA-regulating small RNAs in bacteria.

2 Overview of Approach

An important element, missing in our previous bioinformatics approach, is the role of RNA secondary structure in determination of potential CsrA-binding small RNAs. Given that several CsrA-regulating small RNAs have now been documented, machine learning approaches can be used to identify patterns in both sequence and structure that define small RNA regulators of CsrA. In recent work, we have developed such a machine learning approach which is described in the following [17].

Our approach involves a protocol for training a machine learning classifier for prediction of CsrA-regulating small RNAs. Training of the classifier relies on the fact that functional RNA classes typically have similar structural or sequential features. For instance, in bacterial sRNAs, specific sequential-structural motifs (such as the presence of a Rho-independent terminator at the 3′ end) have a higher probability of appearing in the ensemble of structural conformations (which correspond to low free energy structures) obtained from RNA folding algorithms. We identify similar sequence-structure based signatures and train a binary classifier for the class of CsrA-regulating small RNAs.

As in any supervised binary classifier, our method requires a set of positive and a set of negative examples for training. In supervised learning, labeled positive and negative training examples (such as CsrA-regulating or not CsrA-regulating) are utilized to identify patterns that can separate the positive from negative classes. These patterns are typically transformations of certain features that are calculated from the training examples. Once optimal model parameters are identified, the "fitted" model can be utilized to classify new unlabeled sequences.

For the purpose of classifying CsrA-regulating small RNAs, we utilize previously known CsrA-regulating small RNAs as the positive set. While it is usually difficult to define an appropriate negative set, i.e., small RNAs that do not regulate CsrA, we devise a method for constructing such sequences from the positive examples. Specifically, for each positive sequence, we shuffle the sequence, while preserving the dinucleotide frequencies [18]. Next we examine the minimum free energy structure of the shuffled sequence to ensure that the structure is within similar range of free energy as the positive examples. This is to ensure that sequences are not structurally very different from known small RNAs.

Once positive and negative sequences are defined, the next task is to calculate a set of features that can potentially distinguish the two classes. We calculate a set of 512 features by examining the ensemble of low free energy structures as follows. In the ensemble of low free energy structures, nucleotides are either paired or unpaired. Let us represent the pairing status of the nucleotides with

binary numbers 0 and 1, where 0 indicates unpaired and 1 indicates paired. To account for structural motifs, we consider the pairing status of nucleotide triplets. There are 64 possible triplets consisting of the sequence combinations AAA, AAC, …, UUU. Each of these triplets can be assigned to one of eight possible pairing conformations, i.e., 000, 001, …, 111. For instance, (GGA, 000) indicates the presence of unpaired triplet GGA in the secondary structures (*see* **Note 1**). Taken together, these sequence-structure combinations represent 512 features; (AAA, 000), …, (UUU, 111). We refer to these features as Boltzmann Triplet Features (BTF). Figure 1 shows an example. The probability of BTFs for a given RNA sequence can be calculated by taking a stochastic sample of low free energy structures for the RNA sequence and then computing the frequency of each feature in the ensemble.

We also consider two additional features, namely, (1) the probability of formation of a stem-loop at the 3′ end of the sequence, and (2) presence of a Rho-independent terminator as defined in [19]. These features are computed by examining the ensemble of secondary structures in a similar manner.

The next task in classification is training a binary classifier. There are several choices of supervised classification methods that can be utilized for this task. Some examples include Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), Random Forests, and L_1 regularized logistic regression (LASSO). A particular advantage of LASSO is its high interpretability of the output results. More specifically, in addition to classifying examples as CsrA-regulating or not CsrA-regulating, LASSO will automatically identify the most predictive features that can separate the two classes. For this reason, we utilized LASSO for binary classification (*see* **Note 2**). This model was fitted on training data to identify optimal model parameters. The fitted model was then applied to classify new sequences. To decrease variance in model selections and predictions, we used (1) Ensemble learning on multiple training data, and (2) Bootstrap analysis to identify robust features. Ensemble learning is a machine learning model where several classifiers are trained and predictions are made using each classifier. The final class labels are typically decided by a majority vote or a weighted average (*see* **Note 3**). On the other hand, Bootstrap is a resampling technique that is typically utilized to estimate measures of accuracy. We utilized bootstrapping to assess the robustness of the predictors in classifying the features (*see* **Note 4**). Finally, we generated an initial input list of potential CsrA-regulating small RNAs in a given bacterial genome and the trained model was used to classify each sequence by determining the probability (given the binary classifier) that it belongs to the class of CsrA-regulating small RNAs. Details of the procedures are presented in Methods and Notes.

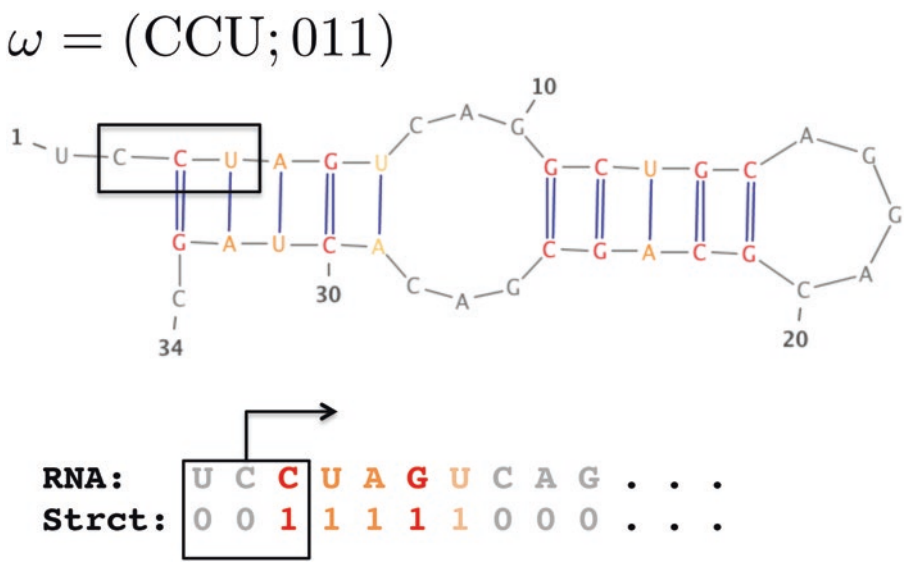


Fig. 1 Boltzmann triplet feature (BTF) example using a randomly selected structure from the ensemble of low free energy secondary structures

3 Methods

The required tools for performing the above procedure are

1. RNAfold program from the Vienna Package [20].
2. RNAsubopt program from the Vienna Package.
3. R programming language.
4. glmnet library for regularized logistic regression in R.
5. tools for obtaining intergenic sequences in bacterial genomes (such as the tools for sequence analysis at http://www.rsat.eu/ [21]).

The procedure for computing the features and learning the model for CsrA-regulating small RNAs is as follows. All algorithmic procedures are implemented in an R package freely available to download at: https://github.com/carltonyfakhry/InvenireSRNA. We also provide a webserver for performing predictions at: http://markov.math.umb.edu/inveniresrna. *See* Subheading 3.8 for further details.

3.1 Positive Set

1. Download the seed sequences of previously known CsrA-regulating small RNAs from RFAM [16] in (Ungapped) FASTA format. This includes sequences from the RsmX, RsmY, RsmZ, CsrB, and CsrA families of noncoding RNAs (ncRNAs). A total of 105 seed sequences were available at RFAM at the time of this study. These constitute the positive examples.

3.2 Negative Set

1. Using RNAfold compute the minimum free energy (MFE) structure of all positive sequences to obtain the range of minimum free energies.
2. Shuffle each positive sequence while keeping the dinucleotide frequencies fixed using the Altschuld-Erickson algorithm [18]. This algorithm is implemented as part of the R package.
3. Using RNAfold compute the minimum free energy (MFE) structure of the shuffled sequence. If the free energy is within the acceptable range, accept the sequence. Else repeat **steps 2** and **3** for a maximum of 100 tries before moving to the next sequence.

3.3 Ensemble Training Sets

1. Construct 100 different negative sets using procedures in Subheading 3.2. Note that random shuffles will result in different negative sets.
2. For each random shuffle, construct a training set consisting of all positives and the newly shuffled negative sequences.

3.4 Feature Generation

1. For each sequence in the training set, sample 1000 structures from the ensemble of low energy secondary structures using the *RNAsubopt* program from the *Vienna* package.
2. Compute the frequencies of the BTFs over all the sampled structures obtained from **step 2**.
3. The empirical probabilities of the possible BTFs are the first 512 features. If a BTF triplet does not occur in any of the sampled structures, its probability is set to 0. Otherwise its probability is the frequency of BTF in the ensemble divided by the sum of frequencies of all BTFs.
4. The 513th feature is the probability of the formation of a stem-loop at the 3′ end of the sequence. This is computed by examining the occurrences of a stem-loop in the random structures obtained from **step 2**. This requires a regular expression to check the formation of a double helix that ends in an unpaired loop at the end of each random structure. The probability for the formation of the stem-loop is thus the number of structures that have a stem-loop divided by 1000 (the number of sampled structures obtained from **step 2**).
5. The 514th feature is a categorical variable indicating whether the RNA has a Rho-independent terminator. The Rho-independent terminator is defined in a similar manner to [19]. For the sequence to have a Rho-independent terminator the last 13 nucleotides must contain at least 3 U nucleotides and no UVVUU stretches (V is an A, C, or G) and the last 7 nucleotides must contain at least 1 U nucleotide.

3.5 Bootstrapping and Robust Feature Identification

1. Using the *R* package *glmnet*, perform a cross validation for LASSO regression for a given training dataset and save the tuning parameter which minimizes the mean-squared error. We refer to the tuning parameter as *lambda.min* (*see* **Note 5**).
2. Take a bootstrap sample from the training data and estimate a LASSO regression using *lambda.min*.
3. Identify significant predictors as those with nonzero coefficient in the regression model.
4. Repeat **step 2.**
5. Re-estimate a LASSO regression using *lambda.min* and note all the coefficients that were not set to 0 in any of the **steps 2–4**.
6. The nonzero coefficients in **step 5** identify the robust features for the model.

3.6 Generation of Input Sets for Prediction

1. Obtain intergenic sequences for the bacterial species of interest using the sequence analysis tools available at http://www.rsat.eu/. These are obtained by downloading the regions that are upstream/downstream of annotated genes avoiding overlap with flanking genes on the same strand. Note that some part of what we call the intergenic sequence can be antisense to a coding sequence on the opposite strand.
2. To initially screen for regions which have potential CsrA-binding small RNAs, identify the regions (from **step 1**) that have 3 or more ANGGA binding motifs, (with at least 3 motifs that are within a distance of 60 bp) followed by a poly T tail (minimum of 3 Ts).
3. The poly T tail (from **step 2**) defines the 3′ end of the input sequence. The 5′ end is determined by sampling from a distance distribution of distances of 5′ ends from the first ANGGA motif in experimentally known CsrA-regulating small RNAs. This approach generates the input sequences for prediction from the regions acquired in **step 1**.
4. To further limit the input set, the sequences obtained in the preceding steps were screened for presence of Rho-independent terminator sequences using the Web tool Arnold (http://rna.igmors.u-psud.fr/toolbox/arnold/) [22].
5. The input sequences thus generated were classified using the trained binary classifier using an ensemble of models as described below.

3.7 Ensemble of Models and Predictions

1. Train 100 models using the *glmnet* package, one per Ensemble training samples, each trained on robust features only (*see* **Note 6**).
2. Given a new sequence, generate the BTF features.

3. Using the features generated in **step 2**, perform 100 predictions using the 100 models learned in **step 1**.
4. The final prediction is the average of all predictions (probabilities) outputted by the 100 models learned in **step 1**.

3.8 InvenireSRNA R Package and Webserver

1. We provide an R package *InvenireSRNA* that implements the outlined algorithm in Subheadings 3.3, 3.4, 3.5, and 3.6.
2. The R package provides functionality to predict the probability of a given RNA sequence (or sequences provided in a FASTA file), belonging to a certain class of small RNAs. The default model used for prediction is for CsrA-regulating small RNAs.
3. The R package is currently downloadable from https://github.com/carltonyfakhry/InvenireSRNA.
4. We also provide the *InvenireSRNA* web server available at http://markov.math.umb.edu/inveniresrna. The web server provides similar functionality as the R package and is intended to increase the ease of use of our method. There is an upper limit of 500 sequences that can be uploaded in a FASTA file for prediction.

4 Notes

1. Based on experiments, a GGA motif in unpaired regions in the RNA secondary structure is considered a key distinguishing feature of the CsrA-binding site [23].
2. Our choice of the regularized logistic regression was motivated by its ability to perform automatic feature selection and the probabilistic interpretation of its output. However, there is a great variety in the choice of classifiers that could have been used as a substitute to LASSO such as Support Vector Machines and Ridge regression.
3. Learning an ensemble of models helps in decreasing variance due to variability in the data set. High model variance can arise from variability in training examples. Training examples are typically assumed to be a fair representation of the entire population. Larger training samples will typically result in lower model variance. In order to account for variability in the negative set, we opted to generate an ensemble of negative sets and train a model on each. Each model is utilized independently to make predictions and the final class labels are decided by a majority vote.
4. Identifying robust features with Bootstrap helps in reducing dimension and variance. More precisely, LASSO will identify a small set of predictive features that can distinguish the classes

using cross-validation analysis. There is some variability in this process that is if the model is retrained using a slightly different training set, different predictive features may be selected by lasso. To account for this source of variability, for each training set from the ensemble, we sample the set with replacement (bootstrapping), train a classifier on the bootstrapped sample, identify the predictive feature, and repeat the process 100 times. We then keep track of the frequency of the selected features by LASSO in the 100 iterations. The features that are consistently selected constitute the robust features. Consistency of features is decided by examining the distribution of the frequencies (e.g., taking the upper 75% quantile).

5. When fitting an L_1 regularized logistic regression (LASSO), a penalty parameter lambda must be selected. This parameter controls model complexity in terms of the number of features that are entered into the model. The optimal value of the lambda parameter is obtained by cross validation, where for a grid of lambda values the model is trained and tested and loss function is recorded. The optimal parameter value called lambda.min is selected as the one that results in minimum loss.
6. As in any regression model, the performance of our method is dependent on the estimated features. While we believe there is strong biological intuition and evidence to support the use of such features, there could potentially be other features, which are more significant in terms of performance.

Acknowledgments

This work was supported by the NCI-funded U54 UMass Boston-Dana Farber/Harvard Cancer Center Partnership Grant [CA156734].

References

1. Vakulskas CA, Potts AH, Babitzke P, Ahmer BM, Romeo T (2015) Regulation of bacterial virulence by Csr (Rsm) systems. Microbiol Mol Biol Rev 79(2):193–224. https://doi.org/10.1128/mmbr.00052-14
2. Timmermans J, Van Melderen L (2010) Post-transcriptional global regulation by CsrA in bacteria. Cell Mol Life Sci 67(17):2897–2908. https://doi.org/10.1007/s00018-010-0381-z
3. Burrowes E, Baysse C, Adams C, O'Gara F (2006) Influence of the regulatory protein RsmA on cellular functions in Pseudomonas aeruginosa PAO1, as revealed by transcriptome analysis. Microbiology 152(Pt 2):405–418. https://doi.org/10.1099/mic.0.28324-0
4. Lawhon SD, Frye JG, Suyemoto M, Porwollik S, McClelland M, Altier C (2003) Global regulation by CsrA in Salmonella typhimurium. Mol Microbiol 48(6):1633–1645
5. Molofsky AB, Swanson MS (2003) Legionella pneumophila CsrA is a pivotal repressor of transmission traits and activator of replication. Mol Microbiol 50(2):445–461
6. Yakhnin H, Pandit P, Petty TJ, Baker CS, Romeo T, Babitzke P (2007) CsrA of Bacillus subtilis regulates translation initiation of the gene encoding the flagellin protein (hag) by blocking ribosome binding. Mol Microbiol 64(6):1605–1620. https://doi.org/10.1111/j.1365-2958.2007.05765.x

7. Babitzke P, Romeo T (2007) CsrB sRNA family: sequestration of RNA-binding regulatory proteins. Curr Opin Microbiol 10(2):156–163. https://doi.org/10.1016/j.mib.2007.03.007
8. Heeb S, Haas D (2001) Regulatory roles of the GacS/GacA two-component system in plant-associated and other gram-negative bacteria. Mol Plant Microbe Interact 14(12):1351–1363. https://doi.org/10.1094/mpmi.2001.14.12.1351
9. Valverde C, Heeb S, Keel C, Haas D (2003) RsmY, a small regulatory RNA, is required in concert with RsmZ for GacA-dependent expression of biocontrol traits in Pseudomonas fluorescens CHA0. Mol Microbiol 50(4):1361–1379
10. Lapouge K, Schubert M, Allain FH, Haas D (2008) Gac/Rsm signal transduction pathway of gamma-proteobacteria: from RNA recognition to regulation of social behaviour. Mol Microbiol 67(2):241–253. https://doi.org/10.1111/j.1365-2958.2007.06042.x
11. Brencic A, McFarland KA, McManus HR, Castang S, Mogno I, Dove SL, Lory S (2009) The GacS/GacA signal transduction system of Pseudomonas aeruginosa acts exclusively through its control over the transcription of the RsmY and RsmZ regulatory small RNAs. Mol Microbiol 73(3):434–445. https://doi.org/10.1111/j.1365-2958.2009.06782.x
12. Kulkarni PR, Cui X, Williams JW, Stevens AM, Kulkarni RV (2006) Prediction of CsrA-regulating small RNAs in bacteria and their experimental verification in Vibrio fischeri. Nucleic Acids Res 34(11):3361–3369. https://doi.org/10.1093/nar/gkl439
13. Edwards RL, Jules M, Sahr T, Buchrieser C, Swanson MS (2010) The Legionella pneumophila LetA/LetS two-component system exhibits rheostat-like behavior. Infect Immun 78(6):2571–2583. https://doi.org/10.1128/iai.01107-09
14. Hovel-Miner G, Pampou S, Faucher SP, Clarke M, Morozova I, Morozov P, Russo JJ, Shuman HA, Kalachikov S (2009) SigmaS controls multiple pathways associated with intracellular multiplication of Legionella pneumophila. J Bacteriol 191(8):2461–2473. https://doi.org/10.1128/jb.01578-08
15. Sahr T, Bruggemann H, Jules M, Lomma M, Albert-Weissenberger C, Cazalet C, Buchrieser C (2009) Two small ncRNAs jointly govern virulence and transmission in Legionella pneumophila. Mol Microbiol 72(3):741–762. https://doi.org/10.1111/j.1365-2958.2009.06677.x
16. Daub J, Eberhardt RY, Tate JG, Burge SW (2015) Rfam: annotating families of non-coding RNA sequences. Methods Mol Biol 1269:349–363. https://doi.org/10.1007/978-1-4939-2291-8_22
17. Fakhry CT, Kulkarni PR, Chen P, Kulkarni RV, Zarringhalam K (2017) Prediction of bacterial small RNAs in the RsmA (CsrA) and ToxT pathways: a machine learning approach. BMC Genomics 18:645
18. Altschul SF, Erickson BW (1985) Significance of nucleotide sequence alignments: a method for random sequence permutation that preserves dinucleotide and codon usage. Mol Biol Evol 2(6):526–538
19. Lesnik EA, Sampath R, Levene HB, Henderson TJ, McNeil JA, Ecker DJ (2001) Prediction of rho-independent transcriptional terminators in Escherichia coli. Nucleic Acids Res 29(17):3583–3594
20. Lorenz R, Bernhart SH, Honer Z, Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL (2011) ViennaRNA Package 2.0. Algorithms Mol Biol 6:26. https://doi.org/10.1186/1748-7188-6-26
21. Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, Thieffry D, van Helden J (2011) RSAT 2011: regulatory sequence analysis tools. Nucleic Acids Res 39(Web Server issue):W86–W91. https://doi.org/10.1093/nar/gkr377
22. Naville M, Ghuillot-Gaudeffroy A, Marchais A, Gautheret D (2011) ARNold: a web tool for the prediction of Rho-independent transcription terminators. RNA Biol 8(1):11–13
23. Valverde C, Lindell M, Wagner EG, Haas D (2004) A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of Pseudomonas fluorescens. J Biol Chem 279(24):25066–25074. https://doi.org/10.1074/jbc.M401870200

Part II

Function and Differential Expression of sRNA

Chapter 4

Host-Pathogen Transcriptomics by Dual RNA-Seq

Alexander J. Westermann and Jörg Vogel

Abstract

Transcriptomics, i.e., the quantification of cellular RNA transcripts, is a powerful way to gauge the physiological state of either bacterial or eukaryotic cells under a given condition. However, traditional approaches were unsuitable to measure the abundance of transcripts across kingdoms, which is relevant for biological processes such as bacterial infections of mammalian host cells. This changed with the establishment of "Dual RNA-seq," which profiles gene expression simultaneously in an infecting bacterium and its infected host. Here, we describe a detailed Dual RNA-seq protocol optimized for—but not restricted to—the study of human cell culture models infected with the Gram-negative model pathogen *Salmonella* Typhimurium. Furthermore, we provide experimental data demonstrating the benefits of some of the key steps of this protocol, including transcriptome stabilization (RNA fixation), FACS-based enrichment of invaded cells, and double rRNA depletion. While our focus is on data generation, we also include a section describing suitable computational methods to analyze the obtained datasets.

Key words Dual RNA-seq, Host-pathogen interaction, Infection, Transcriptomics, RNA-seq, *Salmonella*, Noncoding RNA, Cell sorting, Fixation, rRNA depletion

1 Introduction

To comprehensively describe bacterial infections of mammalian host cells, the underlying gene expression changes in host and pathogen need to be understood. Host-pathogen transcriptomics has traditionally relied on the physical separation of the two infection partners post-infection, followed by their separate analysis via probe- or sequencing-based technologies. In contrast, Dual RNA-seq omits physical separation of host and pathogen cells, with transcripts of the two organisms being discriminated in silico, by assigning sequencing reads to their originating reference genome (reviewed in [1, 2]). To this end, infected cells and infecting bacteria associated with them are lysed jointly, resulting in the release of total RNA. Next, this mixture of eukaryotic and prokaryotic RNA species is purified, ribosomal RNA (rRNA) depleted, the rRNA-free sample converted into cDNA libraries, and sequenced (Fig. 1a). The single-nucleotide resolution of RNA-seq technology

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_4,

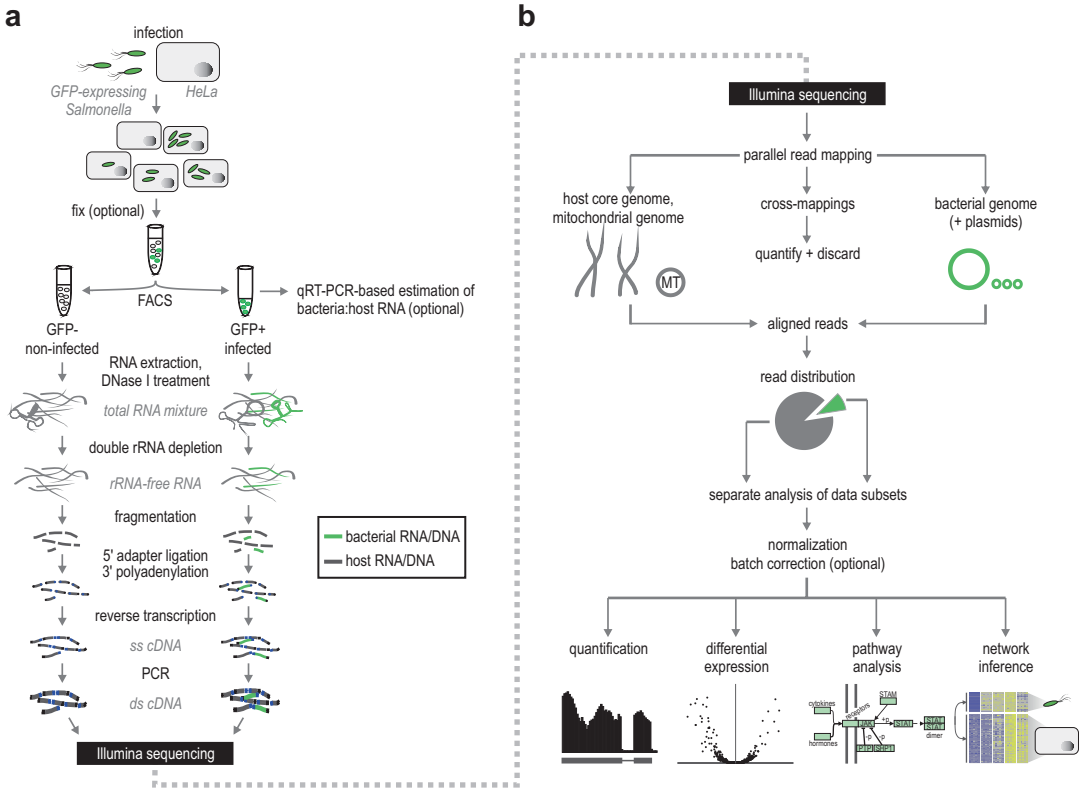


Fig. 1 Experimental workflow for the Dual RNA-seq approach here described. (**a**) Data generation. (**b**) Data analysis. Panel b was modified from [2]

is harnessed to segregate reads derived from host or pathogen with high confidence, followed by the separate quantification of the two resulting data subsets (Fig. 1b).

In the following, we present a detailed Dual RNA-seq protocol for in vitro infection of human cell lines with the facultative intracellular pathogen *Salmonella enterica* serovar Typhimurium (henceforth *Salmonella*). This approach is capable of detecting all major bacterial and human RNA classes, and its application to samples obtained from time course infection experiments with diverse mammalian cell types has recently enabled us to discover a previously uncharacterized small noncoding RNA of *Salmonella* that functions as a timer of expression of this pathogen's major virulence programs, with widespread consequences for the infected host [3].

2 Materials

2.1 Bacterial Cultures

1. *Salmonella enterica* serovar Typhimurium strain SL1344 constitutively expressing GFP from a chromosomal locus (strain JVS-3858) [4].

2. 250 mL Erlenmeyer flasks.
3. Petri dishes.
4. Incubator for bacterial plates.
5. Lennox Broth (LB) liquid medium: 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 85.6 mM sodium chloride; Lennox agar: LB medium (*see* above), 1.2% (w/v) agar.

2.2 Cell Culture

1. HeLa-S3 cells (ATCC CCL-2.2).
2. Class II biological safety cell culture hood.
3. T-75 flasks.
4. 6-well plates.
5. Serological pipets 5, 10, 25 mL (plastic).
6. Incubator for cell culture flasks.
7. Light microscope.
8. Benchtop centrifuge.
9. Neubauer counting chamber.
10. DMEM complete: Dulbecco's Modified Eagle Medium (DMEM), 10% fetal calf serum, 2 mM L-glutamine, 1 mM sodium pyruvate.
11. Phosphate-buffered saline (PBS).
12. Trypsin-EDTA (0.25%), phenol red.

2.3 Infection Assays

1. Benchtop centrifuge.
2. Vortex.
3. Spectrophotometer or cell density meter, and cuvettes.
4. DMEM complete + gentamicin (high): DMEM complete (*see* above), gentamicin sulfate salt to a final concentration of 50 μg/mL.
5. DMEM complete + gentamicin (low): DMEM complete (*see* above), gentamicin sulfate salt to a final concentration of 10 μg/mL.
6. Vacuum system for liquid aspiration and disposal.

2.4 Transcriptome Stabilization and Cell Sorting

1. RNA*later* (Qiagen).
2. PBS.
3. FACSAria III (BD Biosciences).

2.5 RNA Extraction, DNase I Treatment, qRT-PCR, rRNA Depletion

1. *mir*Vana miRNA isolation kit (Life technologies).
2. Safe-lock tubes 1.5, 2 mL.
3. Benchtop microcentrifuge (refrigerated).
4. Ethanol diluted to 70% with RNase-free water.

5. RNase-free water.
6. Heat block for microtubes.
7. Spectrophotometer NanoDrop 2000, or equivalent.
8. Deoxyribonuclease I (DNase I, 1 U/μL; Fermentas).
9. SUPERaseIN RNase Inhibitor (Ambion).
10. Phase-lock gel (PLG) tubes 2 mL (5 PRIME).
11. Roti-Aqua P/C/I (Roth).
12. GlycoBlue (Ambion).
13. 30:1 ethanol/sodium acetate: 30 parts of 100% ethanol, 1 part of 3 M sodium acetate (pH 6.5).
14. *Power* SYBR Green RNA-to-CT 1-Step kit (Life technologies).
15. Gene-specific DNA oligonucleotides.
16. TC Microwell 96F (Thermo Scientific).
17. Real-time PCR cycler CFX96 Real-Time System (BioRad).
18. Ribo-Zero Gold epidemiology kit (Illumina).
19. DynaMag-2 Magnetic Particle Concentrator (Invitrogen).
20. Ethanol 100%.

2.6 cDNA Synthesis and Illumina Sequencing

1. MultiNA microchip electrophoresis system (Shimadzu).
2. M220 Focused-ultrasonicator (Covaris).
3. Agencourt RNAClean XP kit (Beckman Coulter Genomics).
4. Antarctic Phosphatase (NEB).
5. T4 Polynucleotide Kinase (NEB).
6. Poly(A) polymerase (NEB).
7. RNA adaptor.
8. T4 RNA ligase (NEB).
9. M-MuLV reverse transcriptase (NEB).
10. PCR primers.
11. Phusion high fidelity DNA polymerase (NEB).
12. HiSeq2500 (Illumina).

3 Methods

3.1 Scaling the Experiment

1. For each condition, ≥200 ng of total RNA is required prior to Ribo-Zero treatment. Based on our experience, this corresponds to ~2×10^5 sorted HeLa cells. Cells are seeded in a way that, at the time of infection, there will be ~8×10^5 cells per well (6-well format) and thus ~5×10^6 cells per plate. Provided

that ~5% of the HeLa cells get invaded when infection is carried out at a multiplicity of infection (m.o.i.) of 5 [3], as a result of infection there should be ~2.5×10^5 invaded cells on every 6-well plate. Thus, at least one full 6-well plate is seeded for each condition to be sampled (including one plate for the mock-infected controls; *see* **step 1** in Subheading 3.3).

2. In addition, a single well (8×10^5 cells) is required as carrier material for the bacterial reference control (*see* **step 2** in Subheading 3.3).

3.2 Salmonella Infection Assay of Human Epithelial Cells

1. Two days prior to infection, HeLa cells are trypsinized, cell density is determined in a Neubauer counting chamber, and 2×10^5 cells are seeded in 2 mL complete DMEM per each well of a 6-well plate (one plate/condition).
2. Efficient seeding and attachment of the cells to the substrate is confirmed by light microscopy on the subsequent day.
3. For infection, overnight LB cultures of *Salmonella* constitutively expressing the green fluorescent protein (GFP) from the *put* locus in the chromosome [4] are diluted 1:100 in fresh LB medium and grown aerobically at 37 °C (shaking at 220 rpm) to an OD_{600} of 2.0.
4. A volume of 1 mL of the bacterial culture is transferred to a 2 mL reaction tube, pelleted (2 min at 12,000 rpm, room temperature) and bacteria are resuspended in 1 mL of complete DMEM medium and diluted to 8×10^4 bacterial cells/μL (using DMEM). Double-check for correct concentration of the inoculum solution by spectrometry (ideally, the OD_{600}-value measured should be 0.08).
5. To achieve an m.o.i. of 5, 50 μL of this suspension is pipetted into each well. Immediately after addition of the bacterial inoculum, the plates are centrifuged for 10 min at 250 × *g*, room temperature, to enhance pathogen contact with host cells and synchronize the invasion event.
6. For infection to occur, the plates are incubated for 30 min in a 5% CO_2, humidified atmosphere, at 37 °C.
7. Thereafter, medium is replaced for gentamicin-containing complete DMEM (final gentamicin concentration 50 μg/mL) to kill the remaining extracellular bacteria.
8. After a further 30 min incubation step as above, the medium is replaced by complete DMEM containing only 10 μg/mL of gentamicin and incubated for the remainder of the experiment. Note: Time point t_0 is defined as the time when gentamicin is first added to the cells (**step 7** in Subheading 3.2).

3.3 Reference Controls

1. As a reference for host gene expression changes, a non-infected yet mock-treated control needs to be included for every time point to be sampled.
2. The *Salmonella* reference control represents bacteria from the inoculum suspension. To minimize potential biases during RNA isolation using the *mir*Vana kit (*see* **step 3** in Subheading 3.7; optimized for mammalian samples), the bacterial lysate is artificially supplemented with host carrier material prior to joint RNA extraction. In order to obtain a similar ratio of bacterial to human transcripts in the bacterial control compared to the infection samples, 50 μL of the *Salmonella* inoculum suspension is added to the lysate of ~8×10^5 HeLa cells (i.e., one well) in 600 μL of L/B buffer (*see* **Note 1**). From the resulting mixture total RNA is isolated (*see* Subheading 3.7).

3.4 Harvest Infected HeLa Cells

1. For harvest, infected cells are washed with room temperature PBS, trypsinized (100 μL of pre-warmed trypsin/well; stop with each 900 μL of complete DMEM), and collected in 15 mL tubes (pool wells of one plate).
2. If the transcriptome of the infected cells is not meant to be stabilized (e.g., when you carry on directly with cell sorting in Subheading 3.6; *see also* **Note 2**), the harvested cells are pelleted by centrifugation (5 min at 250 × *g*, 4 °C), the supernatant is aspirated, and the cell pellet resuspended in 1 mL of ice-cold PBS (cell density ~5×10^6 cells/mL). *Directly proceed with* Subheading 3.6.

3.5 OPTIONAL: Transcriptome Stabilization

1. When samples from multiple strains are collected in parallel and/or in case that the cells cannot immediately be further processed, their transcriptomes should be preserved during storage. To stabilize the transcriptome of infected cells prior to sorting, cell samples from **step 1** in Subheading 3.4 are pelleted (5 min at 250 × *g*, 4 °C), resuspended in 1 mL of RNA*later* (1 mL/5×10^6 cells; Qiagen), and stored at 4 °C until sorting. We use RNA*later*, because this reagent—despite being initially developed for mammalian cells—also fixes intracellular *Salmonella*. That is, storage of infected HeLa cells in RNA*later* lowers the number of viable bacteria by almost two orders of magnitude (Fig. 2a). On the other hand, the treatment reduces the final RNA yield by approximately two thirds compared to unfixed cells (Fig. 2b), probably due to partial cell lysis or cell loss (i.e., cells are difficult to pellet in the viscous RNA*later* reagent). Importantly, RNA*later* stabilizes the here selected pathogen and host transcripts (Fig. 2c, d). *See also* **Note 2**.
2. To prepare RNA*later*-treated cell samples for sorting (*see* Subheading 3.6), the cell suspension is diluted by the addition

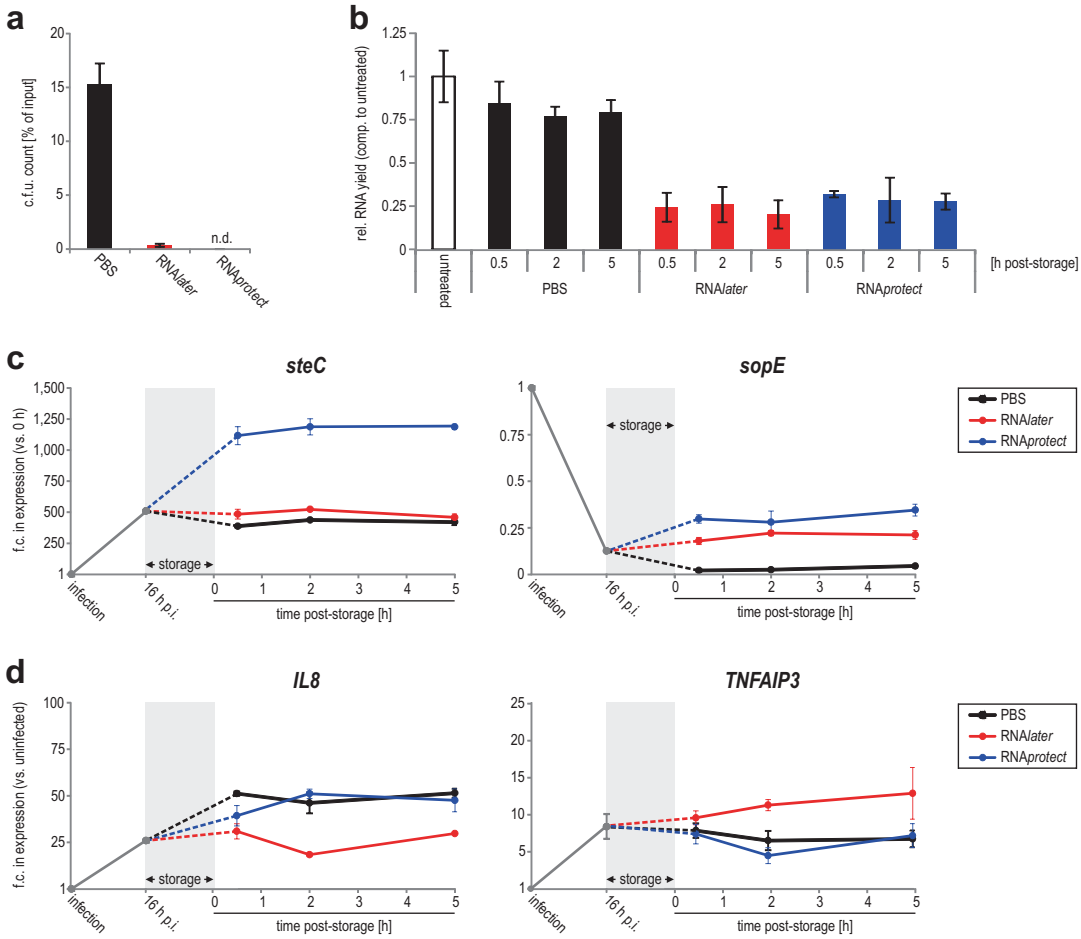


Fig. 2 Evaluation of transcriptome stabilizing reagents. Viable c.f.u. counts in lysates (**a**) and total RNA yield (**b**) of infected HeLa cells (m.o.i. 30; 16 h p.i.) without (PBS) or upon their fixation with RNA*later* or RNA*protect*, as indicated. (**c**, **d**) Infected HeLa cells (m.o.i. 30) were harvested 16 h p.i. and differentially fixed as indicated. Upon storage of the samples overnight at 4 °C, the preservative was removed and cells resuspended and incubated in PBS on ice for the indicated time points ("time post-storage") prior to RNA isolation and qRT-PCR analysis. *steC* and *sopE* are mRNAs of *Salmonella*, and *TNFAIP3* and *IL8* human transcripts. Constitutively expressed *gfp* mRNA served as the *Salmonella*, and U6 snRNA as the host reference. Data in panels a-d represent the mean ± SD from technical triplicates of each single biological sample

of 10 mL of ice-cold PBS (as the RNA*later* reagent is very viscous and would otherwise impede pelleting) and cells are pelleted for 5 min at 500 × *g*, 4 °C.

3. The supernatant is carefully removed and the cell pellet resuspended in 500 µL of ice-cold PBS. Note that this is half the volume as for unfixed cells (*see* **step 2** in Subheading 3.4) to account for the described cell loss during RNA*later* treatment (**step 1** in Subheading 3.5). *Continue with* **step 1** in Subheading 3.6.

3.6 Cell Sorting

1. HeLa cells containing GFP-expressing *Salmonella* emit a fluorescence signal and so can be distinguished from uninfected bystander cells. The FACS gating strategy to separate invaded (GFP-positive) from bystander cells (GFP-negative) is outlined in Fig. 3. GFP intensity is measured in the FITC channel and PE intensity serves as a measure of a cell's auto-fluorescence. Doublet discrimination is difficult to achieve because cell size substantially depends on the number of intracellular bacteria and thus is typically omitted.
2. Sorting is performed using the 100 micron nozzle, at medium flow rate (<7,000 events/s) and under constant cooling to 4 °C (of both the input chamber and the collection tube holder) into 2 mL reaction tubes. Typically ~2 × 10^5 cells are collected per each fraction and applied to RNA isolation.

3.7 Cell Lysis, RNA Extraction, DNase I Treatment

1. The sorted cells are pelleted for 5 min at 1,000 × *g*, 4 °C, and the supernatant is quickly aspirated with care (at best by using a vacuum pump). Note that the cell pellet will be almost invisible.

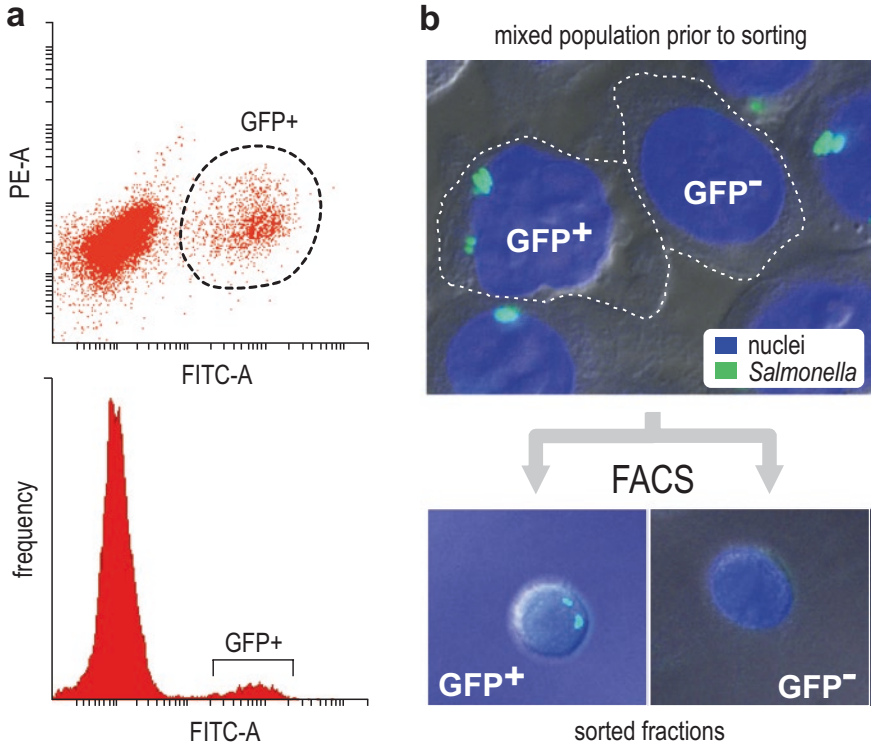


Fig. 3 FACS outline. (**a**) Gating strategy. HeLa cells invaded with GFP-expressing *Salmonella* are distinguished from uninfected bystander cells based on an increased signal in the FITC channel. In the PE channel the auto-fluorescence of the cells is detected. (**b**) Representative fluorescence microscopy images of *Salmonella*-infected HeLa cells before (upper) and after sorting (lower)

2. Lysis is achieved by adding 600 µL of L/B buffer (*see* **Note 1**) to the cell pellet. To efficiently lyse also the intracellular bacteria the sample should be vortexed vigorously.
3. Host-pathogen RNA is extracted using the *mir*Vana kit (Life technologies) following the manufacturer's instructions for total RNA isolation.
4. To remove contaminating genomic DNA remnants, RNA samples are treated with 0.025 U of DNase I per 100 ng of RNA for 45 min at 37 °C in a total volume of 50 µL.
5. The gDNA-free RNA is next purified by ethanol precipitation. To this end, three volumes (150 µL) of Roti-Aqua P/C/I (Roth) are added to each reaction. The samples are mixed by pipetting and transferred into PLG tubes. Upon vortexing for 15 s, the tubes are centrifuged for 15 min at 12,000 rpm, 15 °C. The aqueous phases are transferred into fresh 1.5 mL reaction tubes, supplemented with 1.5 µL of GlycoBlue (Ambion) and three volumes (150 µL) of a 30:1 mixture of ethanol and sodium acetate, and precipitation is achieved by incubating the samples overnight at −20 °C.
6. The next morning, the RNA is pelleted by centrifugation for 30 min at 12,000 rpm and 4 °C. The supernatant is removed from each sample and the pellets are washed with 200 µL of 75% ice-cold ethanol. Upon centrifugation for 10 min as above, the supernatants are removed, the pellets air-dried for 10 min at room temperature with open lids, and resuspended in 28 µL of 65 °C-warm, RNase-free water for 5 min at 65 °C.

3.8 OPTIONAL: Estimation of Bacteria-to-Host RNA Ratio

1. When establishing a Dual RNA-seq approach for a given infection model, it might be reasonable to determine the relative content of bacterial RNA in the sample prior to sequencing. A relatively quick and affordable way of achieving this is via quantitative real-time PCR (qRT-PCR). For instance, for the here-described infection system total RNA from uninfected HeLa cells and *Salmonella* in vitro cultures may be separately isolated and treated with DNase I (using the protocol described in Subheading 3.7).
2. The RNA concentrations of the samples are measured using a NanoDrop device and a serial dilution of *Salmonella*-to-HeLa RNA is set up (Fig. 4a). The final RNA dilutions are adjusted to the same concentration with RNase-free water.
3. In our example, using qRT-PCR primers against *Salmonella rfaH* mRNA and the human marker mRNA *ACTB*, the ΔC_t-value ($\Delta C_t = C_{t;\ \text{host transcript}} - C_{t;\ \text{bacterial transcript}}$) is determined for each dilution in technical triplicates.
4. The fold changes relative to the 1:1 mixture are then calculated based on the $\Delta\Delta C_t$-method by Livak and Schmittgen

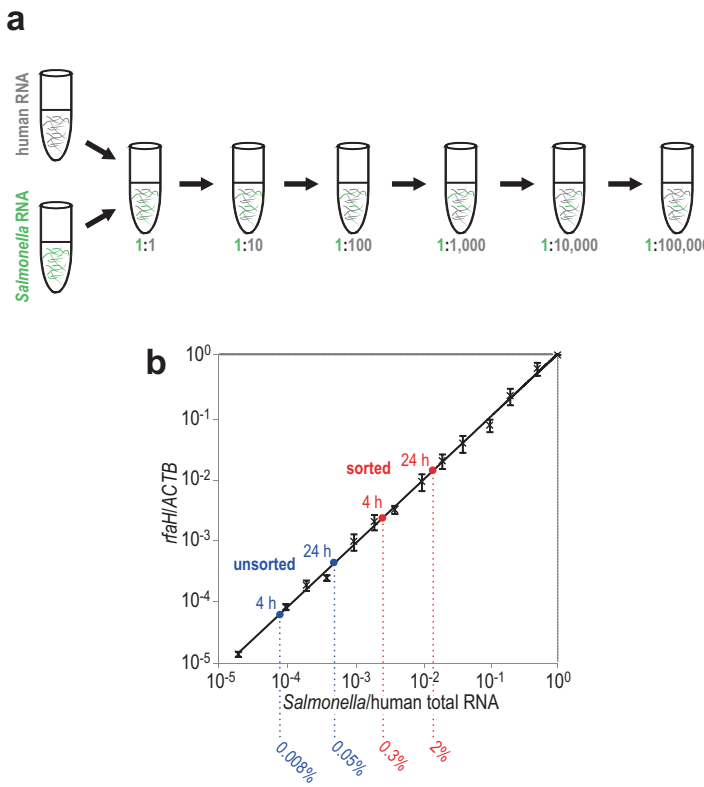


Fig. 4 qRT-PCR-based extrapolation of bacterial RNA proportions. (**a**) Schematic outline of the serial dilution of *Salmonella* to human RNA. (**b**) The obtained relationship between the pre-adjusted bacterial-to-host total RNA ratios and the calculated fold changes in the detection of bacterial and host reference transcripts [5] is used to estimate the relative proportion of *Salmonella* RNA in the infection samples (two different time points after infection; with or without FACS-based enrichment of invaded host cells; mean +/− SD from three replicates). Panel b was modified from [3]

[5], and plotted against the predefined mixing ratios to derive a trend-line equation describing the relation of the measured values with the pre-adjusted transcriptome proportions (black in Fig. 4b).

5. Finally, the same marker transcripts are measured in the infection samples to deduce the approximate ratio of *Salmonella*-to-host RNA (blue and red dots in Fig. 4b).

3.9 Double rRNA Depletion

1. Traditional RNA-seq approaches often rely on rRNA depletion from samples prior to sequencing, thereby reducing sequencing depth requirements. Naturally, Dual RNA-seq samples contain ribosomal transcripts from both infection partners. To remove mammalian and bacterial rRNA in a single step, the

Ribo-Zero Magnetic Gold epidemiology kit (Illumina) is used following the manufacturer's guidelines with the exception that typically less than 500 ng of RNA is used as an input (usually ~200 ng). The initial sample volume is adjusted to 28 μL (*see* **step 6** in Subheading 3.7).

2. Upon ribo-depletion, the final RNA samples are purified by ethanol precipitation as above (**steps 5** and **6** in Subheading 3.7; precipitate for at least 3 h at −20 °C).
3. For elution, the purified, air-dried RNA pellets are dissolved in each 10 μL of pre-warmed, RNase-free water for 2 min at 65 °C. Following this protocol, both human and *Salmonella* rRNA transcripts are efficiently removed (Fig. 5).

3.10 cDNA Library Construction

1. RNA samples are fragmented using ultrasound (four pulses of each 30 s) at 4 °C to generate ~200–400 nt (average) fragmentation products.
2. Short RNA fragments (<20 nt) are removed using the Agencourt RNAClean XP kit (Beckman Coulter Genomics).
3. The remaining fragments are dephosphorylated with Antarctic Phosphatase (NEB), re-phosphorylated with T4 Polynucleotide Kinase (NEB), and poly(A)-tailed using poly(A) polymerase (NEB). *See* **Note 3**.
4. An RNA adapter is ligated to the 5′ monophosphate of the RNA fragments using T4 RNA ligase (NEB).
5. First-strand cDNA synthesis is performed using an oligo(dT) adapter primer and the M-MuLV reverse transcriptase (NEB).
6. The resulting cDNAs are PCR-amplified to about 10–20 ng/μL using the Phusion high fidelity DNA polymerase (NEB) and primers designed for TruSeq sequencing according to the instructions of Illumina. Hexameric barcode sequences ("NNNNNN") required for multiplexing multiple cDNA libraries in a single lane of the flow cell are part of the 3′ sequencing adapter (*see* **Note 3**).

3.10.1 Sense Primer

5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′.

3.10.2 Antisense Primer

5′-CAAGCAGAAGACGGCATACGAGAT-NNNNNN-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC(dT25)-3′.

The combined length of the flanking sequences is 146 nt.

7. The resulting cDNA libraries are purified using the Agencourt AMPure XP kit and analyzed by capillary electrophoresis (Shimadzu MultiNA microchip electrophoresis system).

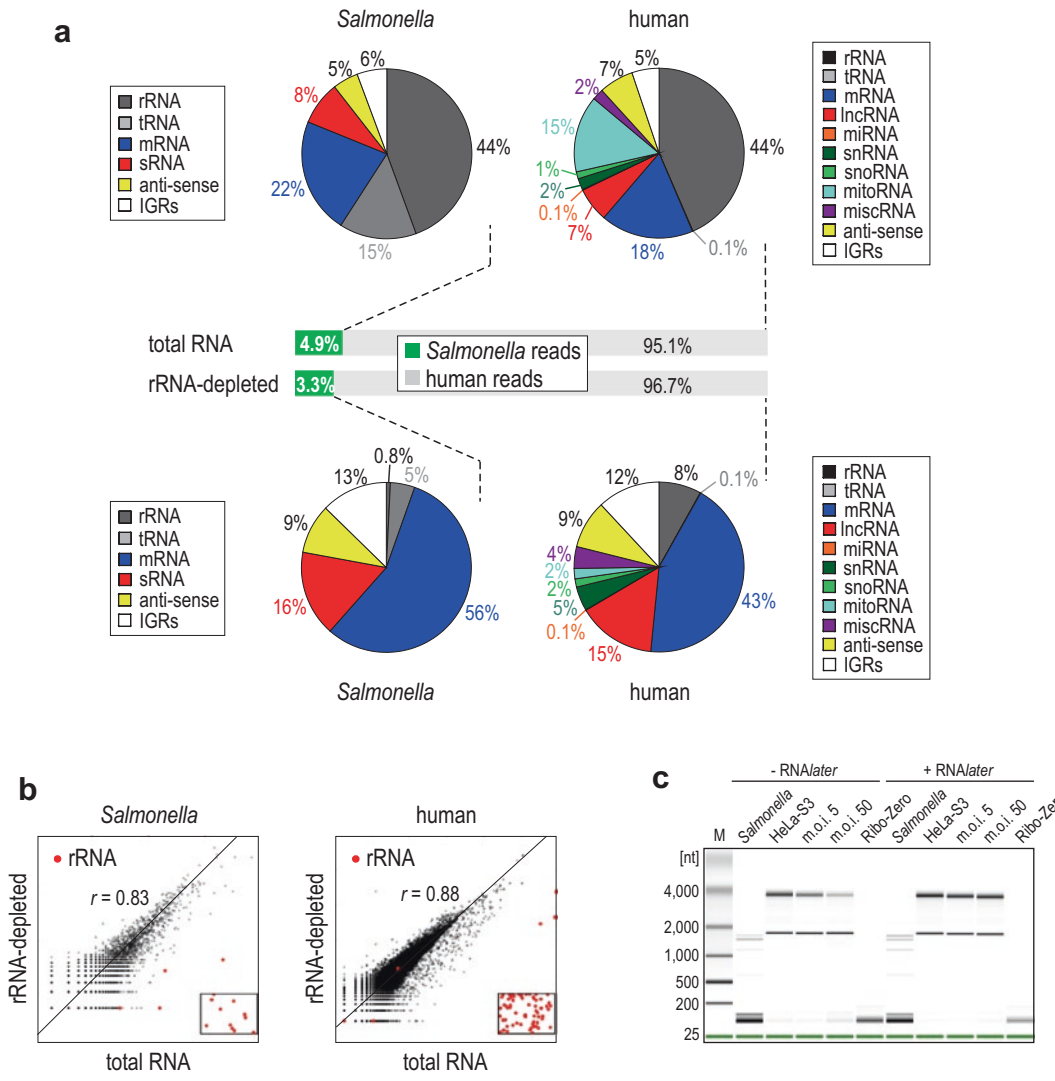


Fig. 5 Host-pathogen rRNA depletion. (**a**) Comparative mapping statistics of Dual RNA-seq samples without (upper panel) or upon the joint depletion of bacterial and eukaryotic rRNA using the Ribo-Zero Gold epidemiology kit (lower panel). (**b**) Gene-wise correlation between read coverages without or upon rRNA removal for *Salmonella* (left) and human (right) data subsets. The Pearson's *r* is given. The red dots in squares represent the rRNA transcripts that had zero reads in the rRNA-depleted sample. (**c**) Bioanalyzer image of gDNA-free RNA samples: *Salmonella* only RNA, HeLa only RNA, RNA from infected samples (m.o.i. 5 or 50), and Ribo-Zero-treated RNA from a *Salmonella*-infected HeLa sample (m.o.i. 50), without or upon RNA*later* fixation. "M" refers to the size marker. Panels a and b were obtained from [3]

3.11 Illumina Sequencing

1. For Illumina HiSeq sequencing, cDNA libraries from **step 7** in Subheading 3.10 are pooled in approximately equimolar amounts.
2. The cDNA pools are size-fractionated in the range of 150–600 bp via a differential clean-up with the Agencourt AMPure kit (Beckman Coulter Genomics).

3. Aliquots of the cDNA pools are analyzed by capillary electrophoresis (Shimadzu MultiNA microchip electrophoresis system) and sequenced on an Illumina HiSeq2500 device in single-end mode with 100 cycles (typically to ~30 million reads/library).

3.12 Bioinformatic Analyses

1. Illumina reads in FASTQ format are trimmed with a Phred quality score cut-off of 20 by the program fastq_quality_trimmer from FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/).
2. Reads shorter than 20 nt after adaptor- and poly(A)-trimming are discarded.
3. The remaining reads are aligned to the respective reference genome sequences in parallel; in our example to the *Salmonella enterica* SL1344 genome (NCBI RefSeq accession numbers: NC_016810.1, NC_017718.1, NC_017719.1, NC_017720.1) and the human genome (hg19—GRCh37; retrieved from the 1000 Genomes Project [6]). The mapping is performed using the READemption pipeline [7] and the short read mapper segemehl and its remapper lack [8] allowing for split reads [9].
4. Mapped reads with an alignment accuracy <90% as well as cross-mapped reads, that is, reads which can be aligned equally well to both host and bacterial reference sequences, are discarded. *See* **Note 4**.
5. Differential gene expression analysis is carried out separately for the host and the pathogen using the edgeR package [10] with an upper-quartile normalization and a prior count of 1.
6. If needed (i.e., to correct for batch effects), sequencing data might further be normalized using the RUVs correction method [11]. *See* **Note 5**.
7. For a detailed description of existing pathway analysis tools and interspecies network inference analyses, *see* [2].

4 Notes

1. We use the L/B buffer of the *mir*Vana kit (Ambion) for the joint lysis of HeLa cells and the *Salmonella* cells contained within (Subheading 3.7). Note, however, that when using bacterial pathogens other than *Salmonella* (especially Gram-positive species), it might be necessary to mechanically lyse the infected cells (e.g., by bead beating) to prevent the loss of the bacterial transcriptome.
2. It is important to assess for any experimental setup whether fixation/transcriptome stabilization is required prior to RNA extraction (*see* Subheading 3.5). Our data suggest that, within the described infection system, samples may be preserved best in RNA*later*. We note, however, that transcriptome stabiliza-

tion always is a compromise. For example, despite RNA*later* stabilizing the bacterial and human transcripts measured here (Fig. 2c, d), overall it leads to reduced RNA yields (Fig. 2b), increases the risk of cell clumping during sorting, and—unless removed completely—may quench the fluorescent signal. Treating infected cells instead with RNA*protect* (Qiagen), which is similar to RNA*later* but is recommended by the manufacturer for cell culture models, did not yield substantially more RNA (Fig. 2b). Importantly, RNA*protect* was inferior to RNA*later* with respect to the stabilization of the selected *Salmonella* and human marker transcripts (Fig. 2c, d). In experimental setups where cell samples do not need to be stored for longer periods but can be immediately processed, transcriptome stabilization seems expendable, especially when the starting material is limited.

3. The present protocol for cDNA library generation (Subheading 3.10) was optimized for sequencing on the HiSeq2500 platform (Illumina). At present, the NextSeq500 instrument (Illumina) is increasingly used for RNA-seq approaches. To render Dual RNA-seq compatible with the NextSeq500, some small modifications should be made to the cDNA synthesis protocol. Most importantly, an adapter consisting of an oligo(A) stretch is ligated to the 3′ end of the RNA fragments (rather than adding a poly(A) stretch with poly(A) polymerase). First-strand cDNA synthesis is then performed as described using M-MuLV and an oligo(dT) adapter primer, followed by the 5′ TruSeq sequencing adapter ligation to the 3′ end of the antisense cDNA. PCR amplification is as described except that the antisense primer contains an octameric (instead of a hexameric) barcode.
4. Cross-mapped reads have to be identified and discarded (*see* Subheading 3.12). For example, the percentage of cross-mapped reads in the previous HeLa time course infection experiment [3] increased as infection progressed, implying that these reads were mainly contributed by *Salmonella* (Fig. 6a). Even if the overall proportion of cross-mapped reads was small in this case (from 0.02 to 1.44%), they were removed in order not to inflate the differential expression analysis.

Fig. 6 (continued) The principle component analysis (PCA) plots (upper) indicate that after batch correction, the three biological replicates of HeLa cells infected with two different *Salmonella* strains (red and blue) segregate along PC 1, which accounts for 66% of the variance. In contrast, before the correction individual replicates did not cluster with one another. Likewise, batch correction led to an increase in significantly differentially expressed genes between the two conditions (lower). The data displayed are derived from the 8 h time point, but are representative for the entire time course

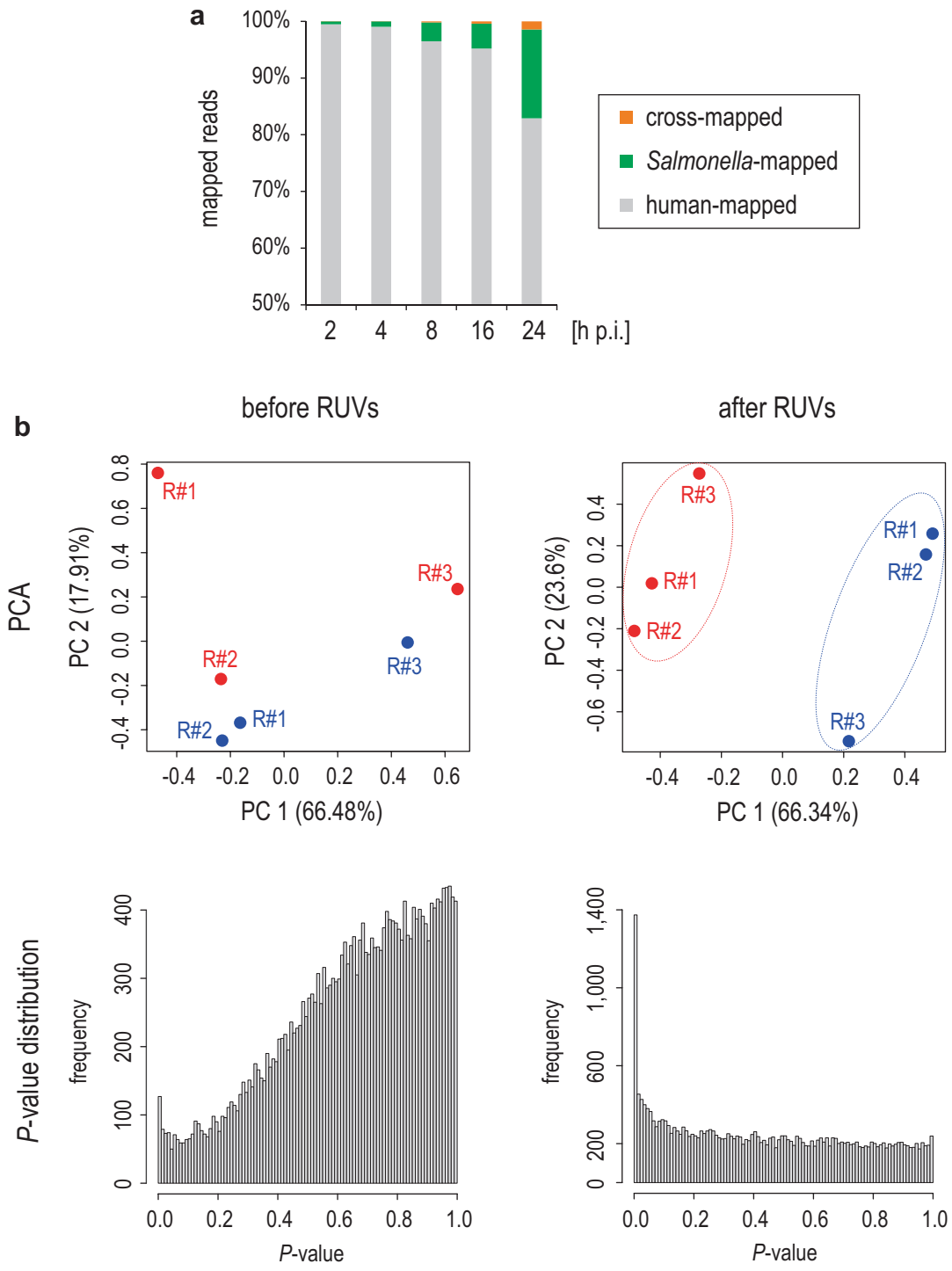


Fig. 6 Quantification of cross-mapped reads and batch correction. (**a**) Distribution of reads that either mapped exclusively to one reference genome, or to both genomes ("cross-mappings") over one replicate of the HeLa infection time course described in [3]. The y-axis was cropped at 50% to enlarge the view on bacterial and cross-mapped reads. (**b**) RUVs sanity plots for Dual RNA-seq data from the same infection experiment [3].

5. There are many factors which are difficult or impossible to control for when working with samples as complex as mammalian cells infected with bacterial pathogens, for instance differences in medium batches, in handling, or the room temperature in the laboratory. However, even minor variations in an experiment can cause widespread changes in high-throughput data [12]. The RUVs approach (*see* Subheading 3.12) performs a factor analysis, extracting factors unrelated to treatment status which can then be included in a standard GLM analysis for differential expression [11]. Application of RUVs to our previous HeLa time course infection assay [3] successfully removed factors of unwanted variation during testing for differential expression between matched time points of the two different *Salmonella* strains used for infection (Fig. 6b).

Acknowledgments

The authors thank Dr. Konrad Förstner and Dr. Lars Barquist for help with Subheading 3.12 and Fig. 6. cDNA library generation and sequencing (Subheadings 3.10 and 3.11) were performed by Vertis Biotechnologie AG (Freising, Germany). AJW was recipient of a stipend from the Elite Network of Bavaria.

References

1. Westermann AJ, Gorski SA, Vogel J (2012) Dual RNA-seq of pathogen and host. Nat Rev Microbiol 10(9):618–630. https://doi.org/10.1038/nrmicro2852
2. Westermann AJ, Barquist L, Vogel J (2017) Resolving host-pathogen interactions by dual RNA-seq. PLoS Pathog 13(2):e1006033. https://doi.org/10.1371/journal.ppat.1006033
3. Westermann AJ, Forstner KU, Amman F, Barquist L, Chao Y, Schulte LN, Muller L, Reinhardt R, Stadler PF, Vogel J (2016) Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature 529(7587):496–501. https://doi.org/10.1038/nature16547
4. Papenfort K, Said N, Welsink T, Lucchini S, Hinton JC, Vogel J (2009) Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. Mol Microbiol 74(1):139–158. https://doi.org/10.1111/j.1365-2958.2009.06857.x
5. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method. Methods 25(4):402–408. https://doi.org/10.1006/meth.2001.1262
6. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491(7422):56–65. https://doi.org/10.1038/nature11632
7. Forstner KU, Vogel J, Sharma CM (2014) READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. Bioinformatics 30(23):3421–3423. https://doi.org/10.1093/bioinformatics/btu533
8. Otto C, Stadler PF, Hoffmann S (2014) Lacking alignments? The next-generation sequencing mapper segemehl revisited. Bioinformatics 30(13):1837–1843. https://doi.org/10.1093/bioinformatics/btu146
9. Hoffmann S, Otto C, Doose G, Tanzer A, Langenberger D, Christ S, Kunz M, Holdt L, Teupser D, Hackermueller J, Stadler PF (2014) A multi-split mapping algorithm for circular RNA, splicing, trans-splicing, and fusion detec-

tion. Genome Biol 15(2):R34. https://doi.org/10.1186/gb-2014-15-2-r34

10. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1): 139–140. https://doi.org/10.1093/bioinformatics/btp616
11. Risso D, Ngai J, Speed TP, Dudoit S (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol 32(9):896–902. https://doi.org/10.1038/nbt.2931
12. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet 11(10):733–739. https://doi.org/10.1038/nrg2825

Chapter 5

Identification of New Bacterial Small RNA Targets Using MS2 Affinity Purification Coupled to RNA Sequencing

Marie-Claude Carrier, Guillaume Laliberté, and Eric Massé

Abstract

Small regulatory RNAs (sRNAs) are ubiquitous regulatory molecules expressed in living cells. In prokaryotes, sRNAs usually bind to target mRNAs to either promote their degradation or interfere with translation initiation. Because a single sRNA can regulate a considerable number of target mRNAs, we seek to identify those targets rapidly and reliably. Here, we present a robust method based on the co-purification of target mRNAs bound to MS2-tagged sRNAs expressed in vivo. After purification of the tagged-sRNA, we use RNAseq to determine the identity of all RNA interacting partners and their enrichment level. We describe how to analyze the RNAseq data through the Galaxy Project Platform bioinformatics tools to identify new mRNA targets. This technique is applicable to most sRNAs of *E. coli* and *Salmonella*.

Key words Small regulatory RNAs, sRNA targets, MAPS, Targetome, Affinity purification, MS2 aptamer, RNA–RNA interaction

1 Introduction

Bacterial small regulatory RNAs (sRNAs) are key actors in the fine-tuning of gene expression, ensuring rapid adaptation of bacteria to their ever-changing environment. sRNAs typically act at the post-transcriptional level by base-pairing to their messenger RNA (mRNA) targets in the 5′ untranslated region (UTR) [1]. Remarkably, limited complementarity between the sRNA and its targets not only allows the regulation of multiple mRNAs by a single sRNA but also the regulation of one mRNA by multiple sRNAs. This added complexity creates an extensive regulatory network where sRNAs act as bridges between various cellular metabolisms [2]. In the last decades, such networks were studied and specific sRNA targetomes were, in part, characterized. Since then, this field of study witnessed an explosion of technological advances [3] that exposed the versatility of sRNAs in terms of possible pairing sites and mechanisms of action. Indeed, these short regulators not only pair in the 5′ UTR of mRNAs, they can also target the

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_5,

coding sequence (CDS) [4, 5] or could even pair in the 3′ UTR of targets. Moreover, sRNAs can regulate the translation of mRNA targets without directly affecting the stability of the transcript [6]. This new knowledge exposed a significant lack in efficacy of the classical techniques used to identify targets of sRNAs, reviving the challenge of sRNA target identification in bacterial cells. To tackle this issue, we developed and optimized a technique that combines RNA affinity purification and RNA sequencing (RNAseq) allowing genome-wide identification of sRNA–mRNA interaction in bacterial cells. The assay is called MAPS: *M*S2 *a*ffinity *p*urification coupled to RNA*s*eq. Here, we describe the MAPS protocol in detail. Briefly, a sRNA is tagged with an MS2 RNA aptamer and expressed in vivo. Following cell lysis, tagged sRNAs are purified through affinity chromatography. Eluted RNAs are analyzed by high-throughput RNAseq and the ratio of enriched mRNAs in the tagged vs. untagged sRNA experiments is representative of the interaction between the two RNAs. Moreover, we describe the bioinformatic pipeline used to analyze MAPS data exploiting the Galaxy Project Platform.

2 Materials

Ultrapure water should be used for every solution. Make sure to work with RNase-free material to avoid degradation of RNAs throughout the experiment.

2.1 Cell Harvesting

1. Bacterial strain: sRNA knock-out strains containing plasmids carrying either a control sRNA or the MS2-sRNA construct (*see* **Note 1**).
2. LB media: 10 g bio-tryptone, 5 g yeast extract, and 5 g NaCl in 1 L of water. Sterilize by autoclaving.
3. 20% w/v L-arabinose. Sterilize by 0.22 μm filtration.
4. Buffer A: 20 mM Tris–HCl (pH 8), 150 mM KCl, 1 mM $MgCl_2$, 1 mM dithiothreitol (DTT), and 1 mM phenylmethylsulfonyl fluoride (PMSF) (*see* **Note 2**).
5. Temperature-controlled agitating water bath.
6. Spectrophotometer.
7. Centrifuge accommodating 50 mL conic tubes.
8. Liquid nitrogen.

2.2 Cell Lysis

1. French Press with a lysis cell that can accommodate at least 3.5 mL.
2. Buffer A: 20 mM Tris–HCl (pH 8), 150 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, and 1 mM PMSF.

2.3 Affinity Purification

1. Amylose resin.
2. Disposable Bio-Spin chromatography columns.
3. Purified MS2-MBP fusion protein. MBP: maltose-binding protein [7].
4. Buffer A: 20 mM Tris–HCl (pH 8), 150 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, and 1 mM PMSF.
5. Elution buffer: Buffer A supplemented with 15 mM maltose.
6. Phenol-water, pH 6: Melt phenol crystals at 65 °C and preheat an equal volume of ultrapure water to the same temperature. Carefully mix equal volume of liquid phenol and ultrapure water. Add 0.1% w/(phenol volume) 8-hydroxyquinoline and mix carefully. Incubate 5 min at 65 °C. Aliquot in 50 mL conical tubes. Keep at 4 °C, protect from light.
7. 25:24:1 (v/v/v) phenol-chloroform-isoamyl alcohol.
8. Glycogen.
9. 95% v/v ethanol.
10. 75% v/v ethanol.

2.4 Samples Preparation for RNA Sequencing

1. 10× TURBO™ DNase Buffer.
2. TURBO™ DNase.
3. 25:24:1 (v/v/v) phenol-chloroform-isoamyl alcohol.
4. 95% v/v ethanol.
5. Agilent Nano Chip.
6. Bioanalyzer 2100.
7. ScriptSeq™ v2 RNA-Seq Library Preparation Kit from Illumina.
8. MiSeq (Illumina).

3 Methods

3.1 Harvesting the Cells

1. Dilute an overnight bacterial culture 1/1000 in 100 mL of fresh LB media supplemented with the appropriate antibiotics. Grow the cultures at 37 °C with agitation.
2. Once the cultures have reached an OD_{600nm} of 0.5, induce expression of the sRNA or MS2-sRNA construct by addition of 0.1% arabinose (*see* **Notes 3** and **4**).
3. Incubate for 10 min at 37 °C with agitation.
4. Transfer the flasks to an ice slurry for 10 min (*see* **Note 5**).
5. Transfer each culture to two 50 mL conic tubes. Centrifuge at 2900 × *g* at 4 °C, 15 min.
6. Discard the supernatant.

7. Resuspend cells from each starting flask in a total of 1 mL of Buffer A.
8. Centrifuge at 16,000 × *g* for 5 min. Discard the supernatant.
9. Freeze the pelleted cells in liquid nitrogen. Keep at −80 °C.

3.2 Cell Lysis

All steps should be performed on ice. All buffers should be at 4 °C.

1. Let the cell pellets thaw on ice for 30 min.
2. Resuspend the pellet in 2 mL of Buffer A (*see* **Notes 6** and 7).
3. Chill the French Press cell by burying it in ice before performing the lysis.
4. Break the bacterial cells using a French Press at 430 psi, 3 times per sample. Keep samples on ice at all times (*see* **Note 8**).
5. Clear the lysates by centrifugation at 16,000 × *g* at 4 °C, 30 min.
6. Transfer the soluble fraction (lysate) to clean tubes. Keep on ice.

3.3 MS2 Affinity Purification

All steps should be performed on ice. All buffers should be at 4 °C.

1. Add 75 µL amylose resin to a Bio-Spin disposable chromatography column.
2. Equilibrate the column three times with 1 mL of Buffer A (*see* **Note 9**).
3. Use the provided stopper to seal the column. Dilute 100 pmol of MS2-MBP coat protein in 1 mL Buffer A. Apply the protein solution to the sealed column and let stand for 5 min.
4. Remove the stopper and let the column drain.
5. Wash the column twice with 1 mL of Buffer A.
6. Load the bacterial lysate onto the column, 1 mL at a time and let the column drain.
7. Wash the column 5 times with 1 mL of Buffer A (*see* **Note 10**).
8. Insert the column in a clean RNase-free collecting tube. Elute with 1 mL of Elution Buffer.
9. Split the column output into two 1.5 mL microtubes.
10. Add 1 volume of phenol-chloroform-isoamyl alcohol to each tube and mix. Centrifuge at 16,000 × *g* at room temperature, 10 min.
11. Transfer the aqueous phase in clean microtubes containing 20 mg of glycogen (*see* **Note 11**).
12. Add two volumes of 95% EtOH. Mix thoroughly and precipitate overnight at −80 °C.
13. Centrifuge the samples at 16,000 × *g* at 4 °C, 30 min.
14. Remove the supernatant very carefully and add 500 µL of ice-cold 75% EtOH to the pellets. Centrifuge the samples at 16,000 × *g* at 4 °C, 5 min (*see* **Note 12**).

15. Remove the supernatant. Let the RNA pellets dry completely.
16. Resuspend the pellets in 86 µL of ultrapure H_2O. Proceed to the next step (*see* **Note 13**).

3.4 Samples Preparation for RNAseq

1. Add 10 µL of 10× TURBO™ DNase Buffer and 4 U of TURBO™ DNase to each sample.
2. Incubate at 37 °C, 30 min.
3. Add 100 µL of phenol-chloroform-isoamyl alcohol to each tube and mix. Centrifuge at 16,000 × *g* at room temperature, 10 min.
4. Add 2.5 volumes of 95% EtOH. Mix thoroughly and precipitate overnight at −80 °C.
5. Centrifuge the samples at 16,000 × *g* at 4 °C, 30 min.
6. Remove the supernatant carefully. Let the RNA pellets dry completely.
7. Resuspend the dried pellets in 6 µL of ultrapure H_2O.
8. Quantify and verify the quality of the RNA using Agilent Nano Chip in a Bioanalyzer 2100.
9. Prepare the cDNA libraries with the ScriptSeq™ v2 RNA-Seq Library Preparation Kit from Illumina.
10. Sequence the libraries in both directions using Illumina MiSeq.

3.5 Data Processing

Bioinformatics tools used are freely available on the Galaxy Platform [8] (https://usegalaxy.org/).

3.5.1 Reads Alignment and Visualization

The following procedure allows the alignment and visualization of the RNA sequencing reads on the genome of interest. Note that the procedure has to be performed independently for the experimental data set (MS2-sRNA) and the control data set (sRNA). Here, the procedure will be detailed for one experimental data set with paired-end sequencing. Refer to Fig. 1a for visual workflow and to Table 1 for Galaxy Project tool details.

1. Log in on the Galaxy Project platform. *Upload files*: FastQ sequencing files downloaded through the Illumina platform.
2. Execute the *FastQ Groomer* tool (Sanger & Illumina 1.8+) [9] to convert each FASTQ files.
3. Run a *FastQC* [10] analysis to visualize the quality of the sequences. If the initial score of the raw sequences is >20, proceed to the next step. If quality of raw sequences data is <20, run a *FastQ Quality Trimmer* [9] to trim the 5′ and 3′ ends using a Quality Score >20 (*see* **Note 14**).
4. Use *Map with Bowtie for Illumina* [11] and select the paired-end analysis. As input, use the *FastQ Groomer* file or the *Quality*

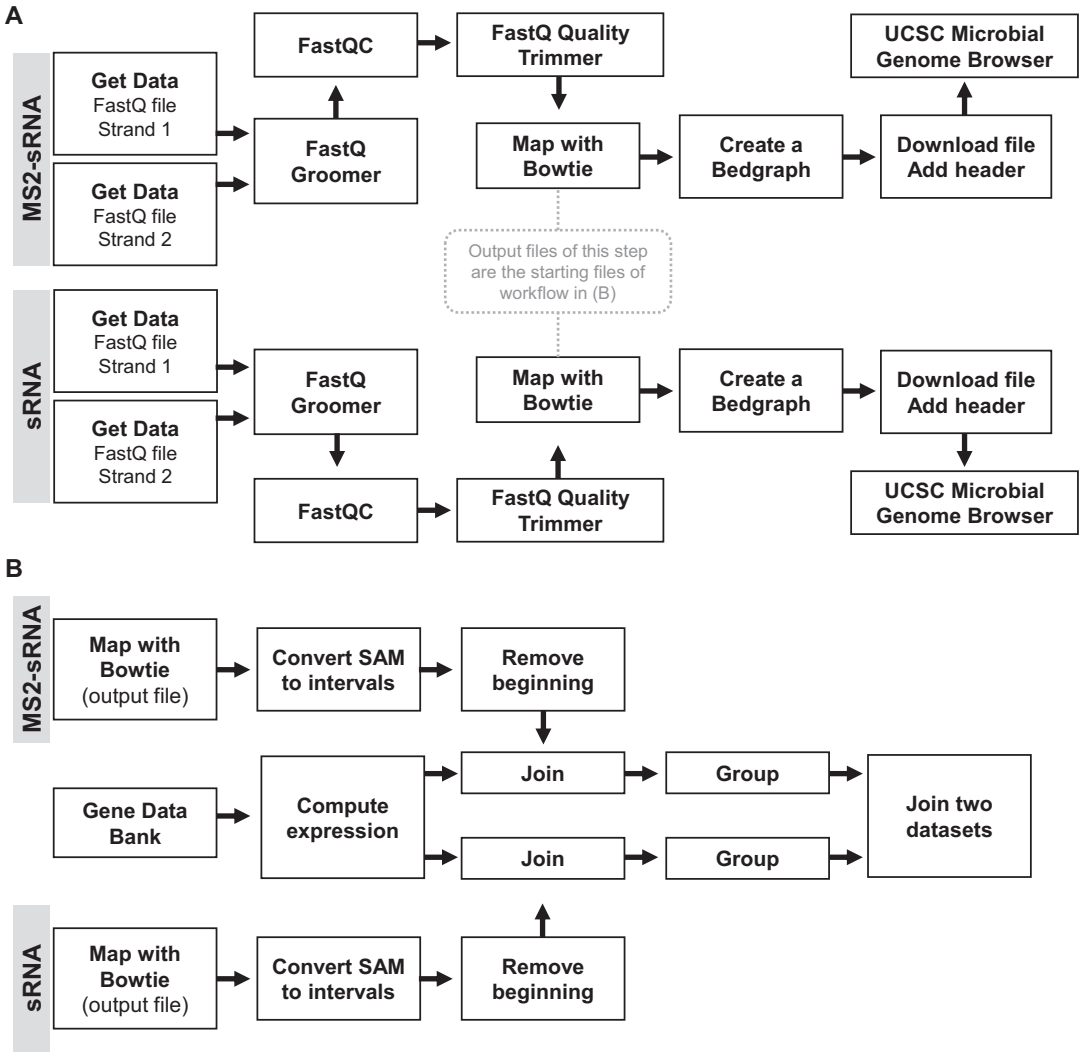


Fig. 1 Schematic representation of MAPS data analysis. (**a**) Workflow allowing the alignment of RNA sequencing reads on the genome. (**b**) Workflow leading to the assignment of reads to specific genes and to the calculation of enrichment ratios between experimental and control conditions

Trimmer file. Select your reference genome. Here, *Escherichia coli* (str. K-12 substr. MG1655): eschColi_K12 was used. Under *Bowtie settings to use*, choose *Full parameters list*. Choose a value for the *Seed* parameter (here, 25) and set the *Maximum number of mismatch allowed in the seed* (here, 1).

5. Navigate to *Genome Coverage* [12]. Select the Bowtie output data sets (SAM files) previously obtained. Select *Report region with zero coverage*.
6. Download the output file to your computer.
7. Open the downloaded file as a text file. Before the first line, add the following header and save the file: track type = bedGraph

Table 1
Detailed information on the Galaxy Project platform tools used for MAPS data analysis

Reference to protocol	Tool name	Category	Version	Reference
Subheading 3.5.1, **step 1** Subheading 3.5.2, **step 4**	Upload files	Get data	–	
Subheading 3.5.1, **step 2**	FastQ Groomer	NGS: QC and manipulation	1.0.4	Blankenberg (20100
Subheading 3.5.1, **step 3**	FastQC	NGS: QC and manipulation	0.67	Andrews (2010)
Subheading 3.5.1, **step 3**	FastQ Quality Trimmer	NGS: QC and manipulation	1.0.0	Blankenberg (2010)
Subheading 3.5.1, **step 4**	Map with Bowtie for Illumina	NGS mapping	1.1.2	Langmead (2009)
Subheading 3.5.1, **step 5**	Create a BedGraph of genome coverage	BedTools	2.26.0.0	Quinlan (2010)
Subheading 3.5.2, **step 1**	Convert SAM to interval	NGS: SAMtools	1.0.1	Galaxy development team
Subheading 3.5.2, **step 2**	Remove the beginning of a file	Text manipulation	1.0.0	Galaxy development team
Subheading 3.5.2, **step 5**	Compute an expression on every row	Text manipulation	1.1.0	Galaxy development team
Subheading 3.5.2, **step 7**	Join	Operate on genomic intervals	1.0.0	Galaxy development team
Subheading 3.5.2, **step 8**	Group	Join, subtract and group	2.1.1	
Subheading 3.5.2, **step 9**	Join two datasets side by side on a specific field	Join, subtract and group	2.0.2	

name = "NAME_EXPERIMENT" description = "BedGraph format" visibility = full (*see* **Note 15**).

8. Navigate to the UCSC Microbial Genome Browser [13] (http://microbes.ucsc.edu/).
9. Enter the name of the reference genome that you need. Here, we used *Escherichia coli* K12.
10. Click on the *Manage custom tracks* button and add your custom tracks (*see* **Note 16**).
11. Click on *View in Genome Browser*. You can now search for a gene name or genomic position to visually compare read alignment on the genome (Fig. 2) (*see* **Note 17**).

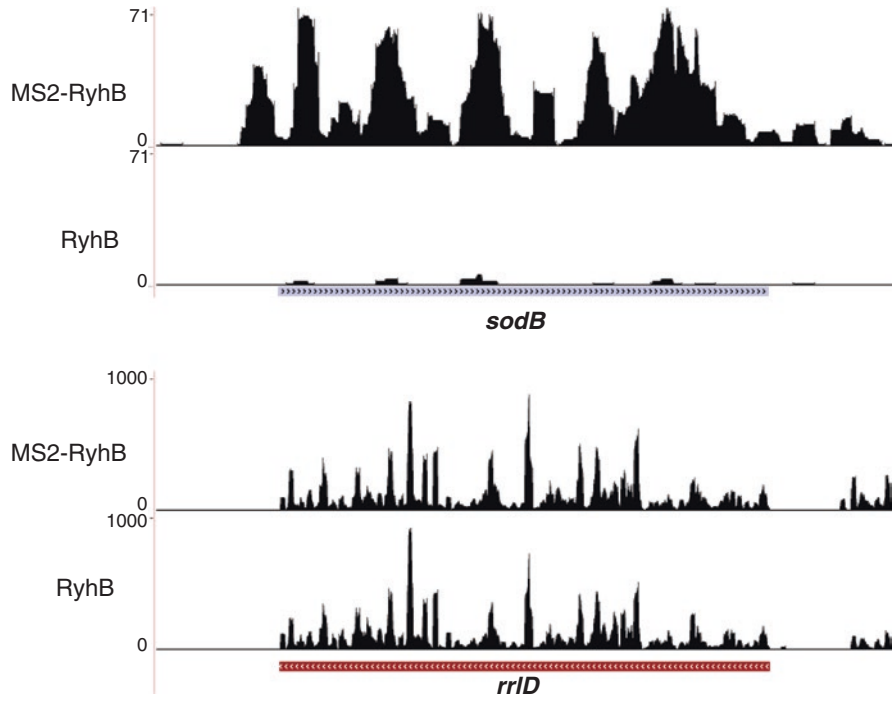


Fig. 2 MAPS reads alignment visualization. (**a**) MS2-RyhB and RyhB MAPS reads aligned on *sodB*, a characterized target of RyhB [17]. (**b**) MS2-RyhB and RyhB MAPS reads aligned on *rrlD* (23S rRNA), serving as a nontarget control. Data was visualized using the microbes and archaea UCSC Genome Browser (http://microbes.ucsc.edu)

3.5.2 Assignment of Reads to Specific Genes and Determination of Enrichment Ratios

The following procedure is performed to assign the RNAseq reads to gene names and to compare the experimental data set (MS2-sRNA) to the control (sRNA) data set.

To perform this step, three files are required and need to be processed. **Steps 1** and **2** are performed on the SAM files from **step 4** in Subheading 3.5.1 for the MS2-sRNA experimental condition and for the sRNA control condition. **Steps 4** and **5** are performed on the file acquired in **step 3** of this section. Refer to Fig. 1b for visual workflow and to Table 1 for Galaxy Project tool details.

1. Run the *Convert SAM to interval* tool with default parameters on each SAM files.
2. On the converted files, run the *Remove beginning of a file* tool. These data sets are ready to use.
3. Through the NCBI database, download the Gene Data Bank of the bacterial strain corresponding to your experimental strain. This is a .txt file.
4. *Upload files*: Gene Data Bank file. Set the file type as Interval.

5. Run *Compute an expression on every row* with the Add Interval parameter c3–c2.
6. On the output from **step 5**, go to *edit attributes* and set the Database/Build as the reference genome used in **step 4** of Subheading 3.5.1. This data file is ready to use.
7. Run *Join* tool. In the parameters, join your sample data set (MS2-sRNA or sRNA; **step 2** in Subheading 3.5.2) with the Gene Data Bank data set (*see* **step 6**). Use default parameters.
8. Run the *Group* tool for each output files obtained at **step 7**. Select the following parameters:
   (a) Group by column: column 23.
   (b) Ignore case by grouping: no.
   (c) Ignore lines beginning with these characters: select all characters except for the dot (.).
   (d) Operation: insert Operation 1, Count Distinct on column 5.
   (e) Operation: insert Operation 2, Mean on column 24.
9. Run *Join two Datasets side by side on a specified field* tool with outputs from **step 8**. Join the MS2-sRNA sample using column 1 with the sRNA control sample using column 1. Set the following parameters:
   (a) Keep lines of first input that do not join with second input: Yes.
   (b) Keep lines of first input that are incomplete: No.
   (c) Fill Empty columns: yes.
   (d) Fill column by: Single fill value.
   (e) Fill value: 1.
10. Download the output file. Open the file with Microsoft Excel (or any similar software). The first 3 columns represent the gene name, the number of reads and the gene length for the MS2-sRNA experiment. The last three columns represent the same for the sRNA control.
11. Relativize the number of reads (*see* **Note 18**):
    (a) From the Illumina Dashboard, note the total number of reads and the total number of reads mapped for each sample.
    (b) For the MS2-sRNA, calculate the relativized number of reads: Reads/((total number of reads) X (the total number of reads mapped for "MS2-sRNA"))
    (c) For the MS2-sRNA, calculate the relativized number of reads: Reads/((total number of reads) X (the total number of reads mapped for "sRNA"))
12. Calculate the enrichment ratio for each gene between the MS2-sRNA experiment and the sRNA control (Fig. 3).

	Reads count		Relativized reads count			
Gene	MS2-RyhB	RyhB	MS2-RyhB	RyhB	Enrichment ratio	Reference
leuZ	542	1	2,15E-04	4,35E-07	495,01	Lalaouna, 2015
glyW	424	2	1,68E-04	8,69E-07	193,62	Lalaouna, 2015
cysT	450	4	1,79E-04	1,74E-06	102,75	Lalaouna, 2015
sodB	818	24	3,25E-04	1,04E-05	31,13	Massé, 2002
sdhC	32	1	1,27E-05	4,35E-07	29,23	Massé, 2002
shiA	108	5	4,29E-05	2,17E-06	19,73	Prévost, 2007
cirA	6	2	2,38E-06	8,69E-07	2,74	Salvail, 2013
fumA	72	30	2,86E-05	1,30E-05	2,19	Massé, 2002
rrlD	23705	22707	9,41E-03	9,87E-03	0,95	N/A

Fig. 3 Example of final results after performing MAPS on MS2-RyhB. Enrichment ratios are obtained using relativized read counts. As a proof of concept, a sample of characterized targets of RyhB recovered using MAPS are shown. *rrlD* (23S rRNA) serves as a nontarget control. All parameters used are those specified in this protocol

3.6 Identification of New sRNA Targets

It is strongly recommended to perform in vivo validation of putative targets identified by MAPS. Various methods are useful to perform such validation. Northern blotting will be effective in assessing the sRNA-dependent modulation of the target at the RNA level [14]. Translational regulation can be determined using translational reporter-gene fusions. These procedures will not be detailed here as they are beyond the scope of this protocol.

4 Notes

1. DNA Sequence of the double MS2 RNA aptamer is as follows: 5′—CGTACACCATCAGGGTACGTTTTTCAGACACC ATCAGGGTCTG—3′.

   For details about the cloning procedure and control experiments that should be carried prior to the MAPS, refer to Corcoran et al. (2012) [15].
2. DTT and PMSF should be added immediately before use. Store Buffer A without DTT and PMSF at 4 °C.
3. Use the inducer required with the chosen vector.
4. Induction of the sRNA expression should be performed during the same growth-phase as endogenous expression to ensure interaction with real targets and avoid artifactual interactions.
5. After the slurry, it is recommended to take a 600 μL aliquot of bacterial culture and perform an extraction of total RNA. This

input sample will be useful at a later step (refer to **Note 11**). Various methods can be used for RNA extraction. We suggest the hot-phenol RNA extraction [16].

6. Do not resuspend all the pellets at once. Follow **steps 2–4** for each sample individually. Keep all samples on ice at all times.
7. Depending on your experimental conditions (for example if the cells were harvested at high OD_{600nm}), the pellets can be resuspended in 3 mL.
8. The number of passages on the French Press can vary according to your experimental conditions. For example, if cells were harvested at high OD_{600nm}, break the cells 4 times per sample.
9. Use a clean 10 mL-syringe to push the first few drops out of the column. Then, let the elution carry on by gravity only.
10. Number of washes is an important parameter and should be optimized for each experiment.
11. Addition of glycogen is very important, if not essential, to be able to recover the small RNA pellets and avoid their loss after precipitation. Glycogen must be in contact with the RNAs (the aqueous phase) before the addition of ethanol.
12. Be very careful as the RNA pellets don't always stick to the bottom of the tube. Remove ethanol using a micropipette. Avoid using a vacuum system.
13. At this step, it is possible and highly recommended to test your samples by Northern blot. Compare the input samples (*see* **Note 5**) with the output samples.
14. This parameter can be adjusted. If using the *FastQ Quality Trimmer* with the threshold at a score of 20 causes the loss of too many sequences, the threshold can be lowered. If the overall quality of sequences is above 20, we do not perform the *FastQ Quality Trimmer*.
15. This header is essential for the next step. It informs the UCSC Microbial Genome Browser on the type of data contained in the file and allows you to name your experiment.
16. Ideally, add both the control track (sRNA MAPS) and the experimental track (MS2-sRNA MAPS) to compare the reads aligned in each condition.
17. Enrichment at a specific location on a gene of interest doesn't always represent the exact pairing site of the sRNA. Additional experimental data is required to validate the pairing site localization.
18. Normally, reads must be relativized taking into consideration the size of the genes they mapped to. However, since we calculate the enrichment ratio of reads between the two conditions, gene size is irrelevant in our analysis.

Acknowledgments

This work was funded by an operating grant from the Canadian Institutes of Health Research (CIHR) to EM. M.-C.C. holds an Alexander Graham Bell Doctoral scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Wagner EGH, Romby P (2015) Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. Adv Genet 90:133–208
2. Modi SR, Camacho DM, Kohanski MA et al (2011) Functional characterization of bacterial sRNAs using a network biology approach. Proc Natl Acad Sci U S A 108:15522–15527. https://doi.org/10.1073/pnas.1104318108
3. Saliba A-E, C Santos S, Vogel J (2017) New RNA-seq approaches for the study of bacterial pathogens. Curr Opin Microbiol 35:78–87. https://doi.org/10.1016/j.mib.2017.01.001
4. Lalaouna D, Morissette A, Carrier M-C, Massé E (2015) DsrA regulatory RNA represses both HNS and RBS D mRNAs through distinct mechanisms in E scherichia coli. Mol Microbiol 98:357–369. https://doi.org/10.1111/mmi.13129
5. Wroblewska Z, Olejniczak M (2016) Hfq assists small RNAs in binding to the coding sequence of *ompD* mRNA and in rearranging its structure. RNA 22:979–994. https://doi.org/10.1261/rna.055251.115
6. Chabelskaya S, Gaillot O, Felden B (2010) A Staphylococcus aureus small RNA is required for bacterial virulence and regulates the expression of an immune-evasion molecule. PLoS Pathog 6:e1000927. https://doi.org/10.1371/journal.ppat.1000927
7. Lalaouna D, Prévost K, Eyraud A, Massé E (2017) Identification of unknown RNA partners using MAPS. Methods 117:28–34. https://doi.org/10.1016/j.ymeth.2016.11.011
8. Afgan E, Baker D, van den Beek M et al (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3–W10. https://doi.org/10.1093/nar/gkw343
9. Blankenberg D, Gordon A, Von Kuster G et al (2010) Manipulation of FASTQ data with Galaxy. Bioinformatics 26:1783–1785. https://doi.org/10.1093/bioinformatics/btq281
10. Andrews S (2010) Babraham bioinformatics - FastQC a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
11. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25
12. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033
13. Chan PP, Holmes AD, Smith AM et al (2012) The UCSC archaeal genome browser: 2012 update. Nucleic Acids Res 40:D646–D652. https://doi.org/10.1093/nar/gkr990
14. Desnoyers G, Massé E (2012) Activity of small RNAs on the stability of targeted mRNAs in vivo. Methods Mol Biol 905:245–255
15. Corcoran CP, Rieder R, Podkaminski D et al (2012) Use of aptamer tagging to identify in vivo protein binding partners of small regulatory RNAs. Methods Mol Biol 905:177–200
16. Aiba H, Adhya S, de Crombrugghe B (1981) Evidence for two functional gal promoters in intact Escherichia coli cells. J Biol Chem 256:11905–11910
17. Massé E, Vanderpool CK, Gottesman S (2005) Effect of RyhB small RNA on global iron use in Escherichia coli. J Bacteriol 187:6962–6971. https://doi.org/10.1128/JB.187.20.6962-6971.2005

Chapter 6

Assessment of External Guide Sequences' (EGS) Efficiency as Inducers of RNase P-Mediated Cleavage of mRNA Target Molecules

Saumya Jani, Alexis Jackson, Carol Davies-Sala, Kevin Chiem, Alfonso Soler-Bistué, Angeles Zorreguieta, and Marcelo E. Tolmasky

Abstract

RNase P is a ribozyme consisting of a catalytic RNA molecule and, depending on the organism, one or more cofactor proteins. It was initially identified as the enzyme that mediates cleavage of precursor tRNAs at the 5′-end termini to generate the mature tRNAs. An important characteristic of RNase P is that its specificity depends on the structure rather than the sequence of the RNA substrate. Any RNA species that interacts with an antisense molecule (called external guide sequence, EGS) and forms the appropriate structure can be cleaved by RNase P. This property is the basis for EGS technology, an antisense methodology for inhibiting gene expression by eliciting RNase P-mediated cleavage of a target mRNA molecule. EGS technology is being developed to design therapies against a large variety of diseases. An essential milestone in developing EGSs as therapies is the assessment of the efficiency of antisense molecules to induce cleavage of the target mRNA and evaluate their effect in vivo. Here, we describe simple protocols to test the ability of EGSs to induce cleavage of a target mRNA in vitro and to induce a phenotypic change in growing cells.

Key words Antisense, Ribozyme, RNase P, Antibiotic resistance, Aminoglycoside

1 Introduction

Antisense inhibition of gene expression by oligonucleotides or oligonucleotide analogs as the basis for the development of therapeutic agents has been researched for over two decades [1]. After approval of the first antisense drug, fomivirsen [2], success was elusive for a long time. However, continuing efforts led to the development of many prospective antisense drugs, and very recently mipomersen, nusinersen, and eteplirsen were approved by FDA for the treatment of homozygous familial hypercholesterolemia, Duchenne muscular dystrophy, and spinal muscular atrophy, respectively [3–5]. Additionally, many compounds are in clinical trials, and most probably new compounds will be approved for a

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_6,

variety of treatments in the near future [6, 7]. As a consequence, research on antisense technologies is experiencing renewed interest and momentum. However, the advances in antisense technologies applied to prokaryotes still lag in comparison to those that target other human diseases.

While the fundamental purpose of antisense technologies is to interfere with expression of specific proteins, several different approaches have been tried such as inhibition of translation by steric hindrance or by inducing mRNA degradation by endogenous RNases like RNase H or RNase P [8–10]. Several strategies involving different mechanisms of antisense inhibition of protein synthesis by a variety of oligonucleotide analogs have been tried in bacteria with mixed results [9–13]. A promising one, known as EGS technology, consists of developing antisense oligomers that inhibit gene expression of bacterial genes by eliciting RNase P-mediated degradation of a target mRNA [11, 14–19]. RNase P plays roles in the processing of RNA molecules such as tRNAs, 4.5S RNA, transfer messenger RNA, some multicistronic mRNAs, phage-related RNAs, and others [20]. The *Escherichia coli* RNase P consists of M1, a 377-nucleotide catalytic RNA subunit, and C5, an 119-amino acid protein that acts as a cofactor [20–22]. These components are coded for by two genes, *rnpB* and *rnpA*, respectively. The utilization of RNase P for an antisense strategy is derived from the characteristics of this enzyme to cleave any RNA molecule (the target) as long as when interacting with an antisense oligoribonucleotide (external guide sequence, EGS) forms a structure similar to that of the pre-tRNA at the cleavage location [14, 20, 23]. Since oligoribonucleotides are very unstable, therapeutic development of EGS technology requires the design of stable, nuclease-resistant oligoribonucleotide analogs. The role of oligonucleotide analogs in all fields of antisense technologies is critical, and as a consequence, research to develop new and more effective analogs is intense [24, 25]. New compounds are used as only components of oligomers or in combination with other analogs, increasing the number of potential drug candidates. Previous studies showed that chimeric compounds composed of locked nucleic acid and deoxyribonucleotide residues with gapmer configurations show EGS activity [17, 19, 26, 27]. These findings make EGS technology a viable antisense approach that could result in the development of a new generation of antimicrobials, which are urgently needed to respond to the current multidrug resistance crisis [11, 28–30]. However, as new analogs are developed, there will be a need to test new mono- and multicomponent oligomers with different configurations. Key milestones to select a potent EGS are the determination of its efficiency to elicit cleavage of the target RNA molecule and their ability to exert the biological activity on cells in culture. In this article we describe a protocol first developed in Altman's laboratory [31] to assess the activity of

EGSs on specific target RNA molecules. We also describe a protocol to assess EGSs activity on the growth of bacterial cells when added to growth medium. We have recently used these protocols to select EGSs that interfere with the expression of the resistance genes *aac(6′)-Ib* and *cat* [17, 32–35], and *ftsZ*, which codes for a protein that is the scaffold for the divisome and generates the constrictive force to initiate cell division [36].

2 Materials

All solutions should be prepared using diethyl pyrocarbonate (DEPC)-treated ultrapure water and RNase-free analytical grade reagents. Prepare reagents at room temperature and store as indicated.

2.1 Components for the RNase P In Vitro Assay

1. M1 RNA and mRNA synthesis: MEGAscript high-yield transcription T7 kit (Life Technologies).
2. 5′ end labeling of mRNA: 5′ EndTag Nucleic Acid Labeling system (Vector Laboratories). Labeling dye: fluorescent dye cyanine 5 maleimide (Lumiprobe).
3. Terrific broth: 12 g/L tryptone, 24 g/L yeast extract, 9.4 g/L potassium phosphate dibasic, 2.2 g/L potassium phosphate monobasic, pH 7.2 [37].
4. Inducer: 100 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Prepare the IPTG stock solution in water and keep aliquoted at −20 °C.
5. Buffer 1: 50 mM Tris–HCl (pH 7.5), 60 mM NH_4Cl, 10 mM magnesium acetate, 0.15% dithiothreitol, 42% urea, 0.2 mM Pefabloc protease inhibitor.
6. Buffer 2: 50 mM Tris–HCl (pH 7.5), 1 M NH_4Cl, 10 mM magnesium acetate, and 0.15% dithiothreitol.
7. Buffer 3: 50 mM Tris–HCl (pH 7.5), 100 mM NH_4Cl, 10 mM $MgCl_2$, and 0.15% dithiothreitol.
8. Buffer 4: 0.05 M sodium acetate (pH 7.2), 0.01 M $MgCl_2$, 7 M urea, and 0.15% dithiothreitol.
9. DNase I stock solution: Prepare at 2000 U/mL concentration and keep at −20 °C.
10. NaCl 0.5 M.

2.2 RNase P In Vitro cleavage assay

1. Reaction buffer: 20 mM HEPES-KOH (pH 8.0), 400 mM ammonium acetate, 10 mM magnesium acetate, and 5% glycerol.
2. Phenol-chloroform (5:1 v/v, pH 4.5).

3. Chloroform: Water saturated chloroform-isoamyl alcohol (24:1).
4. 3 M Sodium acetate (pH 5.5).
5. Ethanol (absolute).
6. EGSs: Dissolve oligonucleotides and oligonucleotide analogs at 100 μM concentration and keep at −20 °C. The appropriate dilution should be prepared just before using each EGS.
7. Reagent Solution 1: 5 pmols of 5′-end-labeled substrate mRNA and 10 pmols of EGS are dissolved in a total volume of 3 μL.
8. Reagent Solution 2: 2.5 pmol of M1 RNA, 70 pmols of C5 protein, 20 mM HEPES-KOH (pH 8.0), 400 mM ammonium acetate, and 10 mM magnesium acetate dissolved in 5% glycerol (v/v) in a total volume of 7 μL (*see* **Note 1**).
9. Denaturing polyacrylamide gel: 5% polyacrylamide gels are prepared containing 16:1 (acrylamide-bis-acrylamide), 7 M urea, 89 mM Tris, 29 mM taurine, and 0.5 mM EDTA (USB Corp.).
10. 2× gel loading buffer: 95% formamide, 1 mM EDTA, pH 8, 0.01% Bromophenol Blue (w/v).
11. Molecular weight standard: Fluorescence-labeled RNA Century Marker or RNA Century Marker-Plus (ThermoFisher).
12. Fluorescence detection: Fluorescence was detected on a Storm 860 Molecular Imager (Molecular Dynamics).

2.3 Bacterial Growth Inhibition Assay

1. Bacterial strains are maintained in 20% glycerol at −80 °C.
2. CPP-EGSs: oligonucleotides and oligonucleotide analogs covalently linked to the $(RXR)_4XB$ (where "X" and "B" stand for 6-aminohexanoic acid and β-alanine, respectively) were purchased from BioSynthesis Inc.
3. Mueller-Hinton broth: 2 g/L beef extract, 17.5 g/L casein hydrolysate, 1.5 g/L starch [38].
4. Amikacin: Prepare stock solution at 10 mg/mL in water (RNase-free is not needed). Keep at 4 °C.
5. Ampicillin: Prepare stock solution at 100 mg/mL in water (RNase-free is not needed). Keep at 4 °C.
6. Chloramphenicol: Prepare stock solution at 34 mg/mL in ethanol and keep at 4 °C.
7. Microtiter plates (100 μL).
8. BioTek Synergy 2 microplate reader.

3 Methods

3.1 Expression and Purification of the RNase P Cofactor C5

1. *E. coli* BL21(DE3)(pLysE, pRHC5) cells were cultured in Terrific broth containing ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL) at 37 °C. When the cells reached OD_{600} 0.6–0.8, 1 mM IPTG was added to trigger expression of the C5 protein (13.8 kD) and the culture was continued overnight at the same temperature.
2. After harvesting by centrifugation, the cells were resuspended in 0.02 volume of Buffer 1, lysed by sonication, and treated with 6 μL/mL DNase I 2000 U/mL for 30 min at 4 °C. After removal of cell debris by centrifugation (7700 × *g*) for 10 min at 4 °C, the soluble extract was centrifuged at 30,000 × *g* for 30 min at the same temperature. The supernatant was harvested and centrifuged again for 2 h at 100,000 × *g* at 4 °C. The pellet was resuspended in Buffer 2 (10 mL), shaken for 2 h at 4 °C, centrifuged at 100,000 × *g* for 2 h (4 °C), and the supernatant was dialyzed overnight against Buffer 3 at 20 °C. After dialysis, the supernatant was centrifuged at 30,000 × *g* for 30 min at 4 °C to collect the protein that tends to precipitate during the process. The pellet containing the C5 protein was resuspended in Buffer 4.
3. The C5 protein was then purified on a Sephadex C50 (Amersham Biosciences) column. The ion-exchange column chromatography was eluted with a linear gradient 0–0.5 M NaCl in Buffer 3 and the C5 protein eluted at 0.3 M NaCl.

3.2 In Vitro RNase P cleavage assay

The in vitro reaction has been modified from a protocol previously described [31].

1. Reagent Solutions 1 (3 μL) and 2 (7 μL) were incubated at 25 °C for 30 min and at 37 °C for 15 min, respectively. Both solutions were then combined and incubated at 37 °C for the times required for each experiment (*see* **Note 2**).
2. The reaction was stopped by addition of 40 μL water and 50 μL phenol-chloroform (5:1 v/v, pH 4.5) followed by chloroform extraction of the aqueous phase and precipitation adding 0.1 volume of 3 M sodium acetate and 2 volume ethanol. The pellet was resuspended in 10 μL water.
3. The reaction products were analyzed using denaturing polyacrylamide gel electrophoresis. The sample was mixed with 10 μL of 2× gel loading buffer before loading. Molecular weight standard were fluorescence-labeled RNA Century Marker or RNA Century Marker-Plus (ThermoFisher). Fluorescence was detected on a Storm 860 Molecular Imager (Molecular Dynamics) and when needed the bands were quantified using Image J [39]. A typical in vitro RNAse P cleavage assay is shown in Fig. 1.

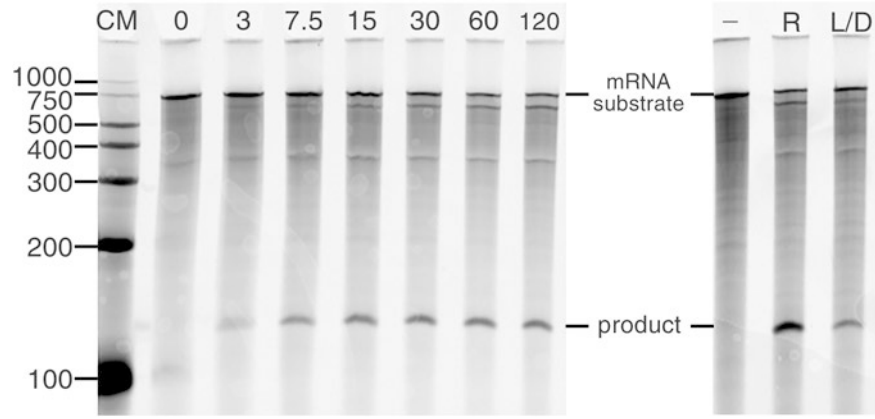


Fig. 1 In vitro RNase P cleavage. The panel to the left shows a time course RNase P digestion of *cat* mRNA in the presence of an RNA EGS. Incubation time in minutes is shown on top. CM: RNA Century markers (nucleotides). The panel to the right shows a comparison of RNase P digestion in the absence (–) or presence of an RNA (R) or a chimeric locked nucleic acid/DNA (L/D) gapmer EGS. In the chimeric gapmer, the five residues at the ends are locked nucleic acids and the seven central residues are deoxyribonucleotides

3.3 Assessment of the Activity of Cell-Penetrating Peptide-Bound EGSs on Growing Bacterial Cells

Intracellular uptake is a critical requirement for successful use of antisense compounds as therapeutic agents. Cell-penetrating peptides (CPP) are known to be able to deliver different covalently bound compounds inside cells [40]. Oligonucleotide analogs of different nature are transported inside bacterial cells when conjugated to CPPs [17, 26, 41–43]. The methodology described in the following paragraphs was applied to chimeric oligonucleotides composed of locked nucleic acid residues at the ends and deoxyribonucleotides in the center (gapmers) covalently linked to the $(RXR)_4XB$ (where "X" and "B" stand for 6-aminohexanoic acid and β-alanine, respectively) (CPP-EGS) [26].

1. Strains used to test the inhibitory action of CPP-EGSs were taken from the stock at −80 °C, plated on Mueller-Hinton agar containing amikacin (10 or 20 μg/mL for strains harboring low or high copy number plasmids, respectively), and incubated overnight at 37 °C (*see* **Note 3**). Cells from an isolated colony were used to inoculate 5 mL Mueller-Hinton broth containing amikacin followed by incubation overnight at 37 °C.
2. The overnight culture was diluted 1:50 in Mueller-Hinton broth with the necessary additions to start the assay. As in the previous cultures, the concentration of amikacin is dependent on the copy number of the resistance gene. Concentrations of the CPP-EGS must be empirically determined.
3. The microtiter plates were incubated at 37 °C with shaking, and OD_{600} measurements were taken every 20 min. Comparison

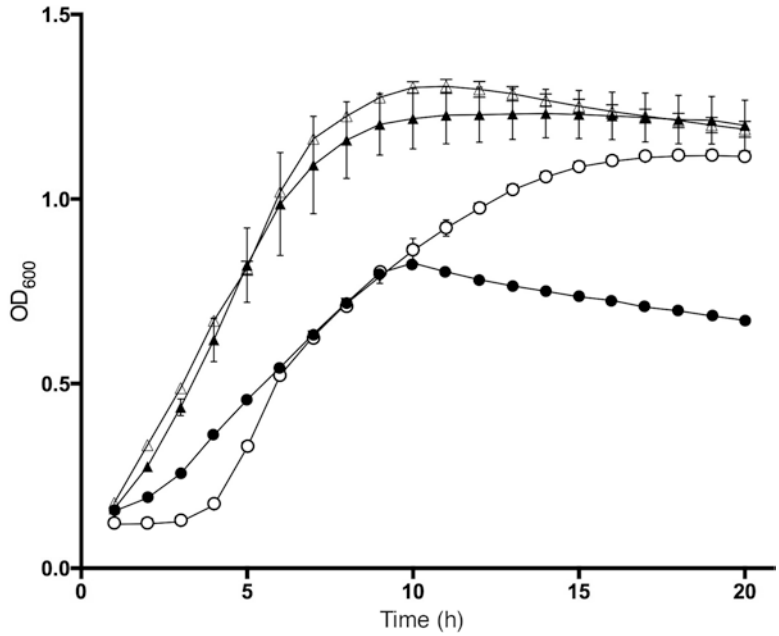


Fig. 2 Effect of an EGS on resistance to amikacin. A CPP-EGS [$(RXR)_4$XB-Cys-SMCC-C6 amino-CGATATGAGATCGACCA] (R, arginine; X, 6-aminohexanoic acid; B, beta-alanine; LNA residues are shown underlined) [26] targeting the *aac(6′)-Ib* mRNA was tested as inhibitor of expression of resistance to amikacin. A clinical *Acinetobacter baumannii* isolate (strain A155) [44] was cultured in microtiter plates at 37 °C with the following additions: filled circles, 16 μg/mL amikacin and 5 μM CPP-EGS; empty circles, 16 μg/mL Amikacin; filled triangles, 5 μM CPP-EGS; empty triangles, none. Values are the average and standard deviation of duplicates

of the growth curves permitted to immediately identify active CPP-EGSs. Fig. 2 shows a typical experiment in which a CPP-EGS targeting the amikacin resistance gene *aac(6′)-Ib* induces an increase in susceptibility.

4 Notes

1. The relative amounts of C5 and M1 may have to be adjusted every time a new C5 protein purification is carried out.
2. For best results, once the substrate mRNA has been labeled with the fluorescent dye cyanine5 maleimide (Lumiprobe), all procedures should be carried out in the dark.
3. It is important that all experiments *in cellulo* are carried with cells that have been removed from the stock and cultured as indicated. Variations in the age of the cultures, the composition of the growth medium, or the additions such as different concentrations of antibiotics affect the reproducibility of the assays.

Acknowledgments

This work was supported by Public Health Service grant 2R15AI047115-04 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. A.J. was funded in part by LA Basin Minority Health and Health Disparities International Research Training Program (MHIRT) 5T37MD001368.

References

1. McClorey G, Wood MJ (2015) An overview of the clinical application of antisense oligonucleotides for RNA-targeting therapies. Curr Opin Pharmacol 24:52–58. https://doi.org/10.1016/j.coph.2015.07.005
2. Marwick C (1998) First "antisense" drug will treat CMV retinitis. JAMA 280(10):871
3. Ricotta DN, Frishman W (2012) Mipomersen: a safe and effective antisense therapy adjunct to statins in patients with hypercholesterolemia. Cardiol Rev 20(2):90–95. https://doi.org/10.1097/CRD.0b013e31823424be
4. Lim KR, Maruyama R, Yokota T (2017) Eteplirsen in the treatment of Duchenne muscular dystrophy. Drug Des Devel Ther 11:533–545. https://doi.org/10.2147/DDDT.S97635
5. Chiriboga CA, Swoboda KJ, Darras BT, Iannaccone ST, Montes J, De Vivo DC, Norris DA, Bennett CF, Bishop KM (2016) Results from a phase 1 study of nusinersen (ISIS-SMN(Rx)) in children with spinal muscular atrophy. Neurology 86(10):890–897. https://doi.org/10.1212/WNL.0000000000002445
6. Burnett JC, Rossi JJ (2012) RNA-based therapeutics: current progress and future prospects. Chem Biol 19(1):60–71. https://doi.org/10.1016/j.chembiol.2011.12.008
7. Sridharan K, Gogtay NJ (2016) Therapeutic nucleic acids: current clinical status. Br J Clin Pharmacol 82(3):659–672. https://doi.org/10.1111/bcp.12987
8. Kole R, Krainer AR, Altman S (2012) RNA therapeutics: beyond RNA interference and antisense oligonucleotides. Nat Rev Drug Discov 11(2):125–140. https://doi.org/10.1038/nrd3625
9. Rasmussen LC, Sperling-Petersen HU, Mortensen KK (2007) Hitting bacteria at the heart of the central dogma: sequence-specific inhibition. Microb Cell Fact 6:24. https://doi.org/10.1186/1475-2859-6-24
10. Sarno R, Ha H, Weinsetel N, Tolmasky ME (2003) Inhibition of aminoglycoside 6′-*N*-acetyltransferase type Ib-mediated amikacin resistance by antisense oligodeoxynucleotides. Antimicrob Agents Chemother 47(10):3296–3304
11. Davies-Sala C, Soler-Bistue A, Bonomo RA, Zorreguieta A, Tolmasky ME (2015) External guide sequence technology: a path to development of novel antimicrobial therapeutics. Ann N Y Acad Sci 1354:98–110. https://doi.org/10.1111/nyas.12755
12. Woodford N, Wareham DW (2009) Tackling antibiotic resistance: a dose of common antisense? J Antimicrob Chemother 63(2):225–229. https://doi.org/10.1093/jac/dkn467
13. Matzov D, Bashan A, Yonath A (2017) A bright future for antibiotics? Annu Rev Biochem 86:567–583. https://doi.org/10.1146/annurev-biochem-061516-044617
14. Lundblad EW, Altman S (2010) Inhibition of gene expression by RNase P. N Biotechnol 27(3):212–221. https://doi.org/10.1016/j.nbt.2010.03.003
15. Davies Sala C, Soler-Bistue AJ, Korprapun L, Zorreguieta A, Tolmasky ME (2012) Inhibition of cell division induced by external guide sequences (EGS technology) targeting *ftsZ*. PLoS One 7(10):e47690. https://doi.org/10.1371/journal.pone.0047690
16. Guerrier-Takada C, Salavati R, Altman S (1997) Phenotypic conversion of drug-resistant bacteria to drug sensitivity. Proc Natl Acad Sci U S A 94(16):8468–8472
17. Shen N, Ko JH, Xiao G, Wesolowski D, Shan G, Geller B, Izadjoo M, Altman S (2009) Inactivation of expression of several genes in a variety of bacterial species by EGS technology. Proc Natl Acad Sci U S A 106(20):8163–8168. https://doi.org/10.1073/pnas.0903491106
18. Soler Bistue AJ, Ha H, Sarno R, Don M, Zorreguieta A, Tolmasky ME (2007) External guide sequences targeting the *aac(6′)-Ib* mRNA induce inhibition of amikacin resistance. Antimicrob Agents Chemother

51(6):1918–1925. https://doi.org/10.1128/AAC.01500-06

19. Soler Bistue AJ, Martin FA, Vozza N, Ha H, Joaquin JC, Zorreguieta A, Tolmasky ME (2009) Inhibition of *aac(6′)-Ib*-mediated amikacin resistance by nuclease-resistant external guide sequences in bacteria. Proc Natl Acad Sci U S A 106(32):13230–13235. https://doi.org/10.1073/pnas.0906529106
20. Altman S (2011) Ribonuclease P. Philos Trans R Soc Lond B Biol Sci 366(1580):2936–2941. https://doi.org/10.1098/rstb.2011.0142
21. Forster AC, Altman S (1990) External guide sequences for an RNA enzyme. Science 249(4970):783–786
22. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35(3 Pt 2):849–857
23. Gopalan V, Vioque A, Altman S (2002) RNase P: variations and uses. J Biol Chem 277(9):6759–6762. https://doi.org/10.1074/jbc.R100067200
24. Deleavey GF, Damha MJ (2012) Designing chemically modified oligonucleotides for targeted gene silencing. Chem Biol 19(8):937–954. https://doi.org/10.1016/j.chembiol.2012.07.011
25. Kurreck J (2003) Antisense technologies. Improvement through novel chemical modifications. Eur J Biochem 270(8):1628–1644
26. Jackson A, Jani S, Sala CD, Soler-Bistue AJ, Zorreguieta A, Tolmasky ME (2016) Assessment of configurations and chemistries of bridged nucleic acids-containing oligomers as external guide sequences: a methodology for inhibition of expression of antibiotic resistance genes. Biol Methods Protoc 1(1):bpw001. https://doi.org/10.1093/biomethods/bpw001
27. Sawyer AJ, Wesolowski D, Gandotra N, Stojadinovic A, Izadjoo M, Altman S, Kyriakides TR (2013) A peptide-morpholino oligomer conjugate targeting *Staphylococcus aureus gyrA* mRNA improves healing in an infected mouse cutaneous wound model. Int J Pharm 453(2):651–655. https://doi.org/10.1016/j.ijpharm.2013.05.041
28. Lin J, Nishino K, Roberts MC, Tolmasky M, Aminov RI, Zhang L (2015) Mechanisms of antibiotic resistance. Front Microbiol 6:34. https://doi.org/10.3389/fmicb.2015.00034
29. Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B, Bartlett J (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. Clin Infect Dis 48(1):1–12. https://doi.org/10.1086/595011
30. Ramirez MS, Traglia GM, Lin D, Tran T, Tolmasky ME (2014) Plasmid-mediated antibiotic resistance and virulence in Gram-negatives: the *Klebsiella pneumoniae* paradigm. Microbiol Spectr 2(5):1
31. Li Y, Guerrier-Takada C, Altman S (1992) Targeted cleavage of mRNA in vitro by RNase P from *Escherichia coli*. Proc Natl Acad Sci U S A 89(8):3185–3189
32. Tolmasky ME (2007) Aminoglycoside-modifying enzymes: characteristics, localization, and dissemination. In: Bonomo R, Tolmasky ME (eds) Enzyme-mediated resistance to antibiotics: mechanisms, dissemination, and prospects for inhibition. ASM Press, Washington, DC, pp 35–52
33. Ramirez MS, Nikolaidis N, Tolmasky ME (2013) Rise and dissemination of aminoglycoside resistance: the *aac(6′)-Ib* paradigm. Front Microbiol 4:121. https://doi.org/10.3389/fmicb.2013.00121
34. Tolmasky ME, Chamorro RM, Crosa JH, Marini PM (1988) Transposon-mediated amikacin resistance in *Klebsiella pneumoniae*. Antimicrob Agents Chemother 32(9):1416–1420
35. Ramirez MS, Tolmasky ME (2010) Aminoglycoside modifying enzymes. Drug Resist Updat 13(6):151–171. https://doi.org/10.1016/j.drup.2010.08.003
36. Mingorance J, Rivas G, Velez M, Gomez-Puertas P, Vicente M (2010) Strong FtsZ is with the force: mechanisms to constrict bacteria. Trends Microbiol 18(8):348–356. https://doi.org/10.1016/j.tim.2010.06.001
37. Ukkonen K, Vasala A, Ojamo H, Neubauer P (2011) High-yield production of biologically active recombinant protein in shake flask culture by combination of enzyme-based glucose delivery and increased oxygen transfer. Microb Cell Fact 10:107. https://doi.org/10.1186/1475-2859-10-107
38. Barry AL, Reller LB, Miller GH, Washington JA, Schoenknect FD, Peterson LR, Hare RS, Knapp C (1992) Revision of standards for adjusting the cation content of Mueller-Hinton broth for testing susceptibility of *Pseudomonas aeruginosa* to aminoglycosides. J Clin Microbiol 30(3):585–589
39. Abramoff M, Magelhaes P, Ram S (2004) Image processing with Image J. J Biophotonics 11:36–42
40. Copolovici DM, Langel K, Eriste E, Langel U (2014) Cell-penetrating peptides: design, synthesis, and applications. ACS Nano 8(3):1972–1994. https://doi.org/10.1021/nn4057269
41. Lehto T, Ezzat K, Wood MJ, El Andaloussi S (2016) Peptides for nucleic acid delivery. Adv

Drug Del Rev 106(Pt A):172–182. https://doi.org/10.1016/j.addr.2016.06.008

42. Reissmann S (2014) Cell penetration: scope and limitations by the application of cell-penetrating peptides. J Pept Sci 20(10):760–784. https://doi.org/10.1002/psc.2672
43. Lopez C, Arivett BA, Actis LA, Tolmasky ME (2015) Inhibition of AAC(6′)-Ib-mediated resistance to amikacin in *Acinetobacter baumannii* by an antisense peptide-conjugated 2′,4′-bridged nucleic acid-NC-DNA hybrid oligomer. Antimicrob Agents Chemother 59(9):5798–5803. https://doi.org/10.1128/AAC.01304-15
44. Arivett BA, Fiester SE, Ream DC, Centron D, Ramirez MS, Tolmasky ME, Actis LA (2015) Draft genome of the multidrug-resistant *Acinetobacter baumannii* strain A155 clinical isolate. Genome Announc 3(2):e00212–e00215. https://doi.org/10.1128/genomeA.00212-15

Chapter 7

Evaluating the Effect of Small RNAs and Associated Chaperones on Rho-Dependent Termination of Transcription In Vitro

Cédric Nadiras, Annie Schwartz, Mildred Delaleau, and Marc Boudvillain

Abstract

Besides their well-known posttranscriptional effects on mRNA translation and decay, sRNAs and associated RNA chaperones (e.g., Hfq, CsrA) sometimes regulate gene expression at the transcriptional level. In this case, the sRNA-dependent machinery modulates the activity of the transcription termination factor Rho, a ring-shaped RNA translocase/helicase that dissociates transcription elongation complexes at specific loci of the bacterial genome. Here, we describe biochemical assays to detect Rho-dependent termination signals in genomic regions of interest and to assess the effects of sRNAs and/or associated RNA chaperones on such signals.

Key words Rho, Transcription, Termination, Ring-shaped, Helicase, Hexamer, sRNA, Chaperone

1 Introduction

In bacteria, transcription termination is triggered by two main types of signals having distinct nucleic acid (NA) and protein cofactor requirements (reviewed in: [1, 2]). Intrinsic (Rho-independent) terminators are encoded by a NA signal that is usually sufficient to destabilize the transcription elongation complex. Intrinsic terminators are relatively easy to detect in bacterial genomes as they most often involve the formation of a GC-rich hairpin followed by a run of U residues at the 3′ end of the nascent RNA transcript. By contrast, Rho-dependent terminators are characterized by a strict requirement for the protein factor Rho (Fig. 1) but have relaxed sequence determinants rendering them difficult to detect in genome sequences [1].

Transcription termination signals are sometimes found in the 5′-untranslated regions (5′-UTRs) of genes or operons where they contribute to the regulation of the gene/operon through conditional attenuation of transcription [3, 4]. Attenuation mechanisms involving intrinsic terminators are diverse and wide-

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_7,

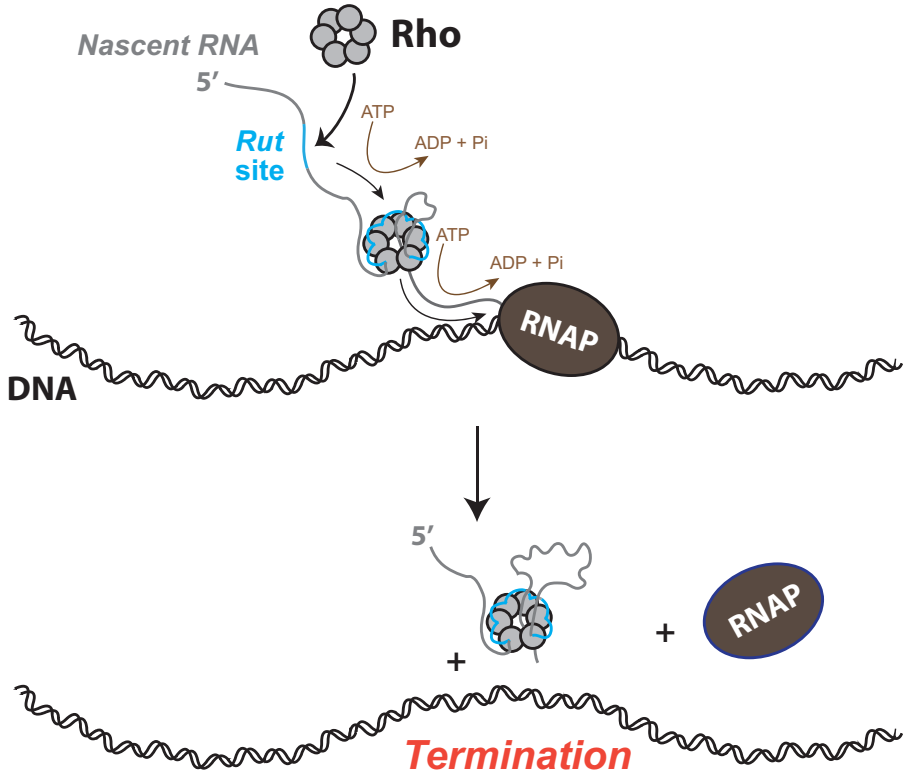


Fig. 1 Rho-dependent termination of transcription. The Rho hexamer binds a *Rut* site within a naked (i.e., untranslated) portion of the nascent transcript. Rho anchoring to the transcript activates the ATP-dependent translocation of RNA within the hexamer central channel (note that the initial interaction with the *Rut* site is maintained throughout) [27]. Then, Rho triggers dissociation of the transcription elongation complex upon catching up with the RNA polymerase (RNAP). Various mechanisms whereby sRNAs and RNA chaperones regulate Rho-dependent termination of transcription have been described (not shown on the figure). These mechanisms include formation of an Hfq-dependent antitermination complex [17], modulation of Rho access to a *Rut* site through inhibition of translation [14] or structural remodeling of the nascent RNA by CsrA [11] as well as less clear modes of structural interference by sRNA:mRNA complexes [10, 15]

spread while, until recently, known instances of comparable Rho-dependent mechanisms were few [5–7]. This situation is, however, quickly evolving with the discovery of riboswitches governing Rho-dependent termination [8, 9] and, more to the point of the present chapter, of conditional Rho-dependent mechanisms governed by sRNAs and/or RNA chaperones. For instance, a Rho-dependent terminator has been identified recently in the 5′-UTR of the *rpoS* gene of *Escherichia coli* where it is regulated negatively by the binding of sRNA DsrA, ArcZ, or RprA to the *rpoS* mRNA leader [10]. A conditional Rho-dependent terminator operating by a distinct mechanism has also been identified recently in the 5′-UTR of the *pgaABCD* operon of *E. coli* [11]. In this case, remodeling of the mRNA leader by the CsrA chaperone allows Rho access to a binding *Rut* (*R*ho *ut*ilization) site that is otherwise

sequestered in a RNA secondary structure [11]. This mechanism contributes to the multilayered regulation of *pgaABCD* by CsrA, which notably includes scavenging of CsrA by CsrB, CsrC, and McaS acting as sRNA sponges [12, 13].

Conditional Rho-dependent termination can also be triggered by the binding of a sRNA to its mRNA target. In the first case described, binding of the sRNA ChiX to the 5′-leader of the *chiPQ* operon of *Salmonella* prevents *chiP* translation [14]. This in turn exposes an intragenic *Rut* site within the *chiP* mRNA that is otherwise hidden by translating ribosomes, and triggers Rho-dependent transcriptional polarity within the operon [14]. Similarly, the sRNA Spf (aka Spot 42) triggers Rho-dependent termination at the end of the *galT* gene within the *galETKM* operon of *E. coli* [15]. These two known cases suggest that the activation of Rho-dependent transcriptional polarity by sRNAs could be a general mechanism contributing to gene/operon silencing. When exploring this possibility, however, one needs to consider the potential contribution of accessory factors. For instance, the Hfq chaperone can inhibit Rho-dependent termination in a manner that is antagonized by NusG—an essential, multi-role transcription factor [16] or by NA ligands that alter Hfq interaction with Rho and RNA [17]. Moreover, some Rho-dependent terminators are effective only in the presence of NusG, probably because their *Rut* sites are too weak to recruit or activate Rho by themselves [14, 18, 19].

Here, we describe in vitro methods to probe whether transcription termination factor Rho is involved in sRNA-mediated regulation of a given bacterial gene or operon. Together with in vivo probing, these methods should prove useful to unravel the potentially complex interplays existing between Rho, sRNAs, and accessory factors such as RNA chaperones.

2 Materials

2.1 Preparation of DNA Templates for Transcription Termination Assays

1. 100 μM stock solutions of DNA primers FWD, REV, and T7A1 (Table 1).
2. Genomic DNA or frozen stock of bacterial strain of interest (*see* **Note 1**).
3. Stock solution of dNTPs (10 mM each).
4. 2 U/μLVent DNA polymerase (*see* **Note 2**).
5. 10× Vent reaction buffer: 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 20 mM $MgSO_4$, 1% Triton X-100, 200 mM Tris–HCl, pH 8.8.
6. PCR thermocycler equipped with a heated lid and a thermal block for 0.5 mL microtubes.
7. PCR purification kit (*see* **Note 3**).
8. Benchtop centrifuge.

Table 1
Sequences of the oligonucleotide primers[a]

Template	Primer	Sequence[b]
rpoS[c]	$FWD_{(rpoS)}$	5′-GTCTAACCTATAGGATACTTACAGCCCTTCGGGTGAACAGAGTGCTAACAAAATG
	$REV_{(rpoS)}$	5′-ATAACAGCTCTTCTTCAGCCAGGTCGTTATCAC
	T7A1[d]	5′-TTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCC
DsrA[e]	$TOP_{(DsrA)}$	5′-TAATACGACTCACTATAGGAACACATCAGATTTCCTGGTGTAACG
	$BOTT_{(DsrA)}$	5′-AAATCCCGACCCTGAGGGGGTCGGGATGAAAC
ArcZ[e]	$TOP_{(ArcZ)}$	5′-TAATACGACTCACTATAGGGGTGCGGCCTG AAAAACAGTG CTGTGCCCTT G
	$BOTT_{(ArcZ)}$	5′-AAAAAATGACCCCGGCTAGACCGGGGTGCGC
FnrS[e]	$TOP_{(FnrS)}$	5′-TAATACGACTCACTATAGGCAGGTGAATGCAACGTCAAGCGATGGGC
	$BOTT_{(FnrS)}$	5′-AAAAAGCCGACTCATCAAAGTCGGCGTCGTACGAATCAATTGTGCTATG
Spot 42[e] (Spf)	$TOP_{(Spf)}$	5′-TAATACGACTCACTATAGGTAGGGTACAGAGGTAAGATGTTCTATCTTTC
	$BOTT_{(Spf)}$	5′-TAAAAAACGCCCCAGTCATTACTGACTGGGGCGG

[a]Primers used to study sRNA effects on Rho-dependent termination in the 5′-leader region of *rpoS*
[b]Promoter sequences are underlined. Other oligonucleotide regions are chosen based on which genomic region needs to be amplified (Fig. 2a)
[c]For transcription termination assays with RNA polymerase from *E. coli*
[d]T7A1 is a universal primer used only in the second round of PCR amplification
[e]For preparation of indicated sRNA by transcription with T7 RNA polymerase

9. 1× $T_{10}E_1$ buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 7.5 (*see* **Note 4**).
10. Agarose.
11. Commercial DNA ladder (e.g., #N3233S ladder from New England Biolabs).
12. 6× agarose gel loading buffer: 15% Ficoll-400, 0.1% SDS, 0.1% bromophenol blue, 20 mM Tris–HCl, 66 mM EDTA, pH 8.0.
13. 20× TAE buffer: dissolve 98 g of Tris base in approximately 800 mL of water, then add 22.8 mL of glacial acetic acid and 40 mL of 0.5 M EDTA pH 8.0, and add water to obtain a final volume of 1 L.
14. 1× TAE buffer: obtained by dilution of the 20× TAE stock buffer.
15. Horizontal gel electrophoresis system and power supply.
16. 10 mg/mL ethidium bromide stock solution (*see* **Note 5**).
17. Microwave oven.
18. UV transilluminator or dedicated gel documentation system.
19. Sephadex G-50 spin columns (e.g., Microspin columns from GE Healthcare).
20. UV spectrophotometer suitable for micro-volumes measurements (e.g., Nanodrop 2000c from Thermo Scientific).

2.2 Preparation of sRNAs

1. Items 2–20 of Subheading 2.1.
2. 100 μM stock solutions of DNA primers TOP and BOTT (Table 1).
3. 5× Transcription buffer: 0.12 M $MgCl_2$, 0.4 M HEPES, pH 7.5, 0.1 M DTT, 0.05% Triton X-100, and 5 mM Spermidine (*see* **Note 4**).
4. 50 U/μL T7 RNA polymerase.
5. 1 U/μL RQ1 DNase (Promega).
6. Dry bath incubator with shaking capability.
7. 20 U/μL SUPERase-IN™ (Thermo Fisher Scientific).
8. Deionized RNA-grade water (*see* **Note 4**).
9. Set of high-grade rNTPs (100 mM each).
10. 0.5 M EDTA solution, adjusted to pH 7.5 with NaOH (*see* **Note 4**).
11. Phenol:Chloroform:Isoamyl alcohol (25:24:1) mix, pH 6.7.
12. Diethyl ether.

13. 3 M sodium acetate (NaAc), adjusted to pH 6.5 with acetic acid (*see* **Note 4**).
14. 20× TBE buffer: dissolve 216 g of Tris base, 110 g of boric acid, 14.9 g EDTA in 1 L of water. Filter on Whatman paper and store at room temperature.
15. 1× TBE buffer: obtained by dilution of the 20× TBE stock.
16. 40% acrylamide:bis-acrylamide [29:1 ratio] commercial stock solution.
17. Denaturing acrylamide solution for sRNA preparation (6% acrylamide:bis-acrylamide [29:1 ratio] and 7 M urea in 1× TBE buffer). Mix 12.6 g of urea, 4.5 mL of 40% acrylamide:bis-acrylamide [29:1 ratio] solution, 1.5 mL 20× TBE, and 10 mL of deionized water. Heat the solution to dissolve urea completely. Adjust volume to 30 mL with deionized water and cool down to room temperature. Prepare fresh solution before use.
18. N,N,N,N′-tetramethyl-ethylenediamine (TEMED).
19. 25% (w/v) ammonium persulfate (APS) in water.
20. Vertical electrophoresis system and power supply for DNA sequencing (e.g., adjustable 20 × 42 cm sequencing kit from CBS scientific).
21. Denaturing loading buffer: 95% formamide, 5 mM EDTA, 0.01% (w/v) bromophenol blue.
22. X-ray intensifying screen or a fluor-coated TLC plate.
23. Hand-held 254 nm UV lamp.
24. 1× Elution buffer: 0.3 M NaAc, 10 mM MOPS, 1 mM EDTA, pH 6.0.
25. 1× $M_{10}E_1$ buffer: 10 mM MOPS, 1 mM EDTA, pH 6.0 (*see* **Note 4**).

2.3 Transcription Termination Assay

1. Items 6–21 of Subheading 2.2.
2. “Protein low binding” and “DNA low binding” 1.5 mL microtubes.
3. Commercial DNA ladder. Select a ladder made of DNA fragments that are not phosphorylated at 5′-ends (e.g., #N3233S ladder from New England Biolabs).
4. 10 U/μL T4 polynucleotide kinase.
5. 10× PNK buffer: 100 mM $MgCl_2$, 50 mM DTT, and 700 mM Tris–HCl, pH 7.6.
6. γ-^{32}P ATP at 3000 Ci/mmol [10 mCi/mL].

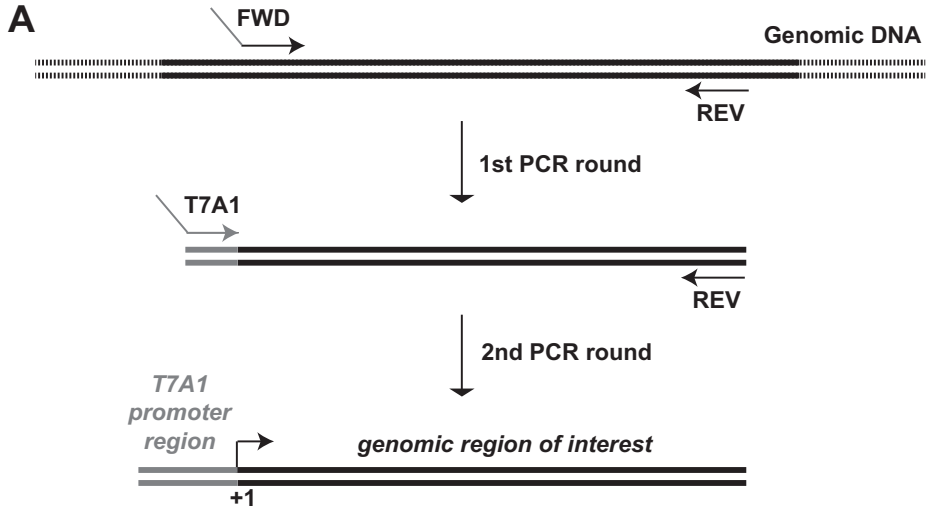
7. Sephadex G-50 spin columns (e.g., Microspin columns from GE Healthcare).
8. 1.4 μM Rho stock solution (*see* **Note 6**) in Rho storage buffer (50% glycerol, 100 mM KCl, 0.1 mM EDTA, 0.1 mM DTT, 10 mM Tris–HCl, pH 7.9). Preparation of the Rho protein from *E. coli* is detailed in the volume 587 of *Methods in Molecular Biology* [20] (*see* **Note 7**).
9. 2 μM NusG stock solution (*see* **Note 8**).
10. 2 μM stock solution of CsrA or Hfq chaperone (optional) (*see* **Note 8**).
11. 1 U/μL *E. coli* RNA Polymerase, Holoenzyme (New England Biolabs).
12. 5 μM stocks of sRNAs (as prepared in Subheading 3.2). Stocks should include the sRNAs under investigation (i.e., the ones expected to pair with the mRNA target) and at least one negative control (sRNA not expected to bind to the mRNA target). We also like to include a shorter, synthetic oligoribonucleotide that is fully complementary to the mRNA sequence targeted by the sRNA(s) and serves as positive control.
13. 5× transcription termination buffer: 250 mM KCl, 25 mM $MgCl_2$, 7.5 mM DTT, 0.25 mg/mL bovine serum albumin, and 200 mM Tris–HCl, pH 8.0 (*see* **Note 4**).
14. 10× initiation mixture: 2 mM ATP, 2 mM GTP, 2 mM CTP, 0.2 mM UTP, 250 μg/mL rifampicin, and 2 μCi/μL ^{32}P-αUTP in 1× transcription termination buffer. Prepare right before use (*see* **Note 9**).
15. 0.25 mg/mL tRNA stock.
16. 1× resuspension buffer: mix 460 μL of $M_{10}E_1$ buffer with 40 μL of 0.5 M EDTA, pH 7.5.
17. Denaturing acrylamide solution (termination assay): 7% acrylamide:bis-acrylamide [19:1 ratio] and 7 M urea in 1× TBE buffer (*see* Subheading 2.2, item 17 for preparation from stocks, using a 19:1 rather than 29:1 acrylamide:bis-acrylamide commercial mixture).
18. Vacuum drying system for electrophoresis gels.
19. Phosphor imager equipped with 35 × 43 cm phosphor imaging plates and dedicated analysis software (e.g., Typhoon Trio imager and ImageQuant TL software from GE-Healthcare).

3 Methods

3.1 Preparation of DNA Templates for Transcription Termination Assays

We prefer working with DNA templates that do not exceed 1000 bp, as longer templates may yield transcription termination products that are not easily separated from runoff transcripts by denaturing polyacrylamide gel electrophoresis. The DNA templates contain the genomic region of interest downstream from the T7A1 promoter (which is efficiently used by the RNA polymerase from *E. coli*) and are prepared in two successive rounds of PCR amplification (Fig. 2a) as follows:

1. Using a sterile loop, scrap and transfer a tiny piece of frozen bacteria (*see* **Note 1**) into a 1.5 mL microtube. Add 100 μL of water, incubate for 5 min at 95 °C, and then keep on ice (this solution can be stored at −20 °C for further use). A diluted solution of purified genomic DNA may be used instead.
2. Transfer 3 μL of above solution into a 0.5 mL PCR microtube kept on ice. Add 39 μL of water, 5 μL of 10× Vent reaction buffer, 0.5 μL of each 100 μM solution of primers FWD and REV (Table 1), 1 μL of the stock mixture of dNTPs (10 mM each), and 1 μL of Vent polymerase.
3. Transfer mixture into a thermocycler and heat it for 3 min at 94 °C. Then, perform 30 cycles of PCR amplification using the following cycle parameters: 94 °C for 1 min, 50 °C for 30 s, 72 °C for 1 min. Incubate for 10 min at 72 °C at the end of the program.
4. Transfer 0.6 μL of the PCR mixture into a new 0.5 mL PCR microtube kept on ice. Add 41.4 μL of water, 5 μL of 10× Vent reaction buffer, 0.5 μL of each 100 μM solution of primers T7A1 and REV (Table 1), 1 μL of the stock mixture of dNTPs (10 mM each), and 1 μL of Vent polymerase.
5. Perform the second PCR round as in **step 3**. At the end of amplification, add 2 μL of 0.5 M EDTA solution and store on ice.
6. Prepare agarose gel by dissolving 1.5 g of agarose into 100 mL of 1× TAE buffer (perform short heating bursts in microwave oven). Add 1 μL of ethidium bromide stock solution (*see* **Note 5**) and pour into gel tray. Once the gel has solidified, install it into a horizontal electrophoresis unit filled with 1× TAE buffer.
7. Mix 1 μL of PCR reaction mixture with 7.3 μL of 1× $T_{10}E_1$ buffer, and 1.7 μL of 6× agarose gel loading buffer. Similarly prepare a sample containing ~0.5 μg of commercial DNA ladder.
8. Load samples on agarose gel and run the gel for 1 h at 100 Volts.
9. Visualize gel with a transilluminator or dedicated gel documentation system. Only one band migrating at the expected rate for the correct DNA fragment should be visible (Fig. 2c).



B

DNA template for *in vitro* transcription

5′.TTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCCTTCGGGTGAACAGAGTGCTA
ACAAAATGTTGCCGAACAACAAGCCAACTGCGACCACGGTCACAGCGCCTGTAACGGTACCAACAGCAAGCACAA
CCGAGCCGACTGTCAGCAGTACATCAACCAGTACGCCTATCTCCACCTGGCGCTGGCCGACTGAGGGCAAAGTGA
TCGAAACCTTTGGCGCTTCTGAGGGGGGCAACAAGGGGATTGATATCGCAGGCAGCAAAGGACAGGCAATTATCG
CGACCGCAGATGGCCGCGTTGTTTATGCTGGTAACGCGCTGCGCGGCTACGGTAATCTGATTATCATCAAACATA
ATGATGATTACCTGAGTGCCTACGCCCATAACGACACAATGCTGGTCCGGGAACAACAAGAAGTTAAGGCGGGGC
AAAAAATAGCGACCATGGGTAGCACCGGAACCAGTTCAACACGCTTGCATTT***TGAAATTCGTTACAAGGGGAAAT***
CCGTAAACCCGCTGCGTTATTTGCCGCAGCGATAAATCGGCGGAACCAGGCTTTTGCTTGAATGTTCCGTCAAGG
GATCACGGGTAGGAGCCACCTT**atg**AGTCAGAATACGCTGAAAGTTCATGATTTAAATGAAGATGCGGAATTTGA
TGAGAACGGAGTTGAGGTTTTTGACGAAAAGGCCTTAGTAGAACAGGAACCCAGTGATAACGATTTGGCCGAAGA
GGAACTGTTAT.3′

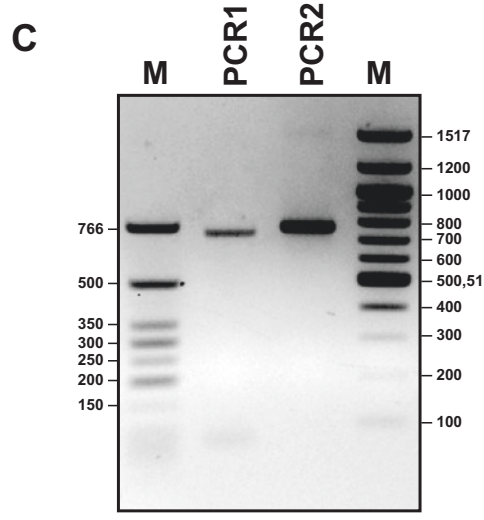


Fig. 2 Preparation of DNA templates for in vitro transcription. (**a**) Outline of the method. (**b**) Sequence of the *rpoS* template used as example in the present chapter. The template contains the 5′UTR as well as the first 139 nucleotides of the *rpoS* coding region. The sequence of the T7A1 promoter is underlined while the region recognized by sRNAs ArcZ and DsrA [28] and the ATG start codon are in bold. (**c**) Representative 1.5% agarose gel of the DNA fragments obtained after rounds 1 and 2 of PCR amplification of the *rpoS* template using primers FWD and REV (*see* Table 1)

10. Purify the PCR reaction mixture in two steps, first with a commercial silica-based purification kit (*see* **Note 3**) and then with a commercial G-50 spin column following manufacturer's instructions.
11. Use a μL spectrophotometer to determine the molar concentration of the DNA template from the absorbance of the solution at 260 nm, assuming ε_{260} ~ [13,200 × number of base pairs] L/mol/cm. Typical yields for a 1000 bp DNA fragment are around 10 pmole.
12. Adjust DNA template concentration to 100 nM with $T_{10}E_1$ buffer and store at −20 °C.

3.2 Preparation of sRNAs

Most sRNAs exceed the size limit of commercial synthetic oligoribonucleotides (~80 nts) and need to be prepared by in vitro transcription with a phage RNA polymerase using either a dedicated commercial kit or the following protocol:

1. Follow instructions in Subheading 3.1 to prepare the DNA template encoding the sRNA of interest. To introduce a promoter for T7 RNA polymerase (instead of the T7A1 promoter recognized by the RNA polymerase from *E. coli*), use oligonucleotides TOP and BOTT instead of primers FWD (or T7A1) and REV in both rounds of PCR.
2. Gently thaw transcription buffer, DNA template, and rNTP stocks on ice. Homogenize each solution by brief vortexing and centrifugation.
3. In a microtube, assemble on ice a mixture of 119 μL of water, 50 μL of 5× transcription buffer, 12.5 μL of each 100 mM rNTP stock, 1 μL of SUPERase-IN™, and 20 μL of the 100 nM DNA template solution.
4. Add 10 μL of T7 RNA polymerase and incubate mixture for 2 h at 37 °C.
5. Add 5 μL of RQ1 DNase to digest the DNA template and incubate for 20 min at 37 °C.
6. Add 12 μL of 0.5 M EDTA and 28 μL of 3 M NaAc.
7. Extract with one volume of Phenol:Chloroform:Isoamyl alcohol mix. Vortex and centrifuge briefly to separate phases. Transfer the aqueous (top) phase to a new tube and extract twice with one volume of ether, discarding the top (ether) phase in each case.
8. Add 900 μL of ice-cold ethanol. Incubate overnight at −20 °C.
9. Centrifuge for 30 min at 20,000 × *g* in a refrigerated centrifuge and discard supernatant. Wash the pellet with 150 μL of ice-cold ethanol and centrifuge for 15 min at 20,000 × *g*.

10. Remove the ethanol and leave the microtube open for ~20 min at room temperature to dry the RNA pellet (the process can be sped up by incubating at 37 °C in a dry bath) (*see* **Note 10**).
11. Dissolve pellet in a mix of 20 µL of $M_{10}E_1$ buffer and 20 µL of denaturing loading buffer. Leave sample on ice while preparing the denaturing polyacrylamide gel for purification.
12. Assemble gel plates and spacers according to manufacturer instructions. We use custom-made 20 × 20 cm gel plates equipped with 0.8 mm spacers, a 10-teeth comb, and a bottom tape seal.
13. Mix 30 mL of 6% denaturing (29:1) polyacrylamide gel solution with 90 µL of APS and 45 µL of TEMED (volumes may need to be adjusted for commercial sets of plates and spacers). Quickly pour the mixture between the gel plates, and insert comb.
14. Once the gel has polymerized (~30 min), remove the comb and wash the gel wells with 1× TBE using a 5 mL syringe. Install the gel into an electrophoresis unit and fill the top and bottom tanks with 1× TBE.
15. Set the power supply at 20 W and perform a pre-electrophoresis for 20 min.
16. Heat-denature the RNA sample (from **step 11**) for 2 min at 95 °C.
17. Turn off the power supply and flush diffusing urea from gel wells using a syringe containing 1× TBE. Distribute sample into two wells using a flat gel loading tip.
18. Run the gel at 20 W until the band corresponding to bromophenol blue reaches the bottom of the gel.
19. Carefully remove the glass plates and wrap the gel in saran sheets.
20. Place the gel on an X-ray intensifying screen (or a fluor-coated TLC plate) and visualize the band corresponding to the transcript by UV shadowing in a dark room with a hand-held 254 nm lamp (*see* **Note 11**).
21. Cut the band with a clean scalpel and crush it by passage through a 1 mL syringe. Soak the gel pieces in 3 mL of 1× elution buffer in a sterile 14 mL culture tube. Shack the tube overnight at 4 °C.
22. Pass the gel slurry through a 5-mL syringe equipped with a glass wool or cotton plug (to retain gel particles) and measure the volume of the resulting solution.
23. Extract the filtered solution with one volume of Phenol:Chloroform:Isoamyl alcohol mix. Vortex and centrifuge briefly to separate phases. Transfer the aqueous (top) phase to a new tube and extract twice with one volume of ether.

24. Add three volumes of ice-cold ethanol and incubate overnight at −20 °C.
25. After centrifugation for 30 min at 20,000 × *g* in a refrigerated benchtop centrifuge, discard supernatant and wash carefully the RNA pellet with 300 µL of 70% ice-cold ethanol. Centrifuge again for 10 min at 20,000 × *g* and discard the ethanol wash.
26. Leave the microtube open for ~20 min at room temperature to dry the RNA pellet (*see* **Note 10**) and dissolve it in 50–100 µL of $M_{10}E_1$ buffer.
27. Use a µL spectrophotometer to determine RNA concentration from the absorbance of the solution at 260 nm, assuming ε_{260} ~ [10^4 × number of nucleotides] L/mol/cm. Typical yields range between 1 and 4 nmoles of purified sRNA for a 250 µL transcription.
28. Store sRNA stock solution at −20 °C.

3.3 Transcription Termination Assay

3.3.1 Detection of Rho-Dependent Signals

To probe the effect of a sRNA or chaperone on a known, well-characterized Rho-dependent termination signal, one may skip this section and proceed directly to Subheading 3.3.2.

If the evidence connecting the effects of Rho and a sRNA (or RNA chaperone) on a given gene or operon is vague or indirect, the first step is to determine if this gene or operon contains a Rho-dependent signal. We advise to also check for the potential effect of NusG as this factor can dramatically stimulate Rho-dependent termination at suboptimal *Rut* sites (as was observed for the ChiX-regulated *chiP* terminator of *Salmonella*) [14]. To undertake these tasks, proceed as follows:

1. To a regular 1.5 mL microtube, add in the following order, 3 µL of deionized water, 1 µL of 10× PNK buffer, 2 µL of DNA ladder (1 µg/µL), 3 µL of γ-^{32}P-ATP (*see* **Note 9**), and 1 µL of T4 polynucleotide kinase. Incubate 40 min at 37 °C.
2. Add 2 µL of 0.5 M EDTA and 78 µL of $T_{10}E_1$ buffer. Pass through a G-50 spin column, following manufacturer's instructions. Discard the G-50 column and store the eluate containing the ^{32}P-labeled DNA ladder at −20 °C (*see* **Note 9**). The ^{32}P-labeled ladder can be used for up to ~2 months.
3. Prepare three 1.5 mL microtubes ("Protein low binding" grade) labeled "T," "Rho," and "NusG," respectively. Place tubes on ice.
4. In a "Protein low binding" microtube prepare a master mix containing 35.2 µL of deionized water H_2O, 10 µL of 5× transcription termination buffer, 3.2 µL of the 100 nM stock solution of DNA template, 0.6 µL of SUPERase-IN™, and 0.9 µL of *E. coli* RNA polymerase.

5. Dispatch 15.6 µL of the master mix in tubes "T," "Rho," and "NusG."
6. Add 2.4 µL of 1× transcription termination buffer to tube "T." To tube "Rho," add 1.4 µL of 1× transcription termination buffer and 1 µL of the 1.4 µM Rho stock. Add 1 µL of the 1.4 µM Rho stock and 1.4 µL of the 2 µM NusG stock to tube "NusG." Vortex gently and centrifuge briefly to homogenize each tube solution.
7. Incubate reaction tubes for 10 min at 37 °C.
8. Add 2 µL of 10× initiation mixture (*see* **Note 12**). From this step to the end of the procedure, samples will contain substantial amounts of radioactive material and should be handled accordingly (*see* **Note 9**).
9. Incubate reactions for 20 min at 37 °C. Then, stop reactions by adding, to each tube, 4 µL of EDTA (0.5 M), 2 µL of tRNA (0.25 mg/mL), 64 µL of deionized water, and 10 µL of 3 M NaAc pH 6.5.
10. Extract each tube solution with one volume of Phenol:Chloroform:Isoamyl Alcohol mix. Vortex and centrifuge briefly to separate phases. Transfer the aqueous (top) phases to new tubes (use only "DNA low binding" tubes from this step) and extract twice with one volume of ether.
11. Add three volumes of ethanol (stored at room temperature) to each tube and incubate for 35 min on ice. Under these conditions, most of the free ^{32}P-αUTP will not precipitate.
12. Centrifuge for 20 min at 20,000 × *g* in a bench centrifuge at room temperature. Discard supernatants and wash carefully the pellets with 150 µL of 100% ethanol (stored at room temperature). Centrifuge again for 10 min at 20,000 × *g* and discard the ethanol washes. Open the tubes and dry the pellets for ~5 min at room temperature (*see* **Note 10**).
13. Following instructions in **steps 12** and **13** of Subheading 3.2 but using 20 × 40 cm (w × l) gel plates and 0.4 mm spacers, prepare a 7% denaturing (19.1) polyacrylamide gel and install it in a vertical electrophoresis unit.
14. Perform a pre-electrophoresis for 45 min at 45 W (gel plates should become warm to the touch).
15. In a 1.5 mL microtube, mix 9 µL of denaturing loading buffer with 1 µL of ^{32}P-labeled ladder from **step 2** (increase volume of ladder preparation if older than a few days).
16. Dissolve pellets from **step 12** in 6 µL of 1× resuspension buffer (incubate tubes at 30 °C for a few minutes to help dissolution) and add 7 µL of denaturing loading buffer.

17. Heat sample and ladder tubes for 2 min at 90 °C. Then, flush diffusing urea from gel wells using a syringe containing 1× TBE and quickly load samples and ladder in the gel wells.
18. Run the gel at 45 W until the band corresponding to xylene cyanol is ~25 cm from the bottom of gel wells.
19. Carefully remove one of the glass plates and replace it with a sheet of Whatman paper. Remove the second glass plate and replace it with saran wrap.
20. Dry the gel in a vacuum gel dryer and expose it overnight to a phosphorimager screen in an exposure cassette.
21. Scan the phosphorimager screen with a dedicated system. The presence of a Rho-dependent signal in the DNA template of interest (Fig. 3a) is made apparent by the apparition of fast migrating bands in the samples containing Rho and by the concomitant decrease in the intensity of the band corresponding to the formation of runoff transcripts (Fig. 3b) (*see* **Note 13**). This trend is usually accentuated in samples containing both Rho and NusG (Fig. 3b) (*see* **Note 14**).

3.3.2 Effects of sRNAs and/or RNA Chaperones

Below we describe our procedure to probe the effects of sRNAs on Rho-dependent termination. The protocol can be easily adjusted to probe the effects of RNA chaperones such as Hfq [17] or CsrA [11]. Because Hfq alone can inhibit Rho [17], we recommend testing Hfq and sRNAs separately, and using control oligoribonucleotides that form perfect duplexes with the mRNA regions targeted by the sRNAs. To find transcription conditions optimal for sRNA-mRNA pairing, we also recommend to run a few exploratory experiments where the concentration of the tested sRNA or the concentration of KCl is varied. In the case of the *rpoS* template, for instance, we observed that the effect of the sRNA DsrA is optimal in the presence of 75 mM KCl (Fig. 3c) and we adjusted the composition of the transcription buffer accordingly.

1. Prepare and label one 1.5 mL "Protein low binding" microtube per planned sample (including "minus Rho" and "minus sRNA" controls) and place tubes on ice.
2. For the n planned samples, prepare a master mix containing $9.4 \times (n + 1)$ µL of deionized water, $2.7 \times (n + 1)$ µL of 5× transcription termination buffer (containing 375 mM KCl in the case of the *rpoS* template), $(n + 1)$ µL of the 100 nM stock solution of DNA template, $0.2 \times (n + 1)$ µL of SUPERase-IN™, and $0.27 \times (n + 1)$ µL of *E. coli* RNA polymerase.
3. Dispatch 13.6 µL of the master mix into each sample tube.
4. Add 2.4 µL of 1× transcription termination buffer (containing 75 mM KCl in the case of the *rpoS* template) to tube "minus Rho."

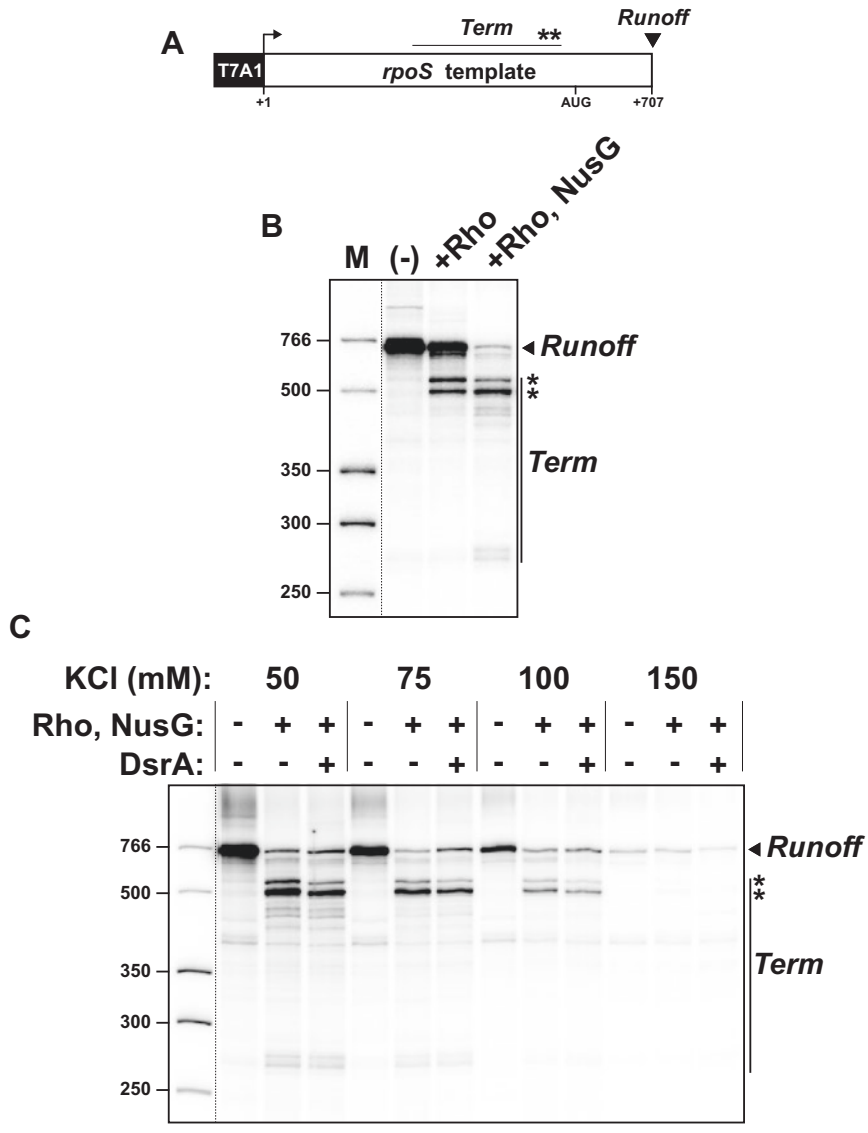


Fig. 3 Representative exploratory transcription termination experiments performed with the *rpoS* template. (**a**) Schematic of the *rpoS* template and transcription termination features. (**b**) Initial detection of Rho-dependent signals within the *rpoS* template. (**c**) Analysis of the effect of KCl. In the example shown, 75 mM KCl represents the best compromise between the effects of the salt on transcription initiation, Rho-dependent termination, and DsrA efficiency. The KCl concentration was increased right from the start of the reaction assay but one may change it at a later stage (to limit the inhibitory salt effects on transcription initiation such as shown on the present gel) by adjusting the composition of the 10× initiation mixture rather than the 5× transcription termination buffer. Note that the contrast of the DNA ladder lane has been adjusted separately as indicated by the dotted line separating it from other gel lanes

To each other tube (including "minus sRNA" tube), add 1 μL of the 1.4 μM Rho stock and 1.4 μL of the 2 μM NusG stocks.

5. Mix reactants by gently pipetting up and down and incubate for 10 min at 37 °C.
6. Add 2 μL of 1× transcription termination buffer to tubes "minus sRNA." To each other tube, add 2 μL of the 5 μM stock of appropriate sRNA or control oligoribonucleotide (*see* **Note 15**). Vortex gently and centrifuge briefly to homogenize each tube solution.
7. Proceed as described in **steps 8–21** of Subheading 3.3.1
8. Determine the apparent percentage of runoff product (*see* **Note 16**) using the phosphorimager analysis software. For instance, one may use the analysis toolbox of ImageQuant TL software (GE Healthcare) to box the runoff and termination product bands (rectangle area function; *see* Fig. 4, inset), select the "local average" option for background and "percent" option for view, and obtain percentages of band intensities directly and in an Excel-compatible format.
9. Evaluate sRNA- or chaperone-dependent effects through the comparison of the apparent percentages of runoff product (*see* **Note 16**) obtained for the different gel lanes (Fig. 4, diagram). We recommend using average percentage values obtained from at least three independent experiments to mitigate potential RNA degradation effects and the experimental variability inherent to such complex biochemical setups (*see* **Note 17**).

4 Notes

1. For the case chosen to illustrate the present chapter (analysis of the *rpoS* region of *E. coli*), we used a 20% glycerol stock of the *E. coli* reference strain MG1655 stored at −80 °C. The strain (or genomic DNA) can be obtained from biological repositories such as DSMZ (www.dsmz.de).
2. The Vent DNA polymerase works well for the preparation of most DNA templates having lengths in the 300–1000 bp range. Other high-fidelity DNA polymerases may be used or better when preparing DNA templates longer than 1 kB.
3. In our hands, the GeneJET PCR purification kit (ThermoFisher Scientific) works optimally when following manufacturer's instructions.
4. To eliminate bacteria, the major source of RNase contamination, stock solutions and buffers for transcription assays should be prepared with RNA-grade chemicals and water in small

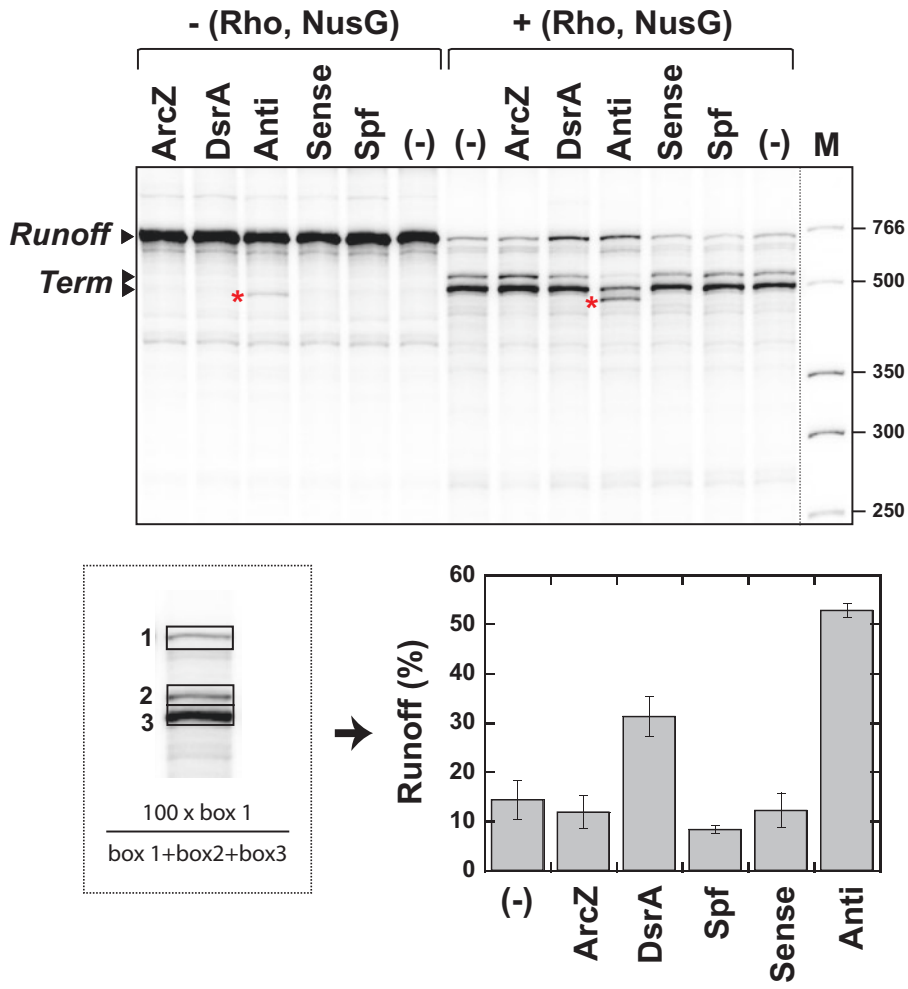


Fig. 4 Effects of sRNAs and control oligoribonucleotides on the *rpoS* terminator. Reactions were performed in the presence of 75 mM KCl and 500 nM of sRNA or control oligonucleotide "Sense" (5′ACUUUAAGCAAUGUUCCCCUUUAGGC) or "Anti" (5′CGGAUUUCCCCUUGUAACGAAUUUCA). The bands identified by asterisks correspond to a Rho-independent product that is formed only in the presence of the "Anti" strand, possibly because pairing of the oligonucleotide to the mRNA triggers intrinsic termination [29]. Runoff percentages shown in the diagram are average values obtained from 3 to 6 independent experiments (*see* **Note 16**). The ArcZ sRNA has no significant effect under the present conditions whereas it inhibited Rho-dependent termination at a higher concentration (1 μM) in a slightly different experimental setup [10]. This is consistent with the observation that ArcZ forms a less stable hybrid with the *rpoS* mRNA than does DsrA [28]

amounts (<50 mL) and sterilized with a 0.22 μm filter unit. We routinely obtain RNA-grade water by filtering ultrapure MilliQ (Millipore) water with 0.22 μm bottle-top sterile filter units. We usually avoid DEPC-treated water because harmful contaminants, such as rust particles, are often introduced during the autoclaving step required to remove excess DEPC.

5. Ethidium bromide is toxic. Solutions and waste (including gels) should be handled with great care following current safety regulation. Potentially less toxic DNA stains are available from various commercial sources and may be preferable for inexperienced users.
6. Concentrations of the Rho factor are expressed in hexamers throughout the chapter.
7. It is important to assess the enzymatic activity of every fresh Rho preparation and to verify that it does not decay upon storage at −20 °C for extended period of times. The simplest test for Rho activity is the determination of the rate of steady-state ATPase turnover as described in volume 1259 of the series [21].
8. NusG, CsrA, and Hfq proteins are now available from commercial sources (e.g., MyBioSource) which we have not tested. We prepare and use His-tagged versions of NusG [22] and CsrA [11] and obtained purified Hfq from V. Arluison [17].
9. Manipulation of ^{32}P-containing materials should be performed exclusively by individuals who have received proper training and authorization from a radiation safety officer.
10. Overdried RNA pellets are difficult to dissolve in aqueous buffers.
11. Full-length RNA should be seen clearly as one strong UV-shadow. If several shadows of comparable intensities are visible, one should assume that RNA degradation and/or formation of abortive transcription products have occurred in large proportion. It is then advisable to repeat the transcription/purification procedure with fresh solutions and reactants.
12. Rifampicin is present in the initiation mix to block subsequent rounds of transcription. Heparin cannot be used instead of rifampicin because it will trap Rho [23] in addition to free RNA polymerase.
13. Degradation of the transcripts (e.g., upon contamination of the samples by RNases) may yield similar band patterns. To ensure that the apparition of shorter-than-runoff termination products is Rho-dependent, it is advisable to repeat the experiment several times. Moreover, to ensure that neither the Rho nor NusG stock is contaminated by RNases, one may prepare control samples containing 150 μM of the Rho inhibitor bicyclomycin (supplemented in the initial mix of reactants). Rho-dependent bands will disappear at the profit of the runoff band whereas degradation product band patterns will not be affected by the presence of bicyclomycin.
14. Strictly speaking, the assay does not distinguish between truncated transcripts resulting from Rho-induced dissociation of transcription elongation complexes (termination) or from

transcription complexes that would be stably paused (or arrested) by Rho along the DNA template. To the best of our knowledge, however, the second scenario (Rho inducing transcriptional pausing or arrest) is purely hypothetical, without experimental support to date. We thus feel confident that the more complicated transcription assays required to distinguish between termination, pausing, and arrest [24] are not necessary in most cases.

15. When testing the effect of an RNA chaperone, the chaperone is added at this step in the place of the sRNA(s).
16. The apparent percentage of runoff product determined under these conditions should be used only to detect sRNA-dependent effects through the comparison of gel lanes (samples), as shown in the diagram of Fig. 4. This is because the apparent percentage of runoff product is not an accurate measure of terminator read-through as transcripts of different lengths contain different numbers of ^{32}P labels upon their internal labeling with ^{32}P-αUTP during transcription. Normalization of the intensities of gel bands for the uracil content of the corresponding transcripts is possible only when the sequences (lengths) of the various transcript species are known [25].
17. Because Rho is a nucleic acid-binding protein displaying some level of sequence specificity [26], indirect "sequestration" effects due to sRNA binding to Rho cannot be excluded beforehand, especially when considering the respective concentrations of Rho (70 nM) and sRNAs (500 nM) used in the assay. We strongly recommend testing this possibility by performing control transcription termination experiments with DNA template(s) encoding Rho-dependent terminator(s) unrelated to the original mRNA target sequence.

Acknowledgments

This work was supported by a PhD scholarship from Région Centre Val-de-Loire to C.N. and by a grant from Agence Nationale de la Recherche (ANR-15-CE11-0024-02) and CNRS core funding to M.B.

References

1. Porrua O, Boudvillain M, Libri D (2016) Transcription termination: variations on common themes. Trends Genet 32:508–522
2. Ray-Soni A, Bellecourt MJ, Landick R (2016) Mechanisms of bacterial transcription termination: all good things must end. Annu Rev Biochem 85:319–347
3. Santangelo TJ, Artsimovitch I (2011) Termination and antitermination: RNA polymerase runs a stop sign. Nat Rev Microbiol 9:319–329

4. Kriner MA, Sevostyanova A, Groisman EA (2016) Learning from the leaders: gene regulation by the transcription termination factor Rho. Trends Biochem Sci 41:690–699
5. Gish K, Yanofsky C (1995) Evidence suggesting cis action by the TnaC leader peptide in regulating transcription attenuation in the tryptophanase operon of Escherichia coli. J Bacteriol 177:7245–7254
6. Matsumoto Y, Shigesada K, Hirano M et al (1986) Autogenous regulation of the gene for transcription termination factor rho in Escherichia coli: localization and function of its attenuators. J Bacteriol 166:945–958
7. Yakhnin H, Babiarz JE, Yakhnin AV et al (2001) Expression of the Bacillus subtilis trpEDCFBA operon is influenced by translational coupling and Rho termination factor. J Bacteriol 183:5918–5926
8. Hollands K, Proshkin S, Sklyarova S et al (2012) Riboswitch control of Rho-dependent transcription termination. Proc Natl Acad Sci U S A 109:5376–5381
9. Takemoto N, Tanaka Y, Inui M (2015) Rho and RNase play a central role in FMN riboswitch regulation in Corynebacterium glutamicum. Nucleic Acids Res 43:520–529
10. Sedlyarova N, Shamovsky I, Bharati BK et al (2016) sRNA-mediated control of transcription termination in E. coli. Cell 167:111–121.e113
11. Figueroa-Bossi N, Schwartz A, Guillemardet B et al (2014) RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. Genes Dev 28:1239–1251
12. Jorgensen MG, Thomason MK, Havelund J et al (2013) Dual function of the McaS small RNA in controlling biofilm formation. Genes Dev 27:1132–1145
13. Wang X, Dubey AK, Suzuki K et al (2005) CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of Escherichia coli. Mol Microbiol 56:1648–1663
14. Bossi L, Schwartz A, Guillemardet B et al (2012) A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. Genes Dev 26:1864–1873
15. Wang X, Ji SC, Jeon HJ et al (2015) Two-level inhibition of galK expression by spot 42: degradation of mRNA mK2 and enhanced transcription termination before the galK gene. Proc Natl Acad Sci U S A 112:7581–7586
16. Yakhnin AV, Babitzke P (2014) NusG/Spt5: are there common functions of this ubiquitous transcription elongation factor? Curr Opin Microbiol 18:68–71
17. Rabhi M, Espeli O, Schwartz A et al (2011) The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. EMBO J 30:2805–2816
18. Peters JM, Mooney RA, Grass JA et al (2012) Rho and NusG suppress pervasive antisense transcription in Escherichia coli. Genes Dev 26:2621–2633
19. Shashni R, Qayyum MZ, Vishalini V et al (2014) Redundancy of primary RNA-binding functions of the bacterial transcription terminator Rho. Nucleic Acids Res 42:9677–9690
20. Boudvillain M, Walmacq C, Schwartz A et al (2010) Simple enzymatic assays for the in vitro motor activity of transcription termination factor Rho from Escherichia coli. Methods Mol Biol 587:137–154
21. D'Heygere F, Schwartz A, Coste F et al (2015) Monitoring RNA unwinding by the transcription termination factor rho from Mycobacterium tuberculosis. Methods Mol Biol 1259:293–311
22. Artsimovitch I, Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc Natl Acad Sci U S A 97:7090–7095
23. Nowatzke W, Richardson L, Richardson JP (1996) Purification of transcription termination factor Rho from Escherichia coli and Micrococcus luteus. Methods Enzymol 274:353–363
24. Kashlev M, Nudler E, Severinov K et al (1996) Histidine-tagged RNA polymerase of Escherichia coli and transcription in solid phase. Methods Enzymol 274:326–334
25. Rabhi M, Gocheva V, Jacquinot F et al (2011) Mutagenesis-based evidence for an asymmetric configuration of the ring-shaped transcription termination factor Rho. J Mol Biol 405:497–518
26. Rabhi M, Rahmouni AR, Boudvillain M (2010) Transcription termination factor Rho: a ring-shaped RNA helicase from bacteria. In: Jankowsky E (ed) RNA helicases, vol 19. RSC Publishing, Cambridge, pp 243–271
27. Gocheva V, Le Gall A, Boudvillain M et al (2015) Direct observation of the translocation mechanism of transcription termination factor Rho. Nucleic Acids Res 43:2367–2377
28. Soper T, Mandin P, Majdalani N et al (2010) Positive regulation by small RNAs and the role of Hfq. Proc Natl Acad Sci U S A 107:9602–9607
29. Yarnell WS, Roberts JW (1999) Mechanism of intrinsic transcription termination and antitermination. Science 284:611–615

Chapter 8

Mapping Changes in Cell Surface Protein Expression Through Selective Labeling of Live Cells

Pierre Fechter

Abstract

ncRNAs are key players in the adaptation of bacteria to new environments, by modulating the composition of the membrane upon changes in the environment. Nevertheless, monitoring the changes in surface protein expression is still a challenge, since these proteins are present in low abundance, and are difficult to extract. Here is described a method to easily, reproducibly, and specifically enrich total protein extracts in surface proteins. This method comprises a direct labeling of surface proteins on living cells using fluorescent dyes, followed by total protein extraction and subsequent separation of these extracts by 2D gel electrophoresis.

Key words CyDye, 2D gel electrophoresis, Surface proteins, Cell labeling, ncRNAs

1 Introduction

Over the past decade, it has become clear that all types of cells encode for small, noncoding RNAs (ncRNAs) that have important roles in regulating gene expression at the posttranscriptional level [1]. Transcriptomic studies allowed the identification of 50–100 ncRNAs in different bacteria, yet the function of only a tiny number has been deciphered. For those ones, an increasing number of studies highlighted their role in the rapid adaptation of the bacteria to new environments. They can directly regulate the expression of target proteins, like outer membrane proteins [2–4], at the posttranscriptional level, or indirectly through the control of regulation factors. More and more evidences show that these ncRNAs form complex networks [5, 6] to modulate the composition of the membrane upon environment changes. To have a more precise picture of these networks and to be able to integrate them in the physiological response of the bacteria to new environments, it is important to increase our knowledge on these networks. A first task would be to uncover the target of the ncRNAs, especially the ones exposed at the surface that are in direct contact with the

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_8,

environment. Thus, profiling the surface proteome is becoming important to understand global changes upon ncRNAs activation.

A major drawback in the study of these proteins is their low abundance. Different methods have been employed to enrich membrane fractions, but they still lead to significant loss of proteins or to contamination with cytoplasmic proteins [7–9]. The recent progress in mass spectrometry (especially in the analysis of membrane proteins) gave unprecedented insight into most protein expression pattern [10, 11]. Nevertheless, membrane proteins present at low concentration are still either not detected, or their detection level is not significant enough to compare their expression under different growth conditions. Thus, to gain insight into the expression and the function of these proteins, different and complementary approaches are still required. In this manuscript is presented an experimental guide for direct labeling of cell surface using fluorescent dyes, which ensures an important enrichment in surface proteins over cytoplasmic ones. The protocol will be described for *Staphylococcus aureus* cell surface labeling. An experimental procedure to easily, reliably, and efficiently extract total proteins, even in the case of bacteria which membranes are resistant to classical lysis procedures, like the one from *S. aureus*, is further described. Finally, protein extracts are analyzed through 2D gel fractionation. A comparison of this cell surface labeling protocol with the standard 2D Fluorescent Difference Gel electrophoresis (2D-DIGE) protocol (post-extraction labeling) showed a better detection of surface proteins with this protocol [12]. This method has already been successfully applied to the detection of surface proteins, and to the comparison of surface protein expression under various experimental conditions, in different bacteria (i.e., *Legionella pneumophila*, *Escherichia coli, Staphylococcus aureus, Xanthomonas citri*) [13–15], but also in eukaryotic cells [12, 16]. This method also revealed new potential targets of *E. coli* MicA ncRNA, and highlighted the role of *S. aureus* RNAIII for the maintenance of the cell wall integrity [14].

2 Materials

2.1 Strain and Culture Broth

1. Gram positive *S. aureus* RN6390.
2. BHI broth (Brain Heart Infusion broth, AES laboratoire, France).

2.2 Cell Labeling

1. PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.76 mM KH_2PO_4, pH = 8.
2. CyDye DIGE Fluor minimal dyes 2, 3, 5, dissolved in dimethylformamide at a concentration of 100 μM and stored at −80 °C. These molecules are derivatives of the fluorescent dyes

cyanine 2, 3, 5 (GE Healthcare, ref. 25-8010-82, 25-8010-83, 25-8010-85).

3. 10 mM lysine.

2.3 Protein Extraction

1. Lysis buffer for *S. aureus* cells prepared extemporaneously: 10 mM Tris–HCl pH 7.5, 20 mM NaCl, 1 mM EDTA, 5 mM $MgCl_2$, 50 µL/mL lysostaphin, protease cocktail inhibitor (Roche), 2 U DNase I (Qiagen), 2 U RNase (Roche).
2. Trizol (Invitrogen).
3. 1-bromo 3-chloropropane.
4. Ethanol.
5. Cold acetone.
6. Cold 70% acetone.
7. Sample buffer: 7 M urea, 2 M thiourea, 50 mM DTT, 4% CHAPS.

2.4 Protein Concentration Estimation

1. Bradford reagent.
2. 10 mg/mL BSA.
3. Sample buffer: 7 M urea, 2 M thiourea, 50 mM DTT, 4% CHAPS.

2.5 IEF Isofocalization

1. PROTEAN® i12™ IEF system (Biorad).
2. 2-iodoacetamide 1 M.
3. Disposable G-25 sephadex column for small volume (PD MiniTrap G-25, GE Healthcare).
4. 17 cm pre-cast pH 4-7 ReadyStrip™ IPG Strip (Biorad).
5. Rehydration/equilibration tray (Biorad).
6. Mineral oil (Biorad).
7. 0.5% bromophenol blue.
8. pH 3–10 ampholytes (Biorad).
9. IEF trays (Biorad).

2.6 SDS-PAGE

1. PROTEAN II XL cell (Biorad), for gel 17 × 20 cm.
2. Equilibration buffer: 6 M urea, 2% SDS, 30% glycerol, 1 M Tris–HCl pH 8.45.
3. Cathode buffer: 0.1 M Tris–HCl pH 8.25, 0.1 M Tricine, 0.1% SDS.
4. Anode buffer: 0.2 M Tris–HCl pH 8.9.
5. Acryl/bis mix: acrylamide 48%, bis-acrylamide 1.5%.
6. Gel buffer (3×): 3 M Tris–HCl pH 8.45, 0.3% SDS.
7. TEMED.

8. 10% ammonium persulfate.
9. Low melting agarose mix: 1% low melting agarose, 0.025% bromophenol blue, in cathode buffer.

2.7 Protein Identification

1. Amersham Typhoon RGB (GE Healthcare) to image the CyDye labeled proteins in the gels.
2. Fixation solution: 40% EtOH, 10% acetic acid.
3. Colloidal blue solution: 10% orthophosphoric acid, 0.025% Coomassie blue G-250, 40% ethanol, 10% (w/v) ammonium sulfate.
4. GS-900™ Calibrated Densitometer (Biorad) to image colloidal blue stained gels.
5. PDQuest software (version 7.4, Biorad).

3 Methods

3.1 Cell Culture

1. Grow *S. aureus* cells in BHI medium until late logarithmic phase (OD = 5).
2. Centrifuge 2.5 mL of culture (4000 × *g*, 15 min), and wash twice with PBS.
3. Resuspend the cells in 200 µL of PBS (*see* **Note 1**).

3.2 Cell Labeling

1. Label accessible lysines of the cell surface by addition of 200 pmol of CyDyes for 30 min at 4 °C in the dark (*see* **Note 2**). The NHS ester reactive group of the CyDye DIGE Fluor minimal dyes covalently attaches to the epsilon amino group of lysine of proteins via an amide linkage.
2. Stop this labeling reaction by adding 20 µL of 10 mM lysine for 10 min at 4 °C in the dark.
3. Centrifuge the cells, and wash them twice with PBS.

3.3 Protein Extraction

1. Resuspend the cells in 200 µL of lysis buffer, and incubate for 30 min at 37 °C (*see* **Note 3**).
2. Add 1 mL of trizol to the sample. Vortex, leave for 5 min at RT.
3. Add 100 µL of 1-bromo 3-chloropropane (chloroform can be used instead). Vortex the solution, leave for 5 min at RT, and centrifuge for 10 min at 10,000 × *g*. The upper phase contains the RNA, and is removed.
4. Add 300 µL EtOH to the lower–pink–phase, vortex smoothly, leave for 5 min at RT, and centrifuge for 5 min at 2000 × *g*. Any remaining DNA is precipitated during this step.
5. Add 7 volumes of cold acetone to the supernatant in a 50 mL Falcon, at least 2 h at −20 °C (better overnight), centrifuge at

4000 × *g* for 20 min, and wash the pellet twice with cold 70% acetone (*see* **Note 4**). Dry the pellet, and dissolve the proteins in 400 µL of sample buffer.

3.4 Protein Concentration Estimation

The Bradford assay can be used to estimate the concentration of complex samples usually in the range 200–900 µg/mL of proteins.

1. Prepare five to eight dilutions of a protein standard (here BSA) in sample buffer with a range of 200 to 900 µg/mL.
2. Dilute the protein extract in sample buffer to obtain 200–900 µg/mL.
3. Add 20 µL of the standard solutions and of the protein extract solutions to appropriately labeled test tubes. Set also a blank tube by adding 20 µL of sample buffer only. Protein extracts are normally assayed in duplicate or triplicate.
4. Add 1.0 mL of Bradford reagent to each tube and mix well.
5. Incubate at room temperature for at least 5 min. Absorbance will increase over time; samples should not be incubated at room temperature for more than 1 h.
6. Measure absorbance at 595 nm.
7. Use the absorbance of the diluted standard protein solutions to draw a standard curve (absorbance over concentration of BSA). Fit the curve, and use the equation to estimate the concentration of the extract.

3.5 IEF Isofocalization

1. Alkylate 300 µg of proteins: complete the volume to 300–400 µL with sample buffer, add 10 µL of 1 M 2-iodoacetamide and incubate at 37 °C for 15 min. This chemical agent will alkylate the thiol group of cysteines.
2. Desalt the protein sample on a disposable G-25 Sephadex column.
3. Complete the volume again to 400 µL if necessary, add 4 µL ampholytes, and 2 µL of 0.5% bromophenol blue.
4. Re-swell a pre-cast pH 4-7 IPG strip with the sample (*see* **Note 5**). Distribute the sample all along a slot of the rehydration tray and add the strip onto the sample, gel-side onto the sample (*see* **Note 6**) for 1–2 h. Then add 3 mL of mineral oil to avoid evaporation of the sample and leave at room temperature for 16 h (this can be done in the Protean IEF system).
5. Transfer the strip into the tray with electrodes, protect the electrodes with small square of wet paper filter, cover with oil (10 mL), put the tray in the Protean IEF system, and perform the isoelectric focusing at 300 V for 2 h (*see* **Note 7**), then at 3000 V for a total of 75 kVh (usually it takes O/N).

3.6 SDS-PAGE

1. Exchange the buffer of the IPG strip. Put the strip gel-side up into a new rehydration/equilibration tray and add 10 mL of equilibration buffer for 15 min. At this step, the strip can be stored at −20 °C for several days.
2. Prepare 50 mL of gel solution (16.6 mL gel buffer, 12.5 mL acryl/bis mix, complete with water to 50 mL and then add 0.05 mL TEMED and 0.5 mL of 10% ammonium persulfate). Cast the gel vertically, to a height of about 2 cm below the top of the glass plates. Carefully overlaid the gel with 1.0–1.5 mL of water (*see* **Note 8**).
3. Remove the water once the gel is polymerized, add the strip on the top of the gel. Add low melting agarose mix. Make sure to avoid the formation of bubbles between the gel and the strip. Use anode and cathode (upper chamber) buffers for the migration. Run is carried out at 10 mA/gel for 1–2 h, and then overnight at 20 mA/gel (set the voltage limit at 300 V) until the bromophenol blue reaches the bottom of the gel.
4. After electrophoresis, scan the 2D gel with a Typhoon RGD imager using filters to select adequate excitation and emission wavelengths specific for CyDye 2 (excitation 489 nm/emission 506 nm), CyDye 3 (548/562 nm), or CyDye 5 (650/680 nm).
5. Fix the gel for 1 h with the fixation solution, wash twice with water, and stain all the proteins from the extract with colloidal blue. After one night wash the gel again with water and scan it on a GS900 scanner. Thus, two images are obtained from the same gel, one with the cell surface proteins labeled with the CyDyes, and one with all the proteins from the extract stained with colloidal blue (*see* Fig. 1).

3.7 Protein Identification

1. When different gels have to be compared, analyze the data with the software PDQuest. Crop and filtrate the images from the Typhoon imager (showing CyDye labeled proteins). Spots should be automatically detected and measured, background subtracted. Compare all gels with one selected gel used as a reference. Spots are matched, and unmatched spots manually added. To compare spots between gels, normalize the spot volumes as a percentage of the total volume of all spots in the gel.
2. Assign the CyDye spots of interest on the colloidal blue stained gel (*see* **Note 9**). An example of such assignment is shown in Fig. 1.
3. Protein spots are excised from the gels, manually or with an automatic spot cutter.
4. The mass spectrometry analysis and the identification of the proteins are usually conducted by a proteomic platform.

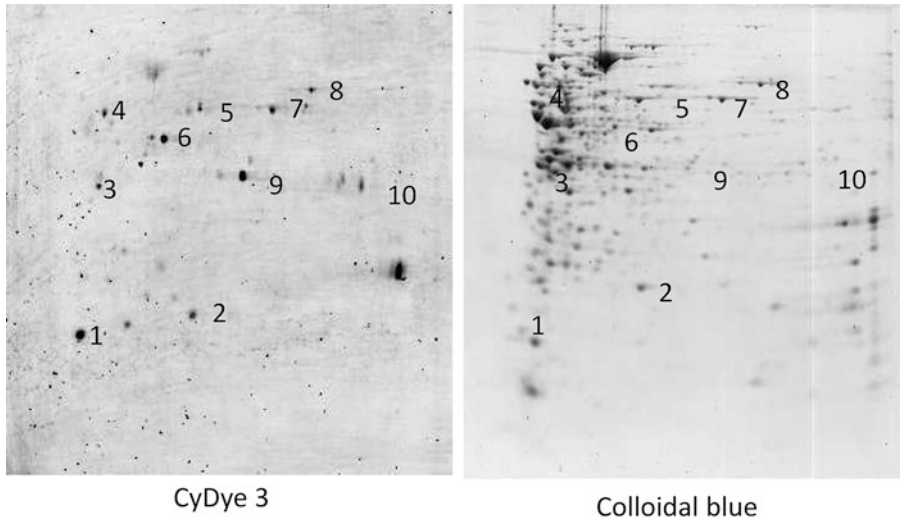


Fig. 1 Assignment of CyDye spots on colloidal blue stained gel. Ten different spots, revealed by CyDye 3 labeling of *S. aureus* RN6390 cells treated as described above are highlighted (left). These spots were assigned on the same gel, stained with colloidal blue (right). Some spots were hardly detected on the colloidal blue gel, or lost among all cytoplasmic protein spots. Without prior labeling treatment, a comparison of the expression of these proteins under different growth conditions would have been difficult

4 Conclusion

Enrichment procedures through labeling of live cells improve the detection, but more over the comparison of the expression of surface proteins under different growth conditions. It should be noted that besides labeling with fluorescent dyes, another surface protein labeling procedure was developed these last years. Sulfo-NH-SS-biotin has been used to label live cells. After cell lysis, biotinylated proteins are purified by affinity chromatography, and identified by mass spectrometry [8, 17–19]. Altogether, one should still remember that all the methods developed up to now to map and compare membrane proteomes bring complementary information, but that none allows full mapping and comparison of these complex proteomes.

5 Notes

1. This labeling protocol can be applied to different bacteria, gram positive as well as gram negative. The number of cells should nevertheless be kept constant. Only the lysis protocol may require some adaptations (*see* **Note 3**).
2. All three CyDyes do not have exactly the same labeling properties, due to subtle differences in their structure. CyDye 3 was

shown to be more efficient than CyDye 2 and 5 on *S. aureus* and *E. coli*. As an important consequence, the comparison of cell surface labeling under different conditions have to be performed with the same CyDye. Nevertheless, it is not possible to know a priori which of the three dyes will give the most accurate information.

3. *S. aureus* membranes are resistant to classical lysis protocols, like sonication, phenol treatment…and protein extraction requires a first step to weaken the membranes, by incubating the cells in the presence of lysostaphin. Lysostaphin is a glycylglycine endopeptidase, an enzyme that cleaves the cross-linking pentaglycine bridges of the cell wall peptidoglycan of some *Staphylococci*. Proteins from gram negative cells are often directly extracted either by a trizol treatment, or even by incubating these bacteria with the sample buffer [14, 15].
4. The washing step with cold 70% acetone is important, since remaining pink coloration may affect the IEF isofocalization.
5. The pH range of the IPG strip used for *S. aureus* proteome fractionation is 4–7, since most *S. aureus* proteins migrate within this range. Different pH ranges are available, the IPG strip used has to be adapted to the bacterial strain studied.
6. The gel is protected by a thin plastic band that should be removed with pliers. Carefully adding the strip onto the sample requires some practice.
7. IEF should always be started at low voltage in order to remove salts.
8. SDS-PAGE can be performed with more classical solutions (classical Tris-Glycine running buffer…). Another acrylamide mixture and running buffer was used that has been shown to improve the quality of spot separation [20].
9. This step has to be carefully undertaken, as confusions when assigning the CyDyes revealed spots on the colloidal blue stained image will lead in identification of unrelated spot slices.

Acknowledgments

I want to thank Laurence Choulier for critical reading of the manuscript. This work was supported by the Centre National de la Recherche Scientifique (CNRS).

References

1. Wagner EG, Romby P (2015) Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. Adv Genet 90:133–208
2. Vogel J, Papenfort K (2006) Small non coding RNAs and the bacterial outer membrane. Curr Opin Microbiol 9:605–611
3. Valentin-Hansen P, Johansen J, Rasmussen AA (2007) Small RNAs controlling outer membrane porins. Curr Opin Microbiol 10:152–155
4. Klein G, Raina S (2015) Regulated Control of the assembly and diversity of LPS by noncoding sRNAs. Biomed Res Int 2015:153561
5. Beisel CL, Storz G (2010) Base pairing small RNAs and their roles in global regulatory networks. FEMS Microbiol Rev 34:866–882
6. Mandin P, Guillier M (2013) Expanding control in bacteria: interplay between small RNAs and transcriptional regulators to control gene expression. Curr Opin Microbiol 16:125–132
7. Becher D, Hempel K, Sievers S, Zuhlke D, Pane-Farre J, Otto A, Fuchs S, Albrecht D, Bernhardt J (2009) A proteomic view of an important human pathogen-towards the identification of the entire *Staphylococcus aureus* proteome. PLoS One 4:e8176
8. Sabarth N, Lamer S, Zimny-Arndt U, Jungblut PR, Meyer TF, Bumann D (2002) Identification of surface proteins of *Helicobacter pylory* by selective biotinylation, affinity purification, and two-dimensional gel electrophoresis. J Biol Chem 277:27896–27902
9. Ythier M, Resch G, Waridel P, Panchaud A, Gfeller A, Majcherczyk P, Quadroni M, Moreillon P (2012) Proteomic and transcriptomic profiling of *Staphylococcus aureus* surface LPXTG-proteins: correlation with *agr* phenotypes and adherence phenotypes. Mol Cell Proteomics 11:1123–1139
10. Savas JN, Stein BD, CC W, Yates JR 3rd (2011) Mass spectrometry accelerates membrane protein analysis. Trends Biochem Sci 36:388–396
11. Gilmore JM, Washburn MP (2010) Advances in shotgun proteomics and the analysis of membrane proteomes. J Proteomics 73:2078–2091
12. Hagner-McWhirter A, Winkvist M, Bourin S, Marouga R (2008) Selective labelling of cell surface proteins using CyDye DIGE fluor minimal dyes. J Vis Exp 21:945
13. Khemiri A, Galland A, Vaudry D, Chan Tchi Song P, Vaudry H, Jouenne T, Cosette P (2008) Outer membrane proteomic maps and surface exposed proteins of *Legionella pneumophila* using cell fractionation and fluorescent labelling. Anal Bioanal Chem 390:1861–1871
14. Hammann P, Parmentier D, Cerciat M, Reimegard J, Helfer AC, Boisset S, Guillier M, Vandenesch F, Wagner EGH, Romby P, Fechter P (2014) A method to map changes in bacterial cell surface composition induced by regulatory RNAs in *Escherichia coli* and *Staphylococcus aureus*. Biochimie 106:175–179
15. Carnielli CM, Artier J, de Oliviera JC, Novo-Mansur MT (2017) *Xanthomonas citri subsp. citri* surface proteome by 2D-DIGE: ferric enterobactin receptor and other outer membrane proteins potentially involved in citric host interaction. J Proteomics 151:251–263
16. Mayrhofer C, Krieger S, Allmaier G, Kerjaschki D (2006) DIGE compatible labelling of surface proteins on vital cells *in vitro* and *in vivo*. Proteomics 11:1123–1139
17. Hörmann K, Stukalov A, Müller AC, Heinz LX, Superti-Furga G, Colinge J, Bennett KL (2016) A surface biotinylation strategy for reproducible plasma membrane protein identification and tracking of genetic drug-induced alterations. J Proteome Res 15:647–658
18. Gesslbauer B, Poljak A, Handwerker C, Schüler W, Schwendenwein D, Weber C, Lundberg U, Meinke A, Kungl AJ (2012) Comparative membrane proteome analysis of three *Borrelia* species. Proteomics 12:845–858
19. Ge Y, Rikihisa Y (2007) Surface exposed proteins of *Ehrlichia chaffeensis*. Infect Immun 75:3833–3841
20. Schagger H, von Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 166:368–379

Chapter 9

Fluorescence-Based Methods for Characterizing RNA Interactions In Vivo

Abigail N. Leistra, Mia K. Mihailovic, and Lydia M. Contreras

Abstract

Fluorescence-based tools that measure RNA-RNA and RNA-protein interactions in vivo offer useful experimental approaches to probe the complex and dynamic physiological behavior of bacterial RNAs. Here we document the step-by-step design and application of two fluorescence-based methods for studying the regulatory interactions RNAs perform in vivo: (i) the in vivo RNA Structural Sensing System (iRS^3) for measuring RNA accessibility and (ii) the trifluorescence complementation (TriFC) assay for measuring RNA-protein interactions.

Key words RNA-RNA interaction, RNA-protein interaction, In vivo fluorescence assay, Hybridization efficacy, RNA accessibility, Complementation assay, RNA regulator, Protein regulator, Target network

1 Introduction

Recent years have been marked by identification and characterization of noncoding RNAs (ncRNAs), such as small RNAs (sRNAs) [1], in all types of bacteria, ranging from model strains such as *E. coli* to extremophiles like *D. radiodurans* [2] and other biotechnologically relevant organisms [3, 4]. The number of tools to probe RNA structure has grown in response to a deeper understanding of ncRNA roles in gene expression, regulatory cascades, and control of critical metabolic and cellular processes [5–7]. Nucleotide-specific chemical modification-based techniques like in vitro and in vivo SHAPE and DMS footprinting have uncovered structure-function paradigms such as sRNA-protein binding footprints [8], riboswitch conformation changes [9], and principles for designing synthetic RNA regulator parts [10]. However, these methods are limited because they do not represent true collective-nucleotide behavior of interactions and thus may overlook

Abigail N. Leistra and Mia K. Mihailovic contributed equally to this work.

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_9,

weak or low-frequency interactions [5]. The two techniques detailed in this work pose an advantage in this area by (i) reproducing RNA-RNA hybridization to capture regional intermolecular interaction preferences and (ii) directly measuring interactions between RNAs and proteins to elucidate physiological reasons for inaccessibility, i.e., structure versus protein occlusion.

This chapter details the design, preparation, and execution of two fluorescence-based methods for monitoring RNA interactions in vivo. The first measures the propensity of unique antisense RNAs (asRNAs) to interact with distinct regions within a target RNA of interest to offer insights into dynamic molecular behavior (Fig. 1a, b) [5]. Specifically, by regionally probing a target RNA in continuous 9–16 nucleotide-long segments, a profile of the RNA's hybridization landscape can be built. This technique, called the in vivo RNA Structural Sensing System (iRS^3) has been used to investigate hybridization landscapes of many RNA types, including *E. coli* sRNAs, mRNAs, and tRNAs as well as non-native RNAs known to host complex interactions. In this way, the iRS^3 captures structure-function information. Particularly, it has demonstrated sensitivity to known RNA-RNA and RNA-protein interactions [11]. The second technique probes RNA-protein interactions via a trifluorescence complementation assay (Fig. 1c, d). Making use of

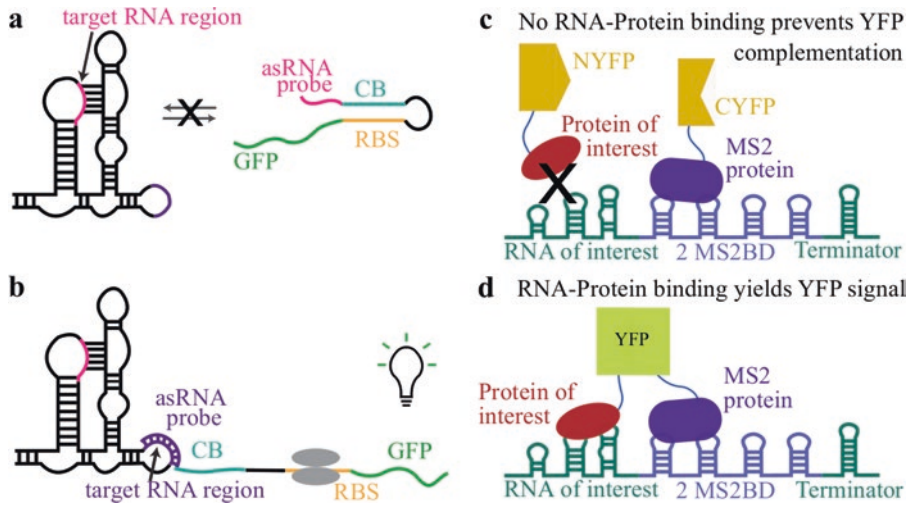


Fig. 1 Schematic of RNA hybridization (**a**, **b**) and RNA-protein probing (**c**, **d**) methods. (**a**, **b**) The iRS^3 assay evaluates the likelihood of an RNA region to engage in RNA-RNA interactions with its asRNA. If the region is inaccessible to the asRNA probe within the iRS^3 transcript, the ribosomal binding site (RBS) remains sequestered by the cis-blocking (CB) region (**a**). If the asRNA interacts with its target region, the hairpin loop of the iRS^3 is disrupted, and the ribosomal binding site is available for GFP translation initiation (**b**). (**c**, **d**) The trifluorescence complementation (TriFC) assay probes RNA-protein interactions by making use of three fusion constructs: the RNA of interest fused to a MS2 binding domain (MS2BD), the protein of interest linked to NYFP, and the MS2 protein linked to CYFP. Only in the presence of a RNA-protein interaction do all three components interact and allow for complementation of NYFP and CYFP to produce yellow fluorescence (**d**)

the high-affinity MS2 protein-MS2 binding domain interaction [12] and the well-established NYFP-CYFP split yellow fluorescent protein system [13], this assay has been used to quantify direct sRNA-protein and mRNA-protein interactions in vivo. In particular, the TriFC assay has quantified regulatory sRNA-protein interactions, capturing differential effects of protein active site mutations on the extent of YFP complementation [14]. Additionally, a dual-fluorescence variation of this assay has been used to test upward of 75 possible mRNA targets for protein binding [15].

We foresee these techniques having broad impact on our understanding of the in vivo functionality of sRNAs. For instance, the iRS^3 approach can be used to map and infer the in vivo functionality of an sRNA's alternative structures and track their relevance to regulatory function under various environmental conditions. Given that many sRNAs are known to confer bacterial virulence [16], this could provide a strong basis for targeted antimicrobial design. The TriFC assay could be applied to screen sRNAs for association with Hfq or any other regulatory proteins. Given the challenges of predicting RNA-protein interactions [17], methods amenable to large-scale screening should prove useful.

The iRS^3 method is capable of measuring regional hybridization efficacy of RNAs at basal levels; however, the TriFC assay requires overexpression of the RNA and protein involved. As with any overexpression system, questions of relevance remain, but with proper negative and positive controls, strong, physiologically relevant conclusions can be made.

2 Materials

2.1 iRS^3 General Materials and Reagents

1. Disposable pipette tips.
2. PCR tubes.
3. Nuclease-free water.
4. 1.7 mL polypropylene microtubes.
5. Thermocycler.
6. Vortex.
7. Microcentrifuge.
8. Luria Broth.
9. Agar.
10. Petri dishes.
11. 2.5–1000 μL pipette set.
12. 25 mL culture tubes.
13. Kanamycin stock solution: 10 or 100 mg/mL in nuclease-free water.

14. UV spectrophotometer.
15. Incubator, set to 37 °C with shaking capability.

2.2 iRS³ Target RNA Insertion

1. Gibson Primers Designed using NEBuilder Assembly Tool.
2. pO-iRS³GG or pN-iRS³GG vector (Addgene plasmids 98589 and 98858, respectively).
3. Deoxynucleotide Solution Mix (NEB).
4. Phusion High-Fidelity (HF) DNA polymerase (2000 U/mL, NEB).
5. 5× Phusion HF buffer (NEB).
6. DpnI (20,000 U/mL, NEB).
7. PCR DNA Purification Kit.
8. Gibson Assembly master mix (NEB or made according to [18]).
9. Electrocompetent *E. coli* cells for cloning (Turbo High Efficiency competent *E. coli*).
10. Plasmid miniprep kit.

2.3 iRS³ Probe asRNA Insertion

1. 50 ng pO-iRS³GG or pN-iRS³GG vector.
2. Linker buffer: 50 mM Tris–HCl pH 8.0, 100 mM NaCl, 1 mM EDTA.
3. Primers diluted in linker buffer to 100 μM.
4. 10× T4 DNA ligase buffer (NEB).
5. T4 DNA ligase (400,000 U/mL, NEB).
6. Nuclease-free water.
7. Bsmb1/Esp3I (10 U/μL, NEB).
8. 0.025 μm nitrocellulose membrane filters.
9. Electrocompetent *E. coli* cells for cloning (Turbo High Efficiency competent *E. coli*).
10. 4-Chloro-DL-phenylalanine.

2.4 iRS³ Experiment

1. K-12 MG1655 *E. coli* or another strain of choice to conduct the assay.
2. Sterile 250 mL Erlenmeyer flasks OR Sterile 200 μL 96-well black clear-bottom plates.
3. Arabinose stock solution: 20% w/v, filter sterilized through 0.22 μm filter.
4. Anhydrotetracycline (aTc) stock solution: 100 μg/mL, filter sterilized through 0.22 μm filter.
5. 1× phosphate-buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4, sterilized through 0.22 μm filter.

6. 20 mL syringes with Leur lock tip.
7. Syringe-tip filters (0.22 μm PVDF 30 mm diameter).
8. 5 mL polystyrene round-bottomed tubes.
9. BD FACSFlow sheath fluid.
10. BD FACSCalibur Flow cytometer and BD CellQuest Pro software.

2.5 TriFC General Materials and Reagents

1. Disposable pipette tips.
2. PCR tubes.
3. 1.7 mL polypropylene microtubes.
4. Luria Broth.
5. Agar.
6. Petri dishes.
7. Kanamycin stock solution: 10 or 100 mg/mL in nuclease-free water.
8. Carbenicillin stock solution: 50 mg/mL in 1:1 nuclease-free water to ethanol.
9. IPTG stock solution (optional): 100 mM in nuclease-free water.
10. 2.5–1000 μL pipette set.
11. Sterile 25 mL culture tubes.
12. Sterile 250 mL Erlenmeyer flasks.
13. Vortex.
14. Microcentrifuge.
15. Thermocycler.
16. Incubator, set to 37 °C with 200 rpm shaking capability.

2.6 TriFC Cloning Supplies

1. Deoxynucleotide Solution Mix (NEB).
2. Nuclease-free water.
3. Phusion High-Fidelity (HF) DNA Polymerase (2000 U/mL, NEB).
4. 5× Phusion HF buffer (NEB).
5. Taq Polymerase (5000 U/mL, NEB).
6. 10× ThermoPol Buffer (NEB).
7. pTriFC (or pTriFC-mStrawberry) and pMS2-CYFP vectors (Addgene plasmids 98584 or 98848 and 98587, respectively).
8. Control pTriFC and pMS2-CYFP vectors (Addgene plasmids 98586, 98585, and 98588).
9. DpnI restriction enzyme (20,000 U/mL, NEB).
10. Gibson Assembly master mix (NEB or made according to [18]).

11. Electrocompetent *E. coli* cells for cloning (NEB 5-alpha or Turbo High Efficiency competent *E. coli*).
12. 0.025 μM nitrocellulose membrane filters.
13. Genomic DNA purification kit.
14. Miniprep kit.
15. PCR DNA purification and/or gel extraction kit.
16. UV spectrophotometer.

2.7 TriFC Experiment

1. K-12 MG1655 *E. coli* or another strain of choice to conduct the assay.
2. 1× phosphate-buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4, sterilized through 0.22 μm filter.
3. Nanopure water.
4. 20 mL syringes with Leur lock tip.
5. Syringe-tip filters (0.22 μm PVDF 30 mm diameter).
6. Biotek Cytation 3 Imaging Reader with Gen5 software.
7. Black 96-well clear flat bottom assay plates.
8. BD FACSCalibur Flow cytometer with BD CellQuest Pro software.
9. BD FACSFlow sheath fluid.
10. 5 mL polystyrene round-bottomed tubes (Note that **items 6** and **7** or **8–10** are required).

3 iRS³ Method

The iRS³ plasmid-based system can be used to investigate the propensity of a synthetic asRNA to interact with its complementary target region on distinct heterologously expressed RNA (pO-iRS³GG plasmid, Fig. 2a) or native RNA (pN-iRS³GG plasmid, Fig. 2b) in vivo, a measurement termed hybridization efficacy [11]. Specifically, the iRS³ system consists of a target region-specific asRNA probe upstream of a hairpin-forming loop and green fluorescent protein (GFP) reporter (Fig. 1a). This asRNA probe is specifically designed to be complementary to a target RNA region of interest. If, upon iRS³ expression, the synthetic asRNA probe is able to bind to its corresponding target RNA region, disruption of the hairpin loop causes exposure of the GFP ribosomal binding site and consequent GFP translation (Fig. 1b). This phenomenon allows the hybridization likelihood of a target RNA region to be quantified by fluorescence shift as measured using flow cytometry [5]. As the system directly mimics in vivo RNA-RNA interactions, results are believed to speak to the structural conformation and availability of RNA regions for regulatory purposes.

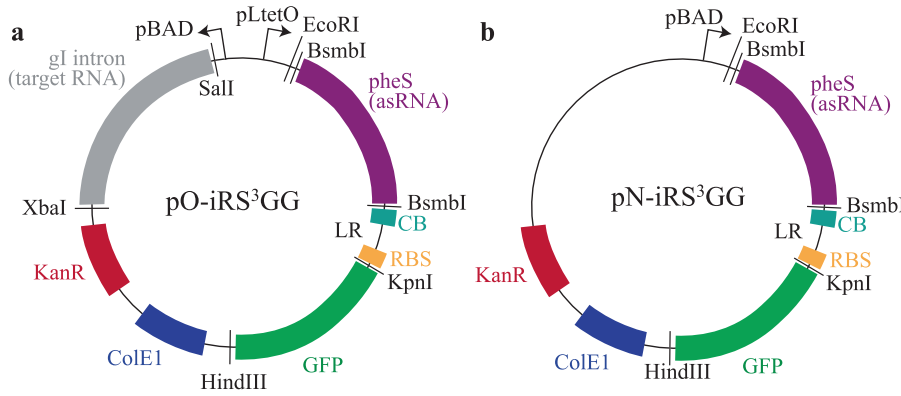


Fig. 2 Overexpression (pO-iRS^{3}GG) or and native (pN-iRS^{3}GG) iRS^3 assay plasmids. (**a**) The pO-iRS^{3}GG plasmid contains the gI intron (default target RNA) expressed by a pBAD promoter. The iRS^3 transcript, consisting of the pheS cassette (to be replaced by an asRNA), cis blocking region (CB), ribosomal binding site (RBS), then GFP, is under pLtetO control. (**b**) The pN-iRS^{3}GG plasmid differs from the O-iRS^3 in that (i) it lacks target RNA overexpression capabilities and (ii) the iRS^3 transcript expression is under pBAD promoter control

As illustrated in Fig. 3, the iRS^3 parent vectors (Fig. 2) must be modified for each unique RNA region being targeted. Depending on whether the RNA expression levels are heterologous or basal, necessary cloning will differ. In the case of heterologous target RNA expression, (i) the target RNA is cloned into the plasmid following the pBAD promoter to replace the default target RNA in pO-iRS^{3}GG (*Tetrahymena* group I, gI, intron) via Gibson Assembly (Fig. 3a). Depending on plasmid chosen, (ii) the asRNA probe is cloned in to follow either the pLtetO (pO-iRS^{3}GG) or pBAD (pN-iRS^{3}GG) promoter via a high-throughput Golden Gate cloning protocol in which the asRNA sequence replaces a 4-chloro-DL-phenylalanine negative selection cassette (pheS) (Fig. 3b). Once the plasmids have been (iii) transformed into the strain of interest (Fig. 3c), (iv) strains are cultured. During early log growth (OD ~ 0.2–0.4), (v) strains are separated into "uninduced" and "induced" samples, the latter in which expression of target RNA and iRS^3 system (pO-iRS^{3}GG) or iRS^3 system only (pN-iRS^{3}GG) is stimulated (Fig. 3d). At the environmental conditions of interest (Fig. 3e), (vi) fluorescence shift between uninduced and induced samples is evaluated via flow cytometry (Fig. 3f) and analyzed (Fig. 3g).

3.1 Rational Experimental Design

RNA regions with considerable overlap have been shown to exhibit significant differences in hybridization efficacy [11]. This is not surprising, considering that binding between single-stranded loops or linear segments of RNA has been implicated in RNA-RNA strand displacement [19]. These observations further support the notion that this hybridization-based system adequately mimics regulatory in vivo antisense-based RNA-RNA interactions, despite

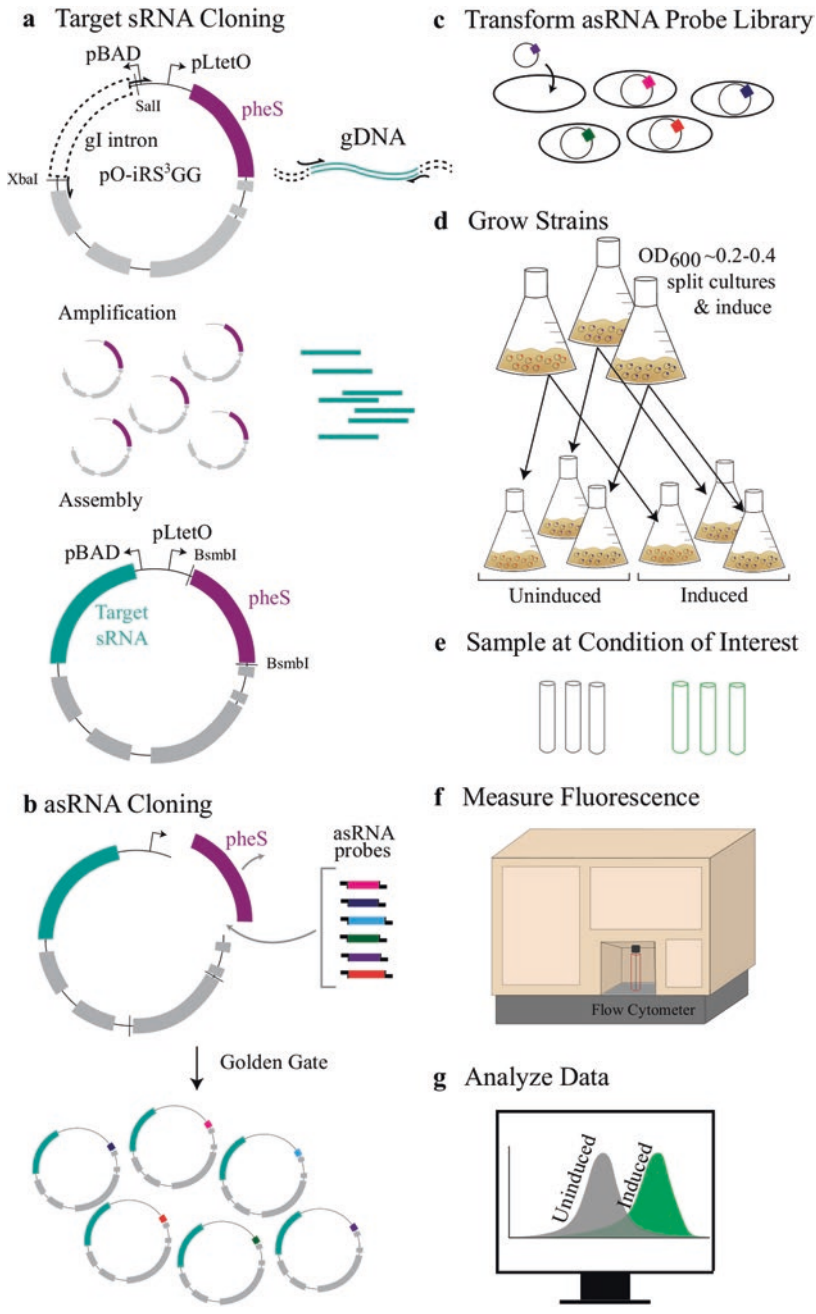


Fig. 3 Work flow for the iRS^3 assay. (**a**) If using pO-iRS^{3}GG, standard two-fragment Gibson cloning is used to amplify and insert the target RNA of interest to replace the default target RNA (gI intron). (**b**) For either pO-iRS^{3}GG or pN-iRS^{3}GG, the pheS selection cassette is replaced by desired asRNAs via Golden Gate cloning via BsmbI restriction digest to create a library of iRS^3 transcripts. (**c**–**g**) Experimental work flow. (**c**) Plasmids are transformed into an *E. coli* strain of choice and grown in triplicate. (**d**) Cultures volumes are split into 2 containers per replicate and the expression of iRS3 transcripts (pO-iRS^{3}GG and pN-iRS^{3}GG) and target RNAs (pO-iRS^{3}GG only) are induced for each representative strain. (**e**) At conditions of interest, cultures are sampled and (**f**) assayed via flow cytometry for fluorescence. (**g**) Finally, induced and uninduced fluorescence data are analyzed compared to an appropriate negative control

the bulkiness of the iRS^3 transcript. Thus, selection of target regions, especially in an uncharacterized RNA, is critical to obtaining valuable hybridization efficacy insights. To this end, we offer loose guidelines. Generally, target RNA regions under investigation should (i) be 9–16 nt long to mimic lengths of seed sequences and known RNA-RNA interactions [20, 21]. Additionally, target regions should be chosen to ensure cognate probe asRNAs (ii) show low complementarity to the genome (besides the target of interest) (i.e., top BLASTn [22] search hits with Expect (E) value ≥1) and (iii) contribute to iRS^3 transcript folding in a previously determined optimal free energy range ($-19.3 < \Delta G < -17.8$ kcal/mol) [11].

Many target region selection schemes will offer valuable insights to the hybridization landscape of an RNA molecule. A few are listed below.

1. Structural Prediction Based Design: Secondary structure predictions may be used to inform probe design. Specifically, we have designed experiments in which the standard deviation of the base pairing probability (as evaluated via Nupack [23]) of each target RNA region is minimized within our length constraint (9–16 nt).
2. Blind Design: If characterization information is sparse, it may be valuable to walk the RNA with overlapping probes, as probes with considerable overlap (≥6 nt) have shown vastly different propensity for hybridization [11]. This design method has been used in previously published work to blindly create a pool of potential asRNA probes [11]. Specifically, random-length regions (within our length constraint) were selected in a manner to represent all possible overlaps and cover the entire length of the RNA of interest. These regions were then filtered based on iRS^3 transcript folding energy (*see* iii under Rational Experimental Design above). A simpler alternative to this blind design would be the exclusion of overlap.

3.2 Target RNA Insertion

Overexpression of the target RNA may be useful if (i) evaluating the functional structure of non-native sRNAs or (ii) titratable control of RNA expression is desired. When using pO-iRS^{3}GG, Gibson Assembly is recommended for insertion of the DNA encoding for the desired target RNA.

1. We recommend designing primers for two-fragment Gibson Assembly of the designated target RNA and pO-iRS^{3}GG using NEBuilder Assembly Tool. Specifically, preferences may be set to Product/Kit = E2611 Gibson Assembly Master Mix, No. of Fragments = 2–3, Total Construct Length = less than 10 kb, Min. Overhang Length (nt) = 20, PCR Product Group = Phusion, PCR Product—Phusion High-Fidelity DNA Polymerase (HF Buffer), PCR Primer Conc. (nM) = 500

(standard), Min. Primer Length (nt) = 18. An example cloning scheme for the replacement of default *Tetrahymena* gI intron with LtrB gII intron [11] is depicted in Fig. 4 and corresponding primer sequences listed in Table 1.

2. Assemble PCR amplification reaction for DNA encoding the target RNA of interest from *E. coli* genomic DNA, previously prepared plasmids, or synthesized DNA fragments using forward and reverse insert primers, with overhang complementary to the vector backbone (*see* example in Fig. 4). Add 50 ng genomic DNA, 10 μM forward and reverse insert primers (2.5 μL each, diluted in nuclease-free water), 10 mM dNTPs (1 μL), 5× Phusion High-Fidelity (HF) buffer (10 μL), Phusion HF DNA polymerase (0.5 μL), and nuclease-free water to 50 μL. Genomic DNA can be obtained with a genomic DNA purification kit or by boiling a colony (single colony, diluted in 50 μL of nuclease-free water, heated at 96 °C for 5 min); 2 μL of this reaction can be used above as genomic DNA template.
3. Assemble PCR amplification reaction for pO-iRS^{3}GG backbone with forward and reverse backbone primers (*see* example in Fig. 4) designed above. Add 50 ng of parent vector (pO-iRS^{3}GG), paired forward and reverse backbone primers (2.5 μL of each, diluted in nuclease-free water), 10 mM dNTPs (1 μL), 5× Phusion High-Fidelity (HF) buffer (10 μL), Phusion HF DNA polymerase (0.5 μL), and nuclease-free water to 50 μL.
4. Cycle all PCR reactions as follows: (i) 98 °C for 30 s (ii) 25 cycles of 98 °C for 10 s, 3° above lowest melting temperature of primer pairs (excluding overhang region) for 30 s, 72 °C for 2.5 min (vector backbone) or 30 s/kb (iii) final extension 72 °C for 10 min (iv) hold at 4 °C (*see* **Note 1**).
5. Digest methylated DNA by adding 1.5 μL DpnI directly to PCR products, incubating at 37 °C for 90 min, then heat deactivating at 80 °C for 20 min.
6. Agarose gel electrophoresis is recommended to confirm the size and amplification of backbone and insert. Using approximately 5.0 μL of each reaction (with EZ Vision or loading dye and ethidium bromide), check for proper band size and amplification specificity (pO-iRS^{3}GG backbone at ~4.5 Kb).
7. Clean-up PCR reactions per PCR DNA purification kit directions. Measure DNA concentrations via spectrophotometry.
8. Insert DNA of target RNA via Gibson fragment assembly. Add 50–100 ng amplified vector backbone, 2× amplified target RNA insert (5× if insert is smaller than nucleotides), and Gibson Assembly Master Mix (10 μL), and nuclease-free water to 20 μL. Additionally, perform the assembly with a negative control omitting the insert: 50–100 ng of amplified vector, Gibson Assembly Master Mix (10 μL), and nuclease-free water to 20 μL. Incubate samples at 50 °C for 45 min.

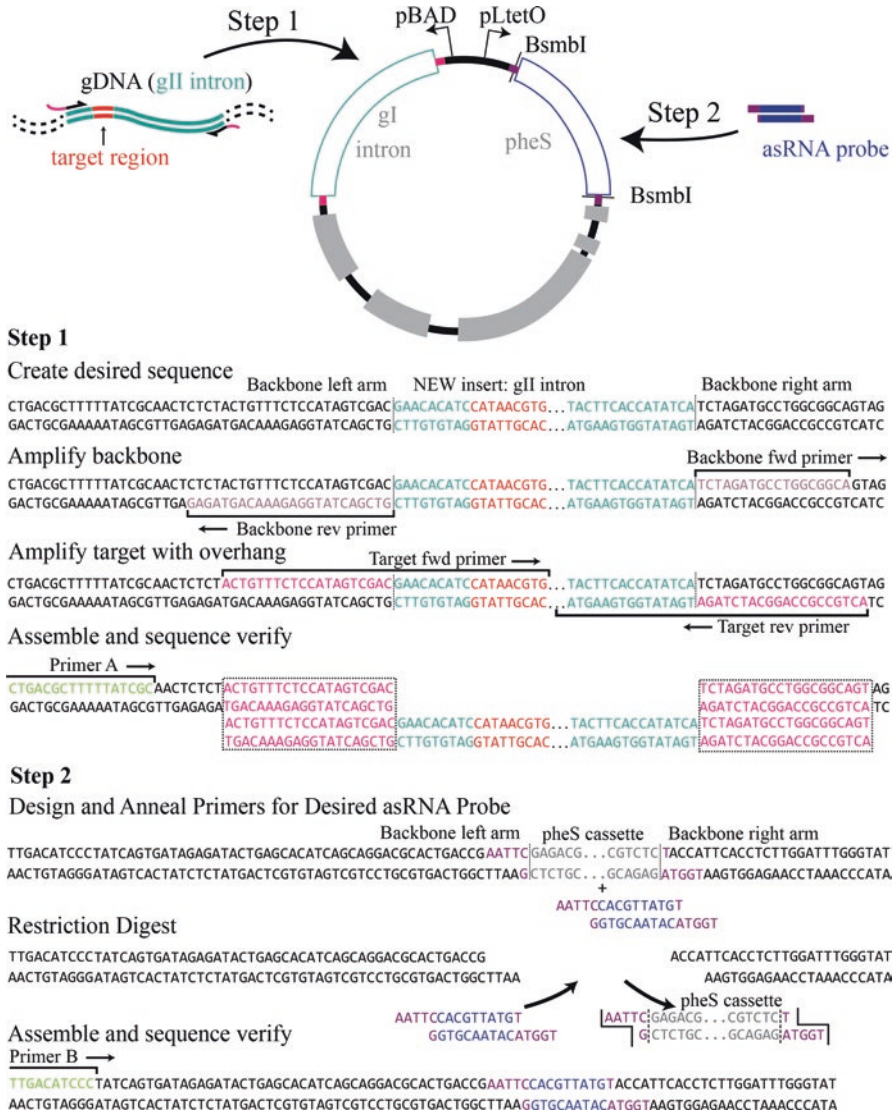


Fig. 4 Example cloning schemes for insertion of target RNA and asRNA probe into pO-iRS^{3}GG. Cloning of target RNA (**step 1**) and a representative cognate asRNA probe (**step 2**) into pO-iRS^{3}GG. Necessary overhangs for target RNA cloning via Gibson and asRNA probe cloning via Golden Gate are shown in pink and purple, respectively. The target region within the DNA of target RNA sequence (turquoise) is shown in orange. DNA corresponding to the asRNA probe, targeting desired region (orange), is pictured in blue. Target RNA sequences and primers corresponding to the cloning scheme example are listed in Table 1

9. Dilute products fourfold in nuclease-free water and use 1 μL for *E. coli* electroporation using standard protocols. After recovery period has elapsed, plate on LB agar with 50 μg/mL kanamycin and let grow overnight at 37 °C (Day 1).
10. Continue to sequencing if significantly greater number of colonies is observed on sample plate than on negative control plate (Day 2). Grow one to two colonies to saturation

Table 1
Example RNA sequences, corresponding Gibson Assembly primers, and recommended sequence confirmation primers for pO-iRS^{3}GG and pN-iRS^{3}GG

Sequence or primer name	Sequence (5′ overhangs in lower case)
Target RNA	GAACACATCCATAACGTG...TACTTCACCATATCA
Target RNA region of interest	CATAACGTG
asRNA probe sequence	CACGTTATG
Backbone Rev. primer	GTCGACTATGGAGAAACAGTAGAG
Backbone Fwd primer	TCTAGATGCCTGGCGGCA
Insert (Target RNA) Fwd primer	actgtttctccatagtcgacGAACACATCCATAACGTG
Insert (Target RNA) Rev. primer	actgccgccaggcatctagaTGATATGGTGAAGTAGGGAG
Primer A	CCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGC
Primer B	CGAGTCCCTATCAGTGATAGAGATTGACATCCC

(overnight, 37 °C) in separate 5 mL tubes of LB supplemented with 50 μg/mL kanamycin per unique target RNA insert (Day 2).

11. The following day, extract plasmid DNA of 2–3 mL saturated culture using plasmid miniprep kit. Measure purified DNA concentration via spectrophotometry. Prepare and submit sequencing samples as prescribed by local facility (Day 3). A forward pBAD primer (Primer A) is recommended for sequencing inserted target RNAs (binding location and sequence shown in Fig. 4 and Table 1).

3.3 asRNA Probe Insertion

Both plasmid systems that can be used for experimentation (pO-iRS^{3}GG and pN-iRS^{3}GG) are Golden Gate cloning compatible. As previously described [24], the cloning design supports elimination of strain sensitivity to 4-chloro-DL-phenylalanine once the pheS selection cassette is replaced by sequences complementary to target RNA regions of interest, referred to as "asRNA probes" [11]. Cloning schemes for the insertion of an example asRNA probe targeting a region in the heterologous and "native" gII intron are shown in Figs. 4 and 5, respectively, and corresponding primer sequences listed in Table 1.

1. asRNA probe primers must be designed to contain the full forward and reverse complement of the target RNA region flanked by BsmbI-compatible sites. Specifically, (i) the forward asRNA Probe Primer should include AATTC(reverse complement of target region sequence)T and (ii) the Reverse asRNA Probe Primer should include TGGTA(target region sequence)G.

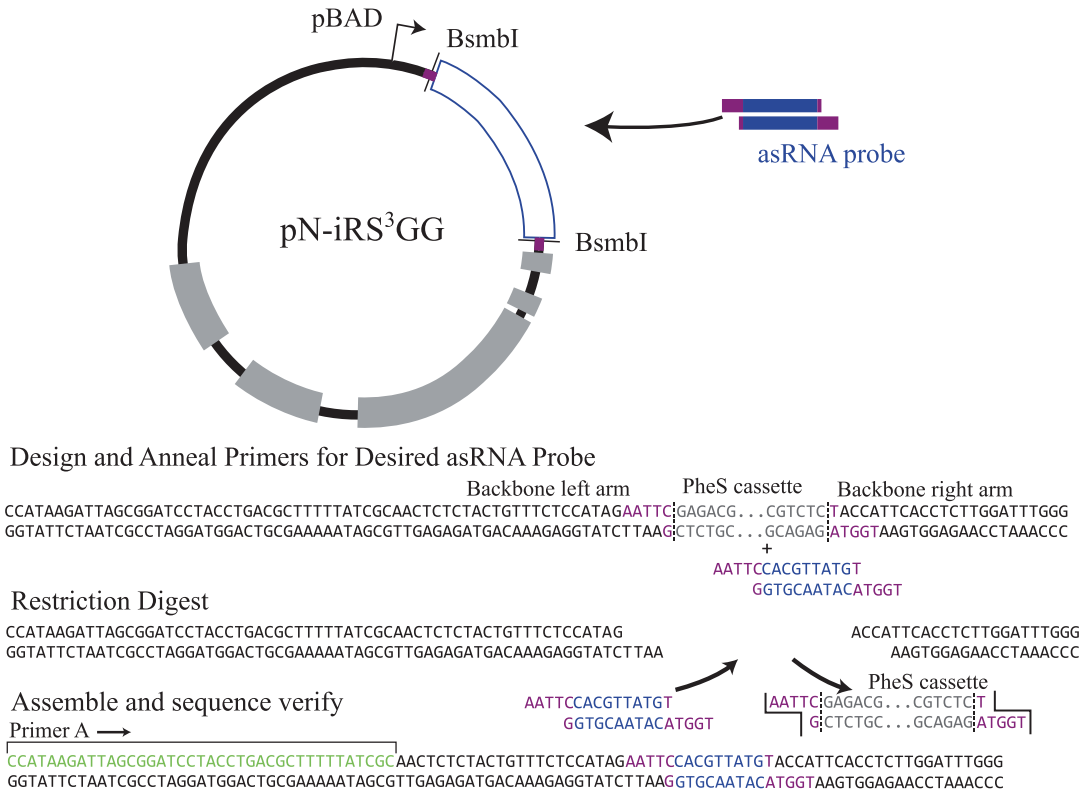


Fig. 5 Example cloning schemes for insertion of asRNA probe into pN-iRS³GG. Cloning of an asRNA probe targeting a region within a natively expressed target RNA into pN-iRS³GG. Necessary restriction sites and DNA corresponding to the asRNA probe are pictured in purple and blue, respectively. Necessary Golden Gate overhang and consecutive cloning steps are pictured. Target RNA sequences and primers corresponding to the cloning scheme example are listed in Table 1

2. In the case that (i) all samples needed for an experiment are split over multiple days or (ii) samples represent multiple unique environmental conditions, control iRS³ transcripts should be used to account for instrument shift or environment-related fluorescence changes, respectively. Specifically, the controls should represent the largest possible fluorescence range. The low-fluorescence "scramble" iRS³ transcript should (i) contain a probe which represents the average length of the target regions within the experimental set and (ii) contain a random sequence, i.e., a "scramble probe," with limited sequence similarity to the genome (i.e., top BLASTn [22] search hits with E value ≥1). The high-fluorescence "open RBS" iRS³ transcript should: (i) contain mutations in the cis-blocking (CB) region of the iRS³ transcript to render the RBS (seen in Fig. 1a) consistently accessible and (ii) contain a "scramble probe" asRNA (*see* low range iRS³, above). Sequences for control iRS³ transcripts previously used can be found in Table 2.

Table 2
Standard asRNA probe and CB region sequences for iRS³ controls

Control name	asRNA sequence	CB sequence
Scramble	CAGCGACAATATCGT	TACCATTCACCTCTTGGAT
Open RBS	CAGCGACAATATCGT	GCATAAATTAGGGAGTCAA

3. Anneal primers. First, dilute primers to 100 μM in linker buffer: 50 mM Tris–HCl pH 8.0, 100 mM NaCl, 1 mM EDTA. Combine 10 μL of forward and reverse primers in unique PCR tubes for each asRNA probe to yield 20 μL total. Thermocycle as follows: (i) 95 °C for 2 min, (ii) 52 °C for 10 min, (iii) hold at 4 °C. Meanwhile, dilute pO-iRS³GG or pN-iRS³GG to 50 ng/μL in nuclease-free water.
4. Insert asRNA probe via Golden Gate cloning. In PCR tubes, add diluted (50 ng/μL) vector (1 μL), annealed primers (2 μL of 100 μM solution), 10× T4 DNA Ligase Buffer (NEB) (1 μL), nuclease-free water (3 μL), 400 U T4 DNA Ligase (NEB) (1 μL), 20 U BsmbI (NEB) (2 μL) (*see* **Note 2**). Incubate at 37° for 45 min.
5. Transform Golden Gate reaction. First, desalt reaction for 20 min on 0.025 μm nitrocellulose membrane filters. Electroporate approximately 5 μL into *E. coli* using standard protocols. Upon outgrowth, plate onto LB agar plates supplemented with 50 μg/mL kanamycin and 2 g/L 4-chloro-DL-phenylalanine. These selection markers will not facilitate growth of bacteria lacking kanamycin resistance or maintaining the pheS cassette, respectively. Let grow overnight at 37 °C (Day 1) (*see* **Note 3**).
6. On the following day, confirm insertion of the asRNA probe. Grow two colonies *for each unique asRNA cloned* to saturation (overnight, 37 °C) in separate 5 mL tubes of LB supplemented with 50 μg/mL kanamycin (Day 2).The next day, extract plasmid DNA of 2–3 mL saturated culture using a plasmid miniprep kit. Measure purified DNA via spectrophotometry. Prepare and submit sequencing samples as prescribed by local facility (Day 3). Forward primers corresponding to iRS³ asRNA promoters (pBAD for pN-iRS3GG (Primer A) and pLtetO for pO-iRS3GG (Primer B)) are recommended (binding locations shown in Figs. 4 and 5, respectively, and sequences listed in Table 1).

3.4 iRS³ Experiment

1. Select an experimental strain. The iRS³ system was developed and exclusively utilized in *E. coli* K-12 MG1655; however, no predetermined hindrances to using this system in other *E. coli* strains exist (*see* **Note 4**). We foresee value in performing iRS³

experiments in genomic sRNA-knockout strains to allow titratable control of sRNA expression that may have an impact on the hybridization landscape. Each strain containing a piRS^{3}GG (O- or N-) that targets a unique RNA region should have at least biological triplicate representation. If splitting samples into multiple experiments, fluorescence shift of strains containing control iRS^3 plasmids should be evaluated at every sampling instance in order to account for instrument shift (scramble and open RBS).

2. When the experimental strain has been selected, transform all confirmed iRS^3 plasmids via standard $CaCl_2$ transformation protocols (or electroporation if desired). Plate each strain on unique kanamycin-containing (50 μg/mL) LB agar plates. Grow at 37 °C overnight.
3. Two distinct cell culture schemes have been successfully used for iRS^3 experimentation and will influence the preparation of overnights with experimental strains. Specifically, cultures supplemented with 50 μg/mL kanamycin can be grown in (i) 40 mL volumes in 250 mL shake flasks (split into 20 mL at induction) [5] or (ii) 200 μL volumes in 200 μL 96-well black clear-bottom plates (split into 100 μL at time of induction) [11] (*see* **Note 5**). Plates are recommended when more than 16 unique strains will be sampled at once, to limit required incubator space and maximize efficiency of seeding and inducing (as it supports use of multiwell pipettes) (*see* **Notes 6** and **7**). Once culturing method has been selected, overnight culturing volumes can be adjusted according to experimental culturing (i.e., 5 mL for flasks and 100 μL for plates). Grow 3 unique colonies from each strain to saturation at 37 °C, 200 rpm overnight (Day 1).
4. Prepare all necessary materials for iRS^3 experiment (Day 1). Specifically, prepare in advance (i) 20% w/v arabinose, (ii) 100 μg/mL aTc, and (iii) LB supplemented with 50 μg/mL kanamycin (*see* **Notes 8** and **9**).
5. On Day 2, inoculate flasks (40 mL LB) or plates (200 μL LB) with saturated culture by adding 1% of total container volume. Shake-incubate at 37 °C and 200 rpm.
6. After approximately 2 h, in early exponential phase (OD ~ 0.2–0.4), split each culture volume into two equal volumes—to serve as uninduced and induced samples. Induce expression of iRS^3 transcript and, if applicable, the target RNA (pO-iRS^{3}GG only) for *designated induced* samples only. For pN-iRS^{3}GG, induce with 20% arabinose for 0.8% final concentration in culture; for pO-iRS^{3}GG, induce with 20% arabinose for 0.8% final concentration in culture and 100 μg/mL aTc for 100 ng/mL final concentration in culture. Continue to shake-incubate at 37 °C and 200 rpm (*see* **Note 10**).

7. At desired sampling OD (at least ~30 min post-induction and with OD < ~2), begin sampling for flow cytometry. Resuspend a small volume of each culture (~1–20 μL, depending on OD at sampling time) in 1 mL of 1× PBS (in 5 mL polystyrene round-bottomed tubes) to achieve a concentration of ~10^7 cells/mL. If testing the strain fluorescence at multiple ODs, return samples to shake-incubator. Evaluate green fluorescence of samples using a flow cytometer (530/30 nm band pass filter on the Benton Dickinson FACSCalibur). Collect fluorescence data of sample until representation of >150,000 active cells has been achieved (*see* **Note 11**).

3.5 Data Analysis

As fluorescence results are expected have a non-normal distribution, median fluorescence is assumed representative of the population. The hybridization efficacy of a target RNA region is defined as the average logarithm of the ratio of the induced to uninduced (background) fluorescence. Because each unique induced sample has a corresponding uninduced sample originating from the same biological replicate, statistics are performed on the ratio (as opposed to each uninduced and induced fluorescence individually) and uncertainty is propagated through to the logarithm of the ratio. Two normalization schemes may be useful for comparing hybridization efficacies to draw conclusions about structure-function relationships (i) within the same molecule when probed under unique growth/environmental conditions which are likely to influence target RNA abundance or (ii) between unique molecules under equivalent growth/environmental conditions.

3.5.1 Intra-RNA Normalization

When evaluating the hybridization efficacy changes of an RNA under various conditions, it may be difficult to separate abundance effects from structure-function effects. This is particularly relevant to sRNAs because of their aptitude to differentially express between conditions associated with their function [25]. One approach for decoupling structure-function insights from abundance effects normalizes each sRNA dataset between the regions exhibiting the highest and lowest observed hybridization efficacies. Specifically, for each sRNA in each unique condition, hybridization data corresponding to the region exhibiting the highest and lowest observed fluorescence ratios will be transformed to 1 and 0, respectively. Uncertainty is propagated to account for errors in the minimum/maximum.

3.5.2 Inter-RNA Normalization

To compare absolute hybridization efficacies between RNAs within unique environmental conditions (i.e., experiments were performed on separate days), normalization to low-end fluorescence (scramble probe) and high-end fluorescence (open RBS) controls is critical. This normalization scheme accounts for instrument shift between days and facilitates comparison of absolute hybridization landscape differences between molecules.

4 TriFC Method

The TriFC assay detects direct RNA-protein binding interactions in vivo by simultaneously expressing three distinct fusions: the RNA of interest with a MS2 RNA binding domain (MS2BD) sequence, the protein of interest with NYFP, and the viral MS2 RNA binding protein with CYFP (Fig. 1c, d). The well-characterized MS2-MS2BD binding interaction is strong (K_d = 3 nM [12]) and renders a direct RNA-protein binding interaction responsible for NYFP-CYFP complementation and thus fluorescent signal generation. Several steps are required to apply the TriFC assay to any RNA-protein interaction of interest: (i) the RNA-MS2BD and protein-NYFP fusions are designed and cloned into the appropriate plasmid by Gibson Assembly, (ii) the two plasmid system is sequentially transformed into a desired *E. coli* strain, (iii) strains are seeded and grown for 18–48 h prior to (iv) fluorescence measurements by plate reader or flow cytometry and (v) data analysis. Construct design, recommended cloning procedures, experimental execution, and data analysis are documented for the TriFC assay below.

A variation of the TriFC assay has also been demonstrated; it includes a third fluorescent protein, mStrawberry, fused to the 3′ end of the RNA fusion construct, to monitor function of RNAs containing Shine-Dalgarno and translation start site sequences. In this way, the TriFC assay can be made particularly useful for measuring interactions between regulatory elements of mRNAs, like 5′ untranslated regions (UTRs), and suspected protein interaction partners. Outside of construct design, this dual-fluorescence variation of the TriFC assay differs from the original assay only in fluorescence measurement and data analysis procedures, which are documented in Subheading 5.

4.1 Construct Design

The TriFC RNA-protein interaction assay uses a two-plasmid system to express three fusion constructs: (i) protein-linker-NYFP, (ii) RNA-MS2BD, and (iii) MS2-linker-CYFP (Fig. 1c, d). The first two components are expressed on the same plasmid, pTriFC, under control of separate pLacO promoters (Fig. 6a). The third is expressed under pLacO control on a second plasmid, pMS2-CYFP (Fig. 6b). Importantly, this design pairs one universal plasmid with one modular plasmid that can be adapted to any RNA-protein interaction of interest. Based on previous successes, we do not anticipate modification the MS2-CYFP fusion being necessary for testing most RNA-protein interactions [14, 15, 26].

When adapting pTriFC for a specific protein-RNA interaction, tag location is a critical consideration as both the RNA and protein of interest need to retain biological function in their respective fused states. An excellent starting place for design of the protein-NFYP fusion is previously published tag placements for the protein

of interest. For instance, in design of a CsrA protein-NYFP fusion [14], the C-terminus of CsrA was linked to NYFP because this arrangement had been shown successful for HIS_6 tagged constructs (*see* **Note 12**) [27, 28]. Regarding linker selection, general guidelines for constructing protein fusions should be considered [29]. Three repeats of a glycine-serine linker (GGGGS) have been successfully used for the protein-NYFP fusion [14, 15]; this linker composition and length imparts sufficient flexibility to the construct (*see* **Note 13**). We foresee that this linker composition and length will be amenable to other protein-NYFP fusions as well (*see* **Note 14**).

Similarly, design of the RNA-MS2BD fusion requires consideration of where to place the MS2BD sequence relative to the RNA of interest and how many repeats of MS2BD to include. Two locations of MS2BD insertion have been demonstrated. In the first instance, the MS2BD sequence was placed in the 3′ portion of an sRNA of interest, between the last predicted non-terminator RNA structure and the terminator hairpin (Fig. 1c, d and Fig. 6a). This selection was made in light of known sRNA co-transcriptional folding patterns (*see* **Note 15**) [14]. Alternatively, a 5′ placement of the MS2BD sequence before an sRNA sequence can be considered [15], particularly if functional 3′ degradation products of the sRNA are known or suspected (*see* **Note 16**). Additionally, two repeats of the MS2BD sequence should be included for either placement, as this maximized RNA-protein interaction signal in the proof-of-concept study [14]. As such, we recommend placing two repeats of the MS2BD sequence (noted as 2MS2BD) flanking the 5′ end of the sRNA of interest if 3′ degradation products are anticipated. Similarly, we suggest inserting the 2MS2BD sequence 3′ of the sRNA of interest, just before the sRNA's terminator, if critical and complex co-transcriptional folding the sRNA is anticipated.

Once designs of the fusion constructs are established, the following series of control constructs should be included to ensure fluorescent signal is from true RNA-protein interactions: (i) pTriFC lacking the protein of interest (NYFP + RNA-2MS2BD present), (ii) pTriFC lacking the RNA of interest (protein-NYFP + 2MS2BD present), and (iii) pMS2-CYFP lacking the MS2 protein. The pTriFC and pMS2-CYFP plasmids, as well as the no-protein pTriFC, no-RNA pTriFC, and no-MS2 pMS2-CYFP control plasmids, are available from Addgene.

4.2 Cloning Procedures

The pTriFC plasmid is amenable to restriction digest and Gibson Assembly cloning approaches. The Gibson Assembly strategy is the focus of this section as it is the most amenable to the large-scale cloning required for screening applications. Cloning procedures are described in reference to Fig. 7, which illustrates an example cloning scheme for adding the CsrA protein and the CsrB

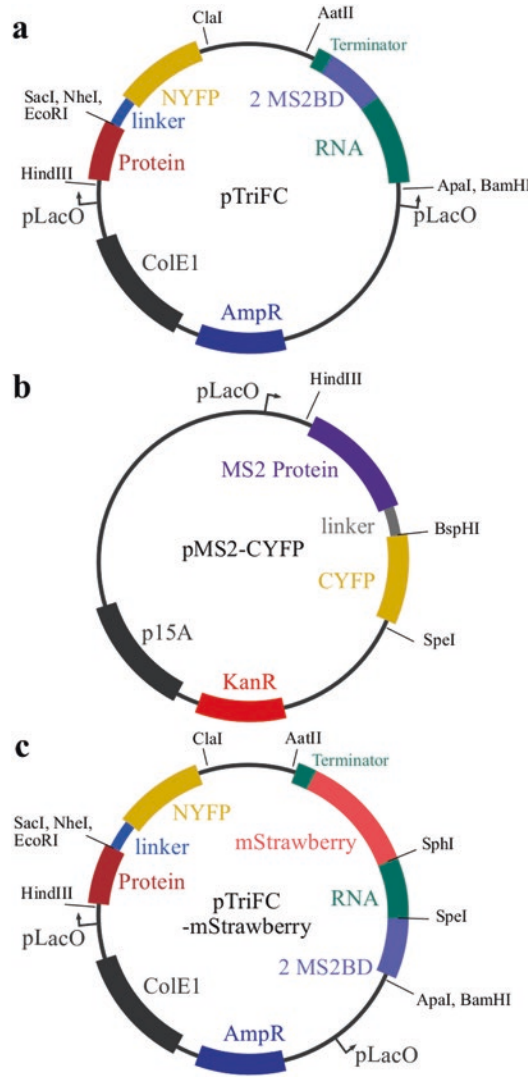


Fig. 6 Plasmids involved in TriFC assay. (**a**) The pTriFC plasmid expresses the protein of interest fused by a linker to NYFP from a pLacO promoter and the RNA fusion from a distinct pLacO promoter. Two repeats of the MS2BD sequence are inserted at the 3′ end of the RNA of interest, just before the RNA's terminator sequence. (**b**) The pMS2-CYFP plasmid contains the MS2-linker-CYFP fusion expressed by a pLacO promoter. (**c**) The pTriFC-mStrawberry plasmid is used in the dual fluorescence version of the complementation assay. It differs from pTriFC in design of the RNA fusion; two repeats of the MS2BD sequence are fused upstream of the RNA of interest and an mStrawberry sequence is fused just downstream of the RNA of interest to confirm functional expression

sRNA to pTriFC. We recommend designing primers for separate two-fragment Gibson Assembly reactions (using the NEBuilder Assembly Tool) to independently insert the protein of interest (Fig. 7, **step 1**) and the RNA of interest (Fig. 7, **step 2**) into pTriFC.

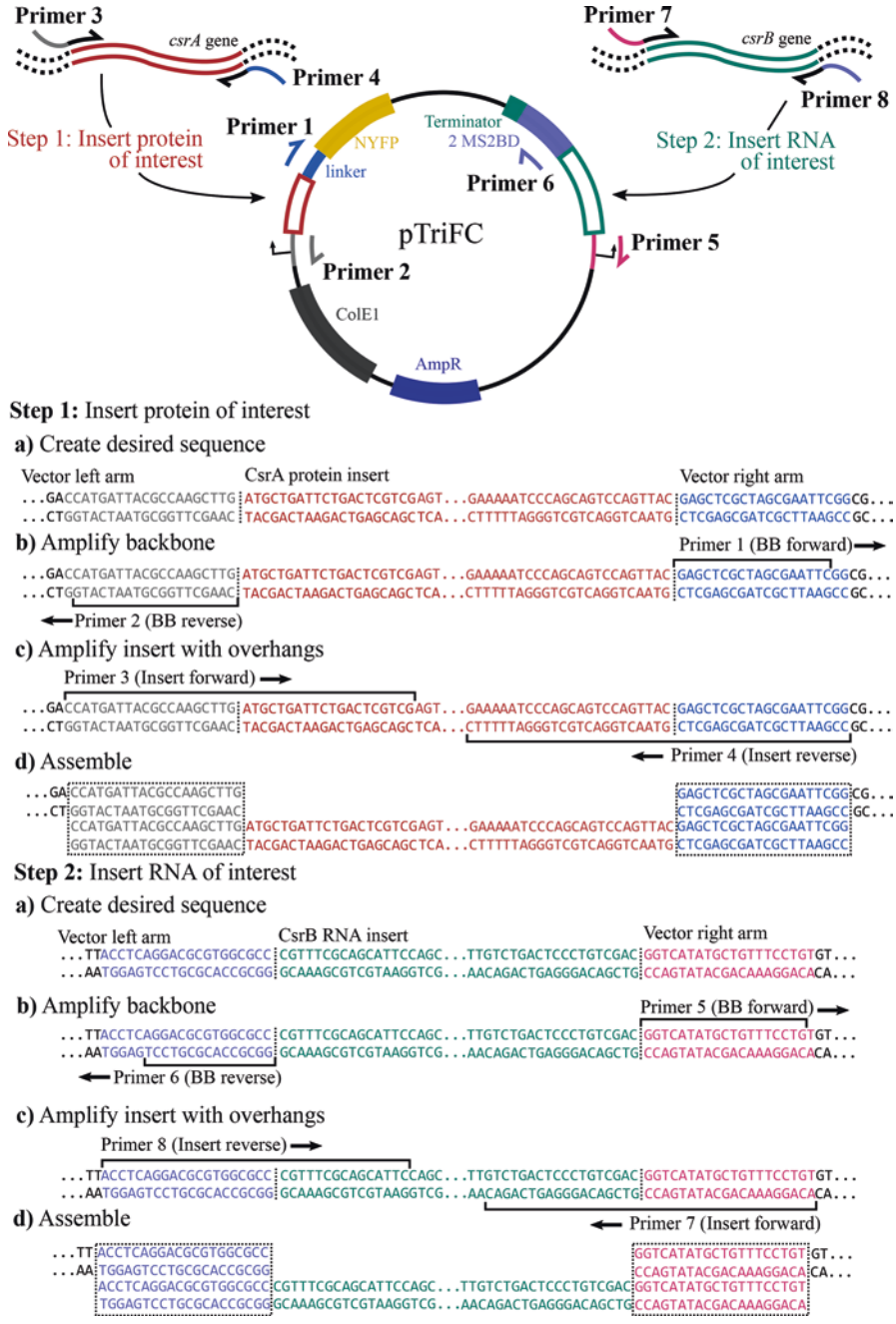


Fig. 7 Example cloning scheme for pTriFC. The TriFC assay can be adapted to investigate an RNA-protein interaction of interest via Gibson cloning; an example cloning scheme for adding the CsrA protein and the CsrB sRNA is illustrated in two steps. (**Step 1a**) The desired composite sequence, left (5′) vector arm + *csrA* gene sequence + right (3′) vector arm, is determined. (**Step 1b**) Primers 1 and 2 are designed (NEBuilder Assembly tool) and used to amplify the vector backbone forward upstream and reverse downstream of the protein insertion site. (**Step 1c**) Primers 3 and 4 are designed to amplify the *csrA* gene with overhangs homologous to the ends of the amplified vector backbone (NEBuilder Assembly tool); they are used to amplify the *csrA* gene (with homologous overhangs) from genomic DNA. (**Step 1d**) Lastly, the amplified vector fragment and the amplified

Table 3
Example primers for Gibson Assembly of RNA and protein of interest with pTriFC ("BB," vector backbone)

Cloning objective	Primer name	Sequence (5′ overhangs in lower case)
pTriFC, protein of interest	Primer 1 (BB forward)	GAGCTCGCTAGCGAATTC
	Primer 2 (BB reverse)	CAAGCTTGGCGTAATCATG
	Primer 3 (insert forward)	ccatgattacgccaagcttgATGCTGATTCTGACTCGTCG
	Primer 4 (insert reverse)	ccgaattcgctagcgagctcGTAACTGGACTGCTGGGATTTTTC
pTriFC, RNA of interest	Primer 5 (BB forward)	GGTCATATGCTGTTTCCTG
	Primer 6 (BB reverse)	GGCGCCACGCGTCCT
	Primer 7 (insert forward)	acaggaaacagcatatgaccGTCGACAGGGAGTCAGAC
	Primer 8 (insert reverse)	acctcaggacgcgtggcgccCGTTTCGCAGCATTC

1. To add the protein of interest to pTriFC, build a composite sequence containing the 5′ arm of the pTriFC vector (sequence just upstream of the protein site), the gene sequence of the protein of interest, and the 3′ arm of the pTriFC vector (sequence just downstream of the protein site) (Fig. 7, **step 1a**).
2. Design Primers 1 and 2 (Table 3) to bind to the pTriFC vector backbone and amplify around the protein site (Fig. 7, **step 1b**).
3. Design Primers 3 and 4 (Table 3) to amplify the gene sequence of the protein of interest from genomic DNA (Fig. 7, **step 1c**). Include overhang sequence homologous to the appropriate arm of the pTriFC vector on the 5′ ends of Primers 3 and 4. Specifically, Primer 3 should contain 20 nucleotides homologous to the 5′ pTriFC vector arm and Primer 4 should contain 20 nucleotides homologous to the 3′ pTriFC vector arm.

Fig. 7 (continued) *csrA* gene (with homologous overhangs) are combined in a standard Gibson Assembly reaction. (**Step 2a**) The desired composite sequence, left (5′) vector arm + *csrB* gene sequence + right (3′) vector arm, is determined. (**Step 2b**) Primers 5 and 6 are designed (NEBuilder Assembly tool) and used to amplify the vector backbone forward upstream and reverse downstream of the RNA insertion site. (**Step 2c**) Primers 7 and 8 are designed to amplify the *csrB* gene with overhangs homologous to the ends of the amplified vector backbone (NEBuilder Assembly tool); they are used to amplify the *csrB* gene (with homologous overhangs) from genomic DNA. (**Step 2d**) Lastly, the amplified vector fragment and the amplified *csrB* gene (with homologous overhangs) are combined in a standard Gibson Assembly reaction

Including homologous overhang regions on just the insert-amplifying primers (Primers 3 and 4), rather than the backbone-amplifying primers (Primer 1 and 2) maximizes amenability of the cloning scheme toward large-scale screening endeavors.

4. To add the RNA of interest to pTriFC, build a composite sequence containing the 5′ arm of the pTriFC vector (sequence just upstream of the RNA site), the gene sequence of the RNA of interest, and the 3′ arm of the pTriFC vector (sequence just downstream of the RNA site) (Fig. 7, **step 2a**). Keep in mind that the RNA fusion is expressed on the reverse strand, relative to expression of the protein-NYFP fusion. **Step 2** illustrates 5′ to 3′ the forward strand of pTriFC, the same orientation as the illustration of **step 1**. Thus, the gene sequence of the RNA of interest is encoded 5′ to 3′ on the reverse strand.
5. Design Primers 5 and 6 (Table 3) to bind to the pTriFC vector backbone and amplify around the RNA site (Fig. 7, **step 2b**).
6. Design Primers 7 and 8 (Table 3) to amplify the gene sequence of the RNA of interest from genomic DNA (Fig. 7, **step 2c**). Include overhang sequence homologous to the appropriate arm of the pTriFC vector on the 5′ end of Primers 7 and 8. Specifically, Primer 7, the forward primer for the RNA, should contain 20 nucleotides homologous to the 3′ pTriFC vector arm and Primer 8, the reverse primer for the RNA, should contain 20 nucleotides homologous to the 5′ pTriFC vector arm.

Fig. 7 depicts 3′ placement of the 2MS2BD sequence relative to the RNA of interest. Note that the RNA terminator sequence annotated in Fig. 7 (and Fig. 6a) is that of CsrB. While this element of the RNA-2MS2BD construct does not have to be altered (*see* **Note 17**), it can be exchanged for the terminator sequence of the RNA of interest by single-fragment Gibson Assembly if desired (*see* **Note 18**). The same primer design steps hold for 5′ placement of the 2MS2BD sequence relative to the RNA of interest; the composite sequence determined in **step 2** A will have different 5′ and 3′ vector arms, thus changing the sequence of the appropriate backbone-amplifying primers (Primers 5 and 6) and the homologous overhang sequences on the insert-amplifying primers (Primers 7 and 8). Lastly, one-fragment Gibson Assembly can be used to build the three control constructs listed above; in this instance, only backbone-amplifying primers that include homologous overhangs to exclude the element to be deleted are needed (for an example, *see* manual for Gibson Assembly Site-Directed Mutagenesis Kit from Synthetic Genomics).

7. PCR amplify the protein and/or RNA of interest from previously prepared genomic DNA with insert primers designed above. Insert amplification reactions should be composed as follows: (i) 10.0 μL of 5× Phusion HF buffer (final concentra-

tion 1×), (ii) 1.0 μL of 10 mM DNTPs (final concentration 200 μM), (iii) 2.5 μL of 10 μM forward primer (Primer 3 or Primer 7, final concentration 0.5 μM), (iv) 2.5 μL of 10 μM reverse primer (Primer 4 or Primer 8, final concentration 0.5 μM), (v) 100 ng of genomic DNA, (vi) 0.5 μL of Phusion polymerase, and (vii) nuclease-free water, up to 50.0 μL to reaction volume. Genomic DNA can be obtained with a genomic DNA purification kit or by boiling a colony (single colony, diluted in 50 μL of nuclease-free water, heated at 96 °C for 5 min); 2 μL of this reaction can be used above as genomic DNA template.

8. Amplify pTriFC with backbone primers designed above. Backbone amplification reactions should be composed as above, except with Primers 1 and 2 or Primers 3 and 5 and approximately 5 ng of purified parent pTriFC DNA rather than 100 ng genomic DNA.
9. Both insert and backbone amplification reactions should be cycled as follows: (i) 98 °C for 30 s, (ii) 98 °C for 10 s, (iii) proper annealing temperature for 30s, with 0.5 °C/s ramp rate, (iv) 72 °C for 30 s/kb, (v) Repeat **steps ii–iv** 34 times, and (vi) 72 °C for 5 min, where proper annealing temperature is 3 °C above the lowest primer initial melting temperature (i.e., initial binding region that excludes 5′ overhangs). We have had success with 4 min and 45 s extension times for backbone and insert (150–300 bp) amplification reactions, respectively.
10. Run approximately 5.0 μL of each reaction on an agarose gel (with EZ Vision or loading dye and ethidium bromide) to check for proper band size and amplification specificity (backbone at ~4.3 Kb, insert at ~100–300 bp, depending on RNA or protein of interest).
11. Add 1.0 μL of DpnI to each reaction and digest at 37 °C for 1–3 h to remove parent plasmid (or digest overnight).
12. PCR clean up according to PCR Purification Kit instructions. Determine concentration of purified linear DNA fragments by UV spectrophotometry (*see* **Note 19**).
13. Compose Gibson Assembly reaction of amplified and purified backbone and insert DNA (illustrated as **step 1d** for protein insert and **step 2d** for RNA insert in Fig. 7) as follows: (i) 15 μL 1.33× (or 10 μL 2×) Gibson master mix, (ii) 50 ng amplified purified backbone DNA, (iii) amplified purified insert at 5–10× molar excess of the backbone, such that the total amount of DNA is less than 0.5 pmol, and (iv) nuclease-free water up to 20 μL. Incubate reaction at 50 °C for 60 min using a thermocycler. For assembly of control constructs, 50 ng of the single fragment (i.e., as backbone) is used without

an insert and more nuclease-free water is added to reach a total volume of 20 μL.

14. Desalt 10 μL of the Gibson Assembly reaction on a 0.025 μM nitrocellulose membrane filter floating on nanopure water for 25–50 min. Recover as much as possible (typically 7–8 μL), and transform into electrocompetent *E. coli* cells by standard protocols. After recovery, plate on LB agar with 100 μg/mL carbenicillin as the selective agent and let grow overnight at 37 °C.
15. Colony PCR, restriction digest, and sequencing can all be used to confirm proper assembly of the insert and backbone fragments. We recommend sequencing for final confirmation and colony PCR for pre-sequencing screening in large-scale cloning endeavors. A sample workflow is detailed in **steps 16–20**.
16. Pick 4–10 colonies from each plated Gibson Assembly reaction and dilute into 50 μL of nuclease-free water. Save each colony by streaking onto a new carbenicillin (100 μg/mL) LB agar plate after dilution (and growing over night at 37 °C). Heat diluted colonies at 96 °C for 5 min.
17. Compose an amplification reaction that is specific to the intended pTriFC using an insert-specific primer and a backbone-specific primer: (i) when adding an RNA to pTriFC with 5′ 2MS2BD placement, use Primer 8 (Table 3) and Primer 18 (Table 4), (ii) when adding an RNA to pTriFC with 3′ 2MS2BD placement, use Primer 7 (Table 3) and Primer 17 (Table 4), (iii) when adding a protein to pTriFC, use Primer 3 (Table 3) and Primer 19 (Table 4). The reaction can be set up

Table 4
Example primers for sequencing pTriFC and pTriFC-mStrawberry

Primer name	Description	Sequence
Primer 17	Binds MS2BD reverse to sequence RNA of interest on pTriFC (with 3′ placement of 2MS2BD to RNA)	CCTTAGGATCCATATATAGGGCCC
Primer 18	Binds MS2BD forward to sequence RNA of interest on pTriFC-mStrawberry (or on pTriFC with 5′ placement of 2MS2BD to RNA)	GGGTTCATTAGATCTGCGCGCG
Primer 19	Binds NYFP reverse to sequence protein of interest on pTriFC and pTriFC-mstrawberry	CCGTTTACGTCGCCGTCCAGCTCGACCAGG
Primer 20	Binds CYFP reverse to sequence MS2 binding protein on pMS2-CYFP	GTTATATCGATTTACAGATCTTCTTCGC

as follows: (i) 2.5 μL 10× ThermoPol buffer, (ii) 0.5 μL 10 mM DNTPs (final concentration 200 μM), (iii) 0.5 μL 100 μM forward primer (final concentration 2 μM), (iv) 0.5 μL 100 μM reverse primer (final concentration 2 μM), (v) 2.0 μL boiled colony, (vi) 0.25 μL Taq polymerase, and (vii) nuclease-free water up to 20.25 μL. Reaction should be cycled as follows: (i) 95 °C for 30s, (ii) 95 °C for 30s, (iii) proper annealing temperature for 60s, (iv) 68 °C for 60s/kb, (v) repeat **steps ii–iv** 29 times, and (vi) 68 °C for 5 min, where proper annealing temperature is 5 °C lower than the lowest primer melting temperature.

18. Run approximately 10.0 μL of each colony PCR reaction on an agarose gel (with EZ Vision or loading dye and ethidium bromide) to check for the band size and specificity. Reactions with only the correct band present are promising colonies to sequence.
19. Retrieve the plate on which colonies picked for colony PCR were saved. Grow each promising colony overnight in liquid culture (5 mL LB in 25 mL test tubes with 100 μg/mL carbenicillin).
20. Miniprep saturated cultures the next day, according to kit directions. Measure the concentration of purified plasmid DNA and prepare sequencing reactions as prescribed by your local sequencing facility. An MS2BD-binding primer (Table 4, Primer 17 or 18) is recommended for sequencing inserted RNAs and an NYFP-binding primer (Table 4, Primer 19) for sequencing inserted proteins.

4.3 Fusion Confirmation

Expression and proper biological function of the RNA and protein of interest in the new RNA-2MS2BD and protein-NFYP fusions should be confirmed. Conventional Western and Northern blotting approaches can be used to confirm protein-NYFP and RNA-2MS2BD expression, respectively. Methods to confirm proper protein and RNA function in their respective fusions vary depending on the molecule. For instance, detecting dimerization in a western blot of a protein-NYFP fusion where the protein is known to dimerize in vivo suggests biological function was retained. Similarly, observation of a phenotype characteristic of overexpressing the RNA of interest when the RNA-2MS2BD fusion is overexpressed suggests biological function was retained.

4.4 TriFC Experiment

1. Select a strain for performing the TriFC experiment. The TriFC assay was developed and tested in a derivative of *E. coli* K-12 MG1655 (*see* **Note 20**); however, we foresee that the assay could be successfully adapted for different bacterial species. If high stoichiometries between the RNA or protein of interest are expected or if high endogenous expression of either

component has been established, the TriFC signal may be diluted by the RNA or protein fusion constructs interacting with endogenous versions of each other, rather than the fused versions. This renders a genomic deletion strain of one or both of the components necessary. While the TriFC assay was developed in a strain without lacIq, we expect that inducible control of expression of the RNA and protein fusion constructs will enhance optimization of assay conditions and impart greater flexibility for probing RNA-protein interactions dynamically (*see* **Note 21** and Subheading 4.4, **step 3**).

2. Sequentially transform the strain of choice with pMS2-CYFP and pTriFC (Fig. 8a) by standard $CaCl_2$ transformation protocols. The pMS2-CYFP plasmid (or version lacking MS2) is transformed first and selected for on kanamycin-containing (50 μg/mL) LB agar plates. The strain harboring pMS2-CYFP is next made competent, transformed with pTriFC, and selected for on kanamycin- and carbenicillin-containing (50 and 100 μg/mL, respectively) LB agar plates.
3. Select triplicate colonies of the *E. coli* strain + pMS2-CYFP + pTriFC and grow in LB cultures (supplemented with 50 μg/mL kanamycin and 100 μg/mL carbenicillin) for 18–48 h prior to measuring fluorescence (Fig. 8b). IPTG induction is not required to express the fusions if the strain is not enhanced with lacIq. However, inducible control of the fusion constructs, though not previously demonstrated, proffers flexibilities and ease for optimizing assay conditions. Two distinct cell culture schemes have been successfully demonstrated for the assay: (i) 40 mL cultures in 250 mL shake flasks, grown at 25 °C with 200 rpm shaking for 24–48 h [14] and (ii) 5 mL cultures in 25 mL test tubes, grown at 37 °C with 200 rpm shaking for 18 h [15]. The first scheme was optimized for a single sRNA-protein interaction; the second scheme was optimized for screening multiple mRNA-protein interactions with the dual-fluorescence variation of the TriFC assay (*see* Subheading 5). The first workflow, i.e., growth at lower temperatures, is a standard approach for optimizing proper protein folding [30]; however, we recommend the second work flow as a starting point for optimizing future interaction-specific cell culture schemes due to experimental convenience of shorter growth times.
4. Fluorescence can be measured by flow cytometry (Fig. 8c) or with a plate reader. For plate reader-based measurements, collect 1 mL of each saturated culture and pellet in a 1.7 mL microtube.
5. Decant the LB supernatant and resuspend the pellet in 1 mL of 1× PBS.

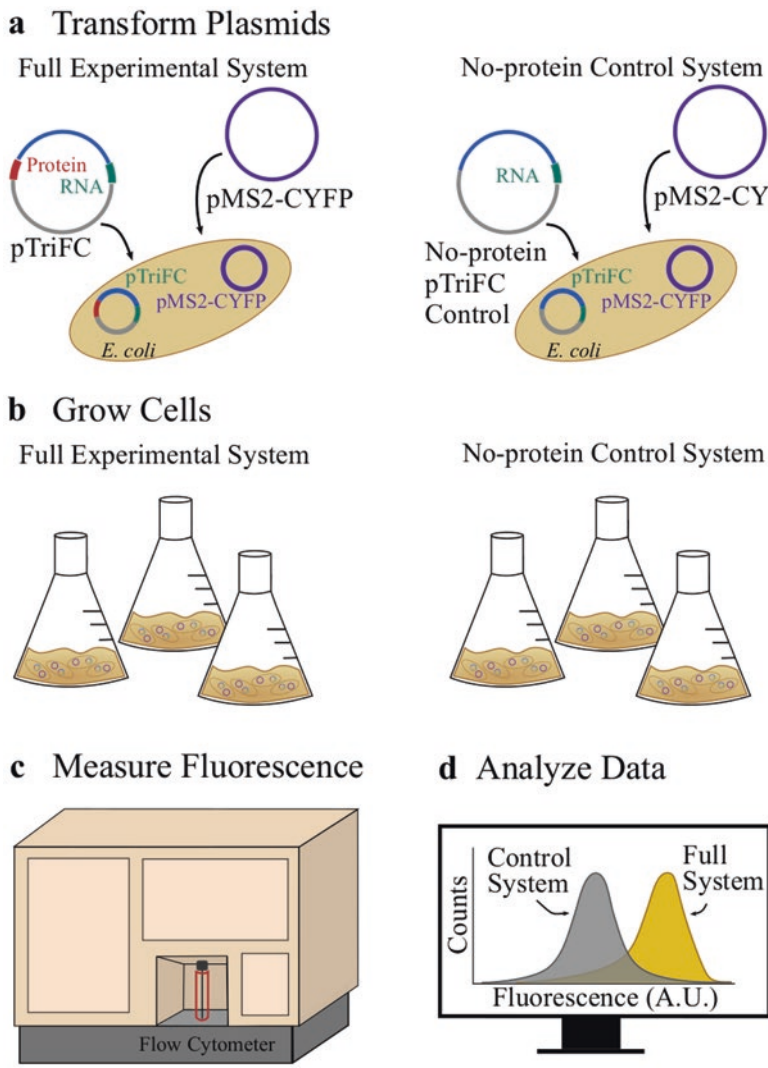


Fig. 8 Work flow for TriFC assay. (**a**) The pTriFC and pMS2-CYFP plasmids are sequentially transformed into an *E. coli* strain of choice to constitute the full experimental system. Similarly, the control systems are transformed into the strain of choice (no-protein, no-RNA, and no-MS2); the no-protein control system is illustrated as an example. (**b**) Cells are grown in triplicate, (**c**) and fluorescence measured by flow cytometry or by plate reader (not shown). (**d**) Finally, fluorescence data are analyzed and the full experimental system is compared to the control system to determine RNA-protein binding

6. Pellet the samples second time and resuspend in a final 1 mL volume of 1× PBS.
7. Transfer 200 uL of each to a 96 well plate (black with clear, flat bottom) for fluorescence and optical density measurements. Be sure to include at least duplicate blank measurements for normalization.
8. Measure yellow fluorescence at the conventional 514/527 nm excitation/emission (ex/em) wavelengths and optical density at 600 nm.

9. Normalize samples' yellow fluorescence measurements by subtracting the average blank reading and dividing by their optical density for further analysis.
10. Alternatively, for flow cytometry, dilute 1–3 μL of the saturated culture into 1 mL of 1× PBS in 5 mL round bottom polystyrene tubes (final concentration ~10^7 cells/mL) and analyze fluorescence by flow cytometry for at least 25,000–50,000 active cells.
11. It is not usually possible to isolate the fluorescent population from non-fluorescent background noise using the forward and side scatter values. Thus, measure fluorescence on both green and yellow channels (530/30 nm and 585/42 nm band pass filter channels, respectively, for the BD FACSCalibur) and track events on a dot plot of the 530 nm channel fluorescence versus the 585 nm channel fluorescence.
12. Distinguish the YFP fluorescent signal from background noise via the 530 nm channel measurement and apply a minimum threshold value to isolate the fluorescent population. The isolated population's 585 nm channel fluorescence measurements are then used for further analysis.

4.5 Data Analysis

Fluorescence measurements of the control constructs, described in Subheading 4.1, are critical for determining whether the recorded fluorescence of a given RNA-protein interaction constitutes true RNA-protein binding (Fig. 8d). We recommend pairing the control and experimental constructs as follows to form three negative control experiments, each in triplicate: (i) pTriFC lacking the protein of interest (NYFP + RNA-2MS2BD present) with pMS2-CYFP, (ii) pTriFC lacking the RNA of interest (protein-NYFP + 2MS2BD present) with pMS2-CYFP, and (iii) pTriFC with pMS2-CFYP lacking the MS2 protein. Additionally, a scramble or known nontarget RNA sequence can be paired with the protein of interest as a negative control [15]. If fluorescence is measured by flow cytometry, geometric means (of the isolated population) of triplicate measurements of the *E. coli* strain + pMS2-CYFP + pTriFC can be compared to those of the control experiments by one-tailed heteroscedastic Student's T-tests. If the full system's average geometric mean fluorescence is significantly greater than that of the control(s), a true RNA-protein interaction is determined to have occurred. Similarly, if fluorescence is measured by plate reader, simply use the normalized yellow fluorescence measurements to compare triplicate samples of the *E. coli* strain + pMS2-CYFP + pTriFC to those of the control experiments by one-tailed heteroscedastic Student's T-tests. If the full system's average normalized yellow fluorescence is significantly greater than that of the control(s), a true RNA-protein interaction is determined to have occurred.

5 Dual-Fluorescence TriFC

The dual-fluorescence variation of the TriFC assay was developed to assess the interaction of mRNAs, particularly anticipated regulatory 5′ UTRs, with a protein. The critical difference between the dual-fluorescence TriFC assay and the original is the addition of a red fluorescent protein, mStrawberry, to the RNA fusion construct (Fig. 6c). Conceivably, the dual-fluorescence TriFC assay could also be applied to any suspected regulatory element of an mRNA, not necessarily just the 5′ UTR, as long as a ribosome binding site and translation start site are included. Tracking red fluorescence along with yellow fluorescence offers (i) confirmation of functional expression of the RNA fusion construct and (ii) potential for tracking the effect of protein-5′ UTR binding on 5′ UTR-controlled red fluorescence. The following details the specifics of the dual-fluorescence TriFC assay where it differs from the original assay.

5.1 Construct Design

In the dual-fluorescence TriFC assay, pTriFC is modified to include an mStrawberry sequence on the 3′ end of the RNA fusion construct and is renamed "pTriFC-mStrawberry" (Fig. 6c). Two repeats of the MS2BD sequence are recommended and, in this instance, required to be placed on the upstream of the RNA of interest as to not interrupt translation of mStrawberry. Due to the usual length of mRNAs (average length of coding sequence of a K-12 *E. coli* gene is 950 base pairs [31]), only a portion of the mRNA sequence can be probed at a time by TriFC. If the protein binding site within the mRNA is known or suspected, this knowledge can guide selection of the RNA sequence to be probed. Otherwise, the 5′ UTR and initial portion of coding sequence of an mRNA are typically of interest. We recommend probing the 5′ UTR sequence, defined by the gene's closest annotated promoter, plus the first 100 nucleotides of coding sequence of the mRNA of interest. RegulonDB can be used to identify annotated promoters for a given *E. coli* gene [32]. In this way, any promoter elements that could affect expression, and thus fluorescence, of the 2MS2BD-RNA-mStrawberry construct are excluded. The SphI restriction site, located between the RNA of interest and mStrawberry in pTriFC-mStrawberry (Fig. 6c), is designed with two spacer nucleotides preceding it so that the 100 nucleotide coding sequence contained in the RNA of interest will be in frame with mStrawberry. For many *E. coli* genes, this definition would yield RNA sequences of interest of about 150 to 350 nucleotides in length. If the mRNA in question does not have a defined promoter sequence or is in the interior of an operon, probing the 100 nucleotides just upstream of the translation start site plus the first 100 nucleotides of coding sequence has been shown reasonable [15].

Recommended control constructs include: (i) pTriFC-mStrawberry lacking the protein of interest (NYFP + 2MS2BD-RNA-mStrawberry present), (ii) pTriFC-mStrawberry with a scramble or known nontarget RNA of interest (protein-NYFP + 2MS2BD-scramble RNA-mStrawberry present), and (iii) pMS2-CFYP lacking the MS2 protein. Using a scramble or known nontarget RNA instead of a no RNA control allows the second control system to retain red fluorescence. The first control construct offers opportunity to compare the effect of RNA-protein binding on red fluorescence controlled by the RNA of interest.

5.2 Cloning Procedures

Cloning procedures are the same as for the original TriFC assay, described in Subheading 4.2. Fig. 9 illustrates an example cloning scheme for adding the CsrA protein and the *glgC* 5′ UTR to pTriFC-mStrawberry. Table 5 documents the sequences of Primers 9–16 shown in Fig. 9. Primers for sequence confirmation of the RNA or protein of interest are given in Table 4.

5.3 Fusion Confirmation

Expression and function of the 2MS2BD-RNA-mStrawberry construct can be confirmed by measuring red fluorescence of an *E. coli* strain harboring pTriFC-mStrawberry and pMS2-CYFP (by plate reader at ex/em 570/590 or 570/600 nm).

5.4 Dual-Fluorescent TriFC Experiment

The experiment is performed according to the protocol of the original TriFC assay (*see* Subheading 4.4 and Fig. 8) with minor modifications. Yellow fluorescence should be measured by plate reader at ex/em 430/510 nm wavelengths and optical density measured at 900 nm to minimize conflict of the YFP-RFP signals. Conceivably, fluorescence can also be measured by flow cytometry, though separation of the YFP and RFP signals by gating has not been attempted.

5.5 Data Analysis

Data analysis proceeds as for the original assay, except that red fluorescence must be present (pass-fail test) to indicate proper construct expression prior to analysis of yellow fluorescence.

Fig. 9 (continued) (**Step 1d**) Lastly, the amplified vector fragment and the amplified *csrA* gene (with homologous overhangs) are combined in a standard Gibson Assembly reaction. (**Step 2a**) The desired composite sequence, left (5′) vector arm + 5′ UTR sequence of the *glgC* gene + right (3′) vector arm, is determined. (**Step 2b**) Primers 13 and 14 are designed (NEBuilder Assembly tool) and used to amplify the vector backbone forward upstream and reverse downstream of the RNA insertion site. (**Step 2c**) Primers 15 and 16 are designed to amplify the *glgC* 5′ UTR sequence with overhangs homologous to the ends of the amplified vector backbone (NEBuilder Assembly tool); they are used to amplify the *glgC* 5′ UTR sequence (with homologous overhangs) from genomic DNA. (**Step 2d**) Lastly, the amplified vector fragment and the amplified *glgC* 5′ UTR (with homologous overhangs) are combined in a standard Gibson Assembly reaction

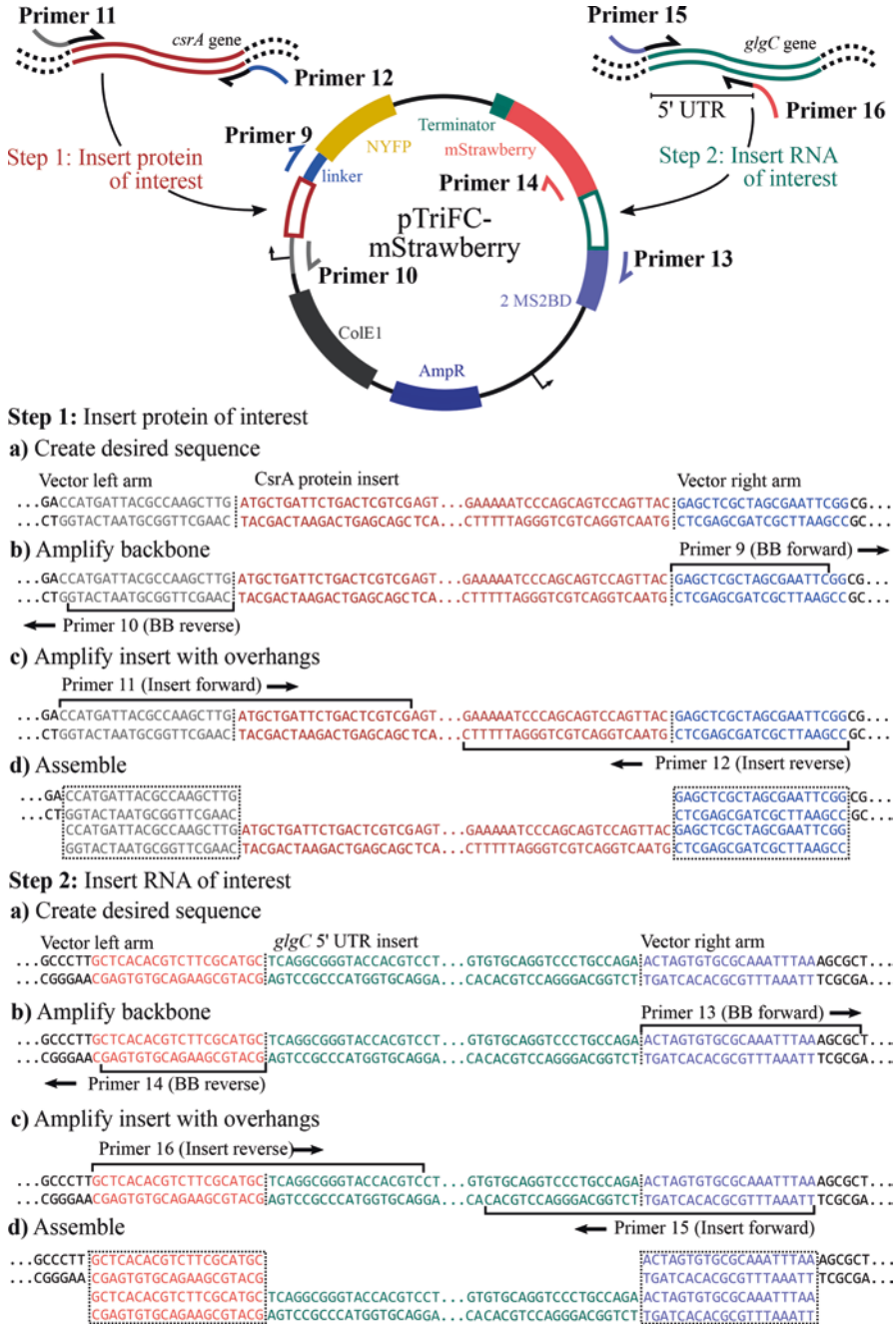


Fig. 9 Example cloning scheme for pTriFC-mStrawberry. The Dual-fluorescence TriFC assay can be adapted to investigate an mRNA-protein interaction of interest via Gibson cloning; an example cloning scheme for adding the CsrA protein and the 5′ UTR of the *glgC* gene is illustrated in two steps. (**Step 1a**) The desired composite sequence, left (5′) vector arm + *csrA* gene sequence + right (3′) vector arm, is determined. (**Step 1b**) Primers 9 and 10 are designed (NEBuilder Assembly tool) and used to amplify the vector backbone forward downstream and reverse upstream of the protein insertion site. (**Step 1c**) Primers 11 and 12 are designed to amplify the *csrA* gene with overhangs homologous to the ends of the amplified vector backbone (NEBuilder Assembly tool); they are used to amplify the *csrA* gene (with homologous overhangs) from genomic DNA.

Table 5
Example primers for Gibson Assembly of RNA and protein of interest with pTriFC-mStrawberry ("BB," vector backbone)

Cloning objective	Primer name	Sequence (5′ overhangs in lower case)
pTriFC-mStrawberry, protein of interest	Primer 9 (BB forward)	GAGCTCGCTAGCGAATTC
	Primer 10 (BB reverse)	CAAGCTTGGCGTAATCATG
	Primer 11 (insert forward)	ccatgattacgccaagcttgATGCTGATTCTGACTCGTCG
	Primer 12 (insert reverse)	ccgaattcgctagcgagctcGTAACTGGACTGCTGGGATTTTTC
pTriFC-mStrawberry, RNA of interest	Primer 13 (BB forward)	ACTAGTGTGCGCAAATTTAAAGCGC
	Primer 14 (BB reverse)	GCATGCGAAGACGTGTGAG
	Primer 15 (insert forward)	ttaaatttgcgcacactagtTCTGGCAGGGACCTGCAC
	Primer 16 (insert reverse)	gctcacacgtcttcgcatgcTCAGGCGGGTACCACGTC

6 Notes

1. For inserts difficult to amplify, ramp temperature at 0.5–2 °C/s during entrance into and exit from annealing stage of cycle.
2. Up to 10 pairs of annealed primers can be used in one reaction if desired. Simply substitute 2 μL of single annealed primer pair for 2 μL of solution made up of equal volumes of the desired annealed primers (each unique annealed primer pair represented at 10 μM).
3. Autoclaving 4-chloro-DL-phenylalanine within LB agar solution promotes dissolution.
4. The most important strain criterion is the detection of a significant fluorescence range between a negative control (scramble asRNA construct) and a positive control (open RBS construct).
5. Although these are the unique methods used to perform iRS[3] assays in our lab, we foresee no issues using other culturing volumes given that (i) the protocol is consistent across strains and (ii) there is ample culture volume at sampling OD.

6. If growing cultures in plates, it is recommended to perform an initial growth curve analysis to gauge time required till desired induction and/or sampling ODs are reached.
7. A higher likelihood of culture contamination is present when growing cultures in plates. If possible, leave blank LB wells that can serve as a qualitative measure of whether or not culturing and splitting protocols are causing well contamination.
8. 2× quantity of flasks or plate wells are required due to splitting of culture at time of induction
9. If using plates, create a map which divides prospective uninduced and induced samples for ease of experimental execution on Day 2.
10. If growing culture in plates, pre-mix arabinose and aTc prior to inducing, as required aTc volumes (per well) will be small.
11. Samples corresponding to the same strain should be run on cytometer in close temporal proximity to avoid data skews due to instrument shift over duration of fluorescence measurements.
12. In general, C-terminal attachment of NYFP to the protein of interest is preferred in light of possible co-translational folding of the protein of interest. If the protein of interest has no previously published tag designs, we recommend first attempting C-terminal attachment of NYFP to the protein of interest. Alternatively, one can use a structure of the protein to predict whether critical N- or C-terminal geometry exists and select the other terminus for linking the protein of interest to NYFP.
13. Flexibility in the protein-NYFP construct is particularly important when working with complex, high-stoichiometry RNA-protein interactions, such as the CsrB sRNA-CsrA protein system, where CsrB binds CsrA with a stoichiometry of ~18 proteins to a single sRNA molecule [33]. A shorter, generic linker derived from a residual multiple cloning site was unsuccessfully attempted for the CsrA-NYFP fusion before the $(GGGGS)_3$ linker was shown to allow expected protein-RNA complementation [14].
14. It should be noted that the optimal linker length for the protein-NYFP fusion is dependent on both the geometry of the protein of interest and the concentration of the fusion expressed from the plasmid. Linker length may need to be altered in conjunction with experimental assay conditions to optimize detection of a RNA-protein interaction of interest.
15. CsrB is a 369 nucleotide-long sRNA that folds into 15 repeating stem-loop structures, most of which are putative binding sites for its target protein, CsrA. Placement of the 2MS2BD

sequence on the 5′ end of CsrB could have disrupted important 5′ to 3′ folding that occurs during CsrB transcription.

16. Importantly, placement of the MS2BD sequence 5′ of the RNA of interest is more amenable to screening applications because the MS2BD sequence can be included as part of the pTriFC backbone, rather than as an insertion before the terminator in each RNA. This arrangement is advantageous for large-scale cloning endeavors that are required for screening applications [15].
17. Given that 3′ placement of the 2MS2BD sequence separates the 3′ portion RNA of interest from its terminator by 200 nucleotides, any contribution of the terminator to RNA function (other than ending transcription) is likely abolished. As such, we do not foresee altering the terminator sequence of the RNA-2MS2BD fusion to match that of a new RNA of interest as necessary.
18. Single-fragment Gibson Assembly can be used to change the terminator sequence of the RNA-2MS2BD fusion in pTriFC from that of CsrB to that of a new RNA of interest. Backbone-amplifying primers are designed to exclude the CsrB terminator and contain overhangs homologous to each other, i.e., the terminator sequence of the new RNA of interest (for an example, *see* manual for Gibson Assembly Site-Directed Mutagenesis Kit from Synthetic Genomics).
19. Alternatively, the full digest could be run in Subheading 4.2, **step 10**, Subheading 4.2, **step 11** skipped, and a full gel extraction of the amplified bands performed in Subheading 4.2, **step 12**. However, we have had most success with DpnI digestion and PCR cleanup rather than gel extraction.
20. Specifically, endogenous expression of CsrB was greater than the pLacO-driven expression of the CsrB fusion construct. Since CsrA binds to CsrB with a stoichiometry of ~18:1 [33], an *E. coli* K-12 MG1655 strain with a genomic deletion of CsrB was used to maximize detectable CsrB-CsrA interactions [14].
21. Expression of pTriFC in a strain without lacIq led to a substantial growth defect. Notably, the defect was only present with the full system; the three control pTriFC constructs that have either the protein, RNA, or MS2 protein missing grew unencumbered. In light of this, we attributed the growth defect to insoluble protein formation due to NYFP-CYFP refolding. We expect that performing the TriFC assay in a strain that contains lacIq and an appropriate genomic deletion will allow for dose-dependent expression of the fusion constructs and maximize the information the TriFC assay can provide.

Acknowledgments

This work is supported by the Welch Foundation (Grant F-1756 to L.M.C.), and the National Science Foundation (Grant MCB 1716777 to L.M.C., and DGE-1610403 to A.N.L. and M.K .M.).

References

1. Vazquez-Anderson J, Contreras LM (2013) Regulatory RNAs: charming gene management styles for synthetic biology applications. RNA Biol 10(12):1778–1797. https://doi.org/10.4161/rna.27102
2. Tsai C-H, Liao R, Chou B, Palumbo M, Contreras LM (2015) Genome-wide analyses in bacteria show small-RNA enrichment for long and conserved intergenic regions. J Bacteriol 197(1):40–50. https://doi.org/10.1128/jb.02359-14
3. Cho SH, Lei R, Henninger TD, Contreras LM (2014) Discovery of ethanol responsive small RNAs in Zymomonas mobilis. Appl Environ Microbiol 80(14):4189–4198. https://doi.org/10.1128/aem.00429-14
4. Jones AJ, Venkataramanan KP, Papoutsakis T (2016) Overexpression of two stress-responsive, small, non-coding RNAs, 6S and tmRNA, imparts butanol tolerance in Clostridium acetobutylicum. FEMS Microbiol Lett 363(8): fnw063. https://doi.org/10.1093/femsle/fnw063
5. Sowa SW, Vazquez-Anderson J, Clark CA, De La Peña R, Dunn K, Fung EK, Khoury MJ, Contreras LM (2015) Exploiting post-transcriptional regulation to probe RNA structures in vivo via fluorescence. Nucleic Acids Res 43(2):e13. https://doi.org/10.1093/nar/gku1191
6. Watters KE, Abbott TR, Lucks JB (2015) Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. Nucleic Acids Res 44(2):e12. https://doi.org/10.1093/nar/gkv879
7. Ignatova Z, Narberhaus F (2017) Systematic probing of the bacterial RNA structurome to reveal new functions. Curr Opin Microbiol 36:14–19. https://doi.org/10.1016/j.mib.2017.01.003
8. Strobel EJ, Watters KE, Loughrey D, Lucks JB (2016) RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. Curr Opin Biotechnol 39:182–191. https://doi.org/10.1016/j.copbio.2016.03.019
9. Garst AD, Edwards AL, Batey RT (2011) Riboswitches: structures and mechanisms. Cold Spring Harb Perspect Biol 3(6):a003533. https://doi.org/10.1101/cshperspect.a003533
10. Takahashi MK, Watters KE, Gasper PM, Abbott TR, Carlson PD, Chen AA, Lucks JB (2016) Using in-cell SHAPE-Seq and simulations to probe structure–function design principles of RNA transcriptional regulators. RNA 22(6):920–933. https://doi.org/10.1261/rna.054916.115
11. Vazquez-Anderson J, Mihailovic MK, Baldridge KC, Reyes KG, Haning K, Cho SH, Amador P, Powell WB, Contreras LM (2017) Optimization of a novel biophysical model using large scale in vivo antisense hybridization data displays improved prediction capabilities of structurally accessible RNA regions. Nucleic Acids Res 45(9):5523–5538. https://doi.org/10.1093/nar/gkx115
12. Lim F, Peabody DS (1994) Mutations that increase the affinity of a translational repressor for RNA. Nucleic Acids Res 22(18): 3748–3752. https://doi.org/10.1093/nar/22.18.3748
13. C-D H, Chinenov Y, Kerppola TK (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. Mol Cell 9(4):789–798. https://doi.org/10.1016/S1097-2765(02)00496-3
14. Gelderman G, Sivakumar A, Lipp S, Contreras LM (2015) Adaptation of tri-molecular fluorescence complementation allows assaying of regulatory Csr RNA-protein interactions in bacteria. Biotechnol Bioeng 112(2):365–375. https://doi.org/10.1002/bit.25351
15. Sowa SW, Gelderman G, Leistra AN, Buvanendiran A, Lipp S, Pitaktong A, Vakulskas CA, Romeo T, Baldea M, Contreras LM (2017) Integrative FourD omics approach profiles the target network of the carbon storage regulatory system. Nucleic Acids Res 45(4):1673–1686. https://doi.org/10.1093/nar/gkx048
16. Gripenland J, Netterling S, Loh E, Tiensuu T, Toledo-Arana A, Johansson J (2010) RNAs: regulators of bacterial virulence. Nat Rev Microbiol 8(12):857–866

17. Storz G, Vogel J, Wassarman Karen M (2011) Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 43(6):880–891. https://doi.org/10.1016/j.molcel.2011.08.022
18. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6(5):343–345. https://doi.org/10.1038/nmeth.1318
19. Green AA, Silver PA, Collins JJ, Yin P (2014) Toehold switches: de-novo-designed regulators of gene expression. Cell 159(4):925–939. https://doi.org/10.1016/j.cell.2014.10.002
20. Bandyra KJ, Said N, Pfeiffer V, Górna MW, Vogel J, Luisi BF (2012) The seed region of a small RNA drives the controlled destruction of the target mRNA by the Endoribonuclease RNase E. Mol Cell 47(6):943–953. https://doi.org/10.1016/j.molcel.2012.07.015
21. Gorski SA, Vogel J, Doudna JA (2017) RNA-based recognition and targeting: sowing the seeds of specificity. Nat Rev Mol Cell Biol 18(4):215–228. https://doi.org/10.1038/nrm.2016.174
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
23. Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, Dirks RM, Pierce NA (2011) NUPACK: analysis and design of nucleic acid systems. J Comput Chem 32(1):170–173. https://doi.org/10.1002/jcc.21596
24. Engler C, Marillonnet S (2014) Golden gate cloning. In: Valla S, Lale R (eds) DNA cloning and assembly methods, vol 1116. Humana Press, New York, pp 119–131
25. Gottesman S, McCullen C, Guillier M, Vanderpool C, Majdalani N, Benhammou J, Thompson K, FitzGerald P, Sowa N, FitzGerald D (2006) Small RNA regulators and the bacterial response to stress. Cold Spring Harb Symp Quant Biol 71:1–11. https://doi.org/10.1101/sqb.2006.71.016
26. Kostecki JS, Li H, Turner RJ, DeLisa MP (2010) Visualizing interactions along the Escherichia coli twin-arginine translocation pathway using protein fragment complementation. PLoS One 5(2):e9225. https://doi.org/10.1371/journal.pone.0009225
27. Liu MY, Romeo T (1997) The global regulator CsrA of Escherichia coli is a specific mRNA-binding protein. J Bacteriol 179(14):4639–4642. https://doi.org/10.1128/jb.179.14.4639-4642.1997
28. Edwards AN, Patterson-Fortin LM, Vakulskas CA, Mercante JW, Potrykus K, Vinella D, Camacho MI, Fields JA, Thompson SA, Georgellis D, Cashel M, Babitzke P, Romeo T (2011) Circuitry linking the Csr and stringent response global regulatory systems. Mol Microbiol 80(6):1561–1580. https://doi.org/10.1111/j.1365-2958.2011.07663.x
29. Klein JS, Jiang S, Galimidi RP, Keeffe JR, Bjorkman PJ (2014) Design and characterization of structured protein linkers with differing flexibilities. Protein Eng Des Sel 27(10):325–330. https://doi.org/10.1093/protein/gzu043
30. Makrides SC (1996) Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol Rev 60(3):512–538
31. Xu L, Chen H, Hu X, Zhang R, Zhang Z, Luo ZW (2006) Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. Mol Biol Evol 23(6):1107–1108. https://doi.org/10.1093/molbev/msk019
32. Gama-Castro S, Salgado H, Santos-Zavaleta A, Ledezma-Tejeida D, Muñiz-Rascado L, García-Sotelo JS, Alquicira-Hernández K, Martínez-Flores I, Pannier L, Castro-Mondragón JA, Medina-Rivera A, Solano-Lira H, Bonavides-Martínez C, Pérez-Rueda E, Alquicira-Hernández S, Porrón-Sotelo L, López-Fuentes A, Hernández-Koutoucheva A, Moral-Chávez VD, Rinaldi F, Collado-Vides J (2016) RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. Nucleic Acids Res 44(D1):D133–D143. https://doi.org/10.1093/nar/gkv1156
33. Liu MY, Gui G, Wei B, Preston JF 3rd, Oakford L, Yuksel U, Giedroc DP, Romeo T (1997) The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli. J Biol Chem 272(28):17502–17510. https://doi.org/10.1074/jbc.272.28.17502

Chapter 10

Mutational Analysis of sRNA–mRNA Base Pairing by Electrophoretic Mobility Shift Assay

Eva Maria Sternkopf Lillebæk and Birgitte Haahr Kallipolitis

Abstract

Small regulatory RNAs (sRNAs) in bacteria often act by base pairing to mRNAs. Direct interactions between an sRNA and its target mRNA can be investigated by electrophoretic mobility shift assay. In this assay, regions engaged in base pairing are analyzed by introducing mutations in one of the RNAs that prevent sRNA–mRNA complex formation, followed by the introduction of complementary mutations in its partner RNA that restore base pairing. Here, we describe the design of a mutational strategy used to analyze the base pairing between two CU-rich regions of the sRNA Rli22 and the AG-rich Shine-Dalgarno region of the mRNA *oppA* in *Listeria monocytogenes*. The protocol can be employed for mutational studies of base pairing between any sRNA and its mRNA target(s).

Key words Electrophoretic mobility shift assay, Regulatory sRNA, Target mRNA, Base pairing

1 Introduction

Small noncoding RNAs (sRNAs) in bacteria are known as important regulators of gene expression [1]. In many cases, sRNA-mediated control occurs at the posttranscriptional level and involves direct base pairing to specific regions of target mRNAs. Typically, the sRNA binds to the Shine-Dalgarno (SD) region of a mRNA, leading to inhibition of translation initiation and mRNA degradation [1]. Electrophoretic mobility shift assay (EMSA) is a simple and fast method to analyze direct interactions between a specific sRNA and its target mRNA(s). EMSA was originally developed to monitor interactions between proteins and nucleic acids [2] but can also be used to study RNA–RNA interactions. In brief, one of the interacting RNAs is labeled with ^{32}P and mixed with unlabeled partner RNA. The samples are analyzed by gel electrophoresis using a native polyacrylamide gel. If the labeled RNA is engaged in a RNA–RNA complex, it will migrate more slowly through the gel, relative to unbound RNA. Using this method, the effect of mutations within base

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_10,

pairing regions can easily be examined. Mutations designed to disrupt base pairing between RNA partners will prevent the formation of a RNA–RNA complex. By introducing complementary mutations that restore base pairing, the RNA partners will regain their ability to engage in complex formation.

We have used EMSA to investigate the base pairing of LhrC sRNAs and their target mRNAs [3–5]. The LhrC regulatory case of *Listeria monocytogenes* LO28 is complicated by the fact that the LhrC family consists of seven homologous sRNAs: LhrC1-5, Rli22, and Rli33-1. Furthermore, each sibling sRNA holds multiple binding sites for target mRNAs. The LhrC1-5 sRNAs are capable of binding up to three mRNAs, whereas Rli22 and Rli33-1 each hold two mRNA-binding sites. By employing the EMSA method, we were able to study the base pairings between individual sibling sRNAs and specific mRNAs, using mutant variants of the partner RNAs. First, a mutant mRNA derivative was designed to disrupt base pairing with its partner sRNA. Then, mutant variants of the sRNA were designed that disrupt base pairing with wild-type mRNA, but restored binding to the mutant mRNA. In this chapter, we will focus on the interactions between the LhrC family member Rli22 and the target mRNA *oppA*. However, the method can be applied to the analysis of base pairing between any pair of RNAs.

2 Materials

Make sure that RNase-free H_2O, demineralized H_2O (dH_2O), 96% ethanol, and 3 M sodium acetate pH 4.5 are available for repeated use throughout the protocol.

2.1 DNA Template

1. Overlapping DNA primer sets of interest (10 μM) (*see* **Note 1**).
2. dNTP mix: 10 mM of dATP, dCTP, dGTP, and dTTP.
3. High-fidelity DNA polymerase and reaction buffer. We used Phusion High-Fidelity DNA Polymerase and 5× Phusion HF buffer (provided by the supplier) (New England Biolabs).
4. 2% agarose gel mix with ethidium bromide (1 μg/mL) with TAE buffer (0.04 M Tris–base, 0.02 M sodium acetate, 1 mM ethylene-diamine-tetra-acetic-acid (EDTA)).
5. DNA purification kit. We used illustra GFX PCR DNA and gel band purification kit (GE Healthcare) (*see* **Note 2**).

2.2 In Vitro Transcription and Purification of RNA

1. Purified DNA templates of interest (*see* **Note 3**).
2. In vitro transcription kit. We used MEGAscript® T7 transcription kit (Thermo Fisher Scientific).
3. RNA fragments of interest (*see* **Note 4**).

4. TBE running buffer (1× TBE): prepare by dilution of 10× TBE buffer with dH_2O. 10× TBE consist of 0.89 M Tris–base, 0.89 M boric acid, and 0.02 M EDTA. We used UltraPure™ TBE buffer (10×) (Thermo Fisher Scientific).
5. 6% denaturing polyacrylamide gel mix (50:1 acrylamide:bisacrylamide): mix 150 mL of acrylamide solution (40%), 60 mL of bisacrylamide solution (2%), 100 mL of 10× TBE buffer, 420 g urea, and dH_2O to 1 l. Filtrate the solution and store at 4 °C.
6. Ammonium persulfate (APS). Prepare a 10% solution by dissolving 1 g APS in dH_2O to a total volume of 10 mL (*see* **Note 5**).
7. *N, N, N′, N′*-tetramethylethylene diamine (TEMED).
8. Glass plates (16 × 19 cm), spacers, and comb (1 mm thick). The comb should optimally have wells large enough to accommodate a volume of 63 μL.
9. Formamide loading dye: formamide, 5 mM EDTA, and 0.01% bromophenol blue (BPB) (*see* **Note 6**).
10. Vita-wrap (*see* **Note 7**).
11. TLC plate Polygram PE SIL G/UV_{254} (Macherey-Nagel).
12. UV lamp. We use a Mineralight® Multiband Ultraviolet lamp 254/366 nm.
13. Surgical blades.
14. 2 M NH_4-acetate pH 5.5. Prepare 100 mL by dissolving 15.42 g of NH_4-acetate in approx. 50 mL of dH_2O. Adjust pH with CH_3COOH and adjust volume to 100 mL. Sterile filtrate and store at 4 °C.
15. RNA phenol pH 4.5. Dissolve phenol in dH_2O (*see* **Note 8**). Mix 100 mL phenol phase, 50 mL aqueous phenol phase, and 0.1 g 8-hydroxyquinoline in a dark glass bottle. Add 3.33 mL 3 M sodium acetate, pH 4.5. Shake to mix and leave to separate overnight (ON). Check the pH of the aqueous phase with pH strips and adjust with sodium acetate if necessary.

2.3 Labeling of RNA

1. Polynucleotide kinase (PNK) (10 U/μL) supplied with 10× PNK buffer (New England Biolabs).
2. Shrimp alkaline phosphatase (SAP) (1 U/μL) (Affymetrix).
3. [γ- ^{32}P]ATP (6000 Ci/mmol) (Perkin Elmer).
4. EDTA (7.5 mM).
5. RNA purification kit. We use NucleoSpin® miRNA kit (Machery-Nagel).

2.4 EMSA

1. Unlabeled and labeled RNA fragments of interest (*see* **Note 9**).
2. EMSA loading buffer: 50% glycerol, sterile filtered.

3. EMSA control loading dye: 50% glycerol with 0.02% BPB.
4. 5× binding buffer: 100 mM HEPES (pH 8.0), 500 mM KCl. Prepare 10 mL with RNase-free H_2O and store at room temperature (RT).
5. *E. coli* MRE 600 tRNA (Sigma Aldrich). Prepare a 10 mg/mL stock by dissolving in RNase-free H_2O. Aliquote in 100 µL portions and store at −80 °C.
6. 5% native polyacrylamide gel (38:1 acrylamide:bisacrylamide): mix 50 mL 10× TBE, 125 mL acrylamide (40% solution), 65 mL bisacrylamide (2% solution), and dH_2O to a total volume of 1 L. Filtrate the solution and store at 4 °C.
7. Glass plates (16 × 19 cm), comb, and spacers (1 mm thick). The comb should accommodate at least 19 samples.
8. Whatman Grade 2 Chr Cellulose Chromatography Paper (Frisenette).

3 Methods

3.1 In Silico Prediction of sRNA–mRNA Interactions and Design of Mutant RNA Derivatives

1. Putative base pairing regions in full length sRNA and target mRNA are predicted using the IntaRNA program [6, 7]. For the sRNA Rli22, two CU-rich regions were predicted to interact with the AG-rich SD region of the *oppA* mRNA (Fig. 1a). The CU-rich regions reside within loop A and a single-stranded stretch of Rli22 (Fig. 1b). Both CU-rich regions contain a conserved UCCC motif.
2. Mutant derivatives of sRNA and target mRNAs are designed by substituting selected nucleotides within the base pairing region. First, design a mutant mRNA derivative that is predicted to disrupt base pairing with the sRNA. Next, design a mutant sRNA derivative that restores base paring with the mutant mRNA, but disrupts interaction with wild-type mRNA. The mutant derivatives should be carefully designed to prevent changes of the secondary structure of the RNA. Potential effects of nucleotide (nt) substitutions on secondary structure are tested by using the Mfold web server [8]. To investigate the importance of the SD region in *oppA* mRNA for interaction with Rli22, we designed a mutant version (*oppA*_mut) carrying 3 nt substitutions (5′GGGA to 5′CGCU) within the base pairing region (Fig. 1a). For Rli22, the importance of the two UCCC motifs for base pairing with *oppA* mRNA was assessed by designing three mutant Rli22 variants carrying 3 nt substitutions (5′UCCC to 5′AGCG) in the CU-rich regions of loop A (Rli22_mut_loopA), the single-stranded stretch (Rli22_mut_ss), or both regions (Rli22_double_mut) (*see* Fig. 1a, b). The substitutions in Rli22 did not alter

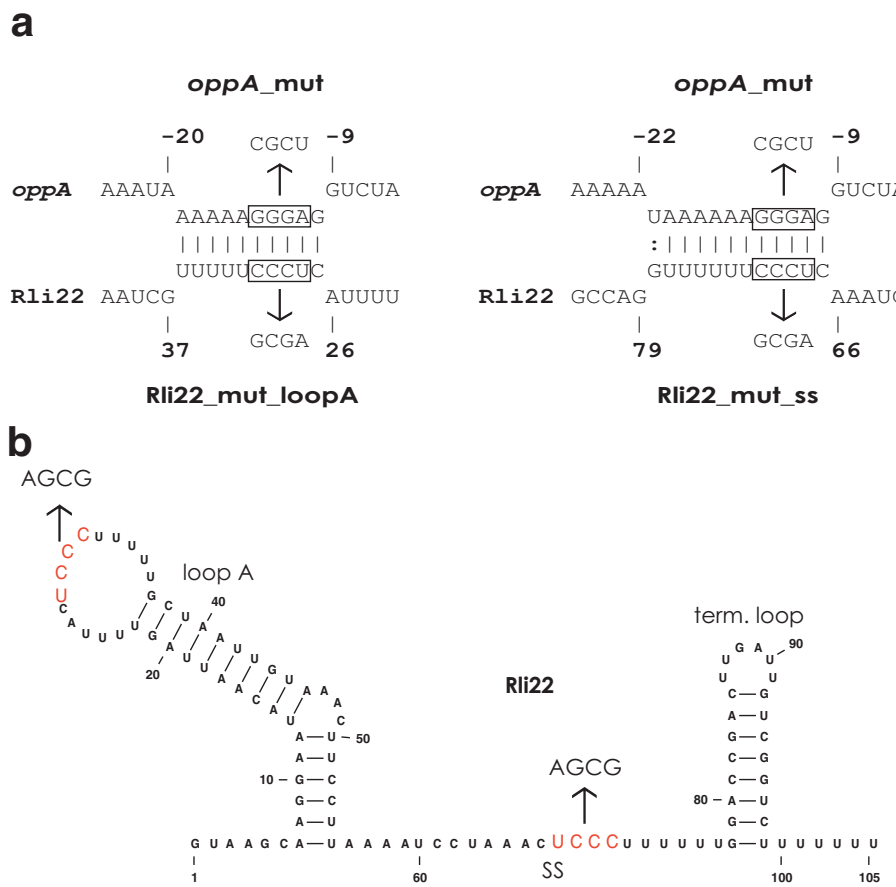


Fig. 1 Mutational analysis of base pairing between Rli22 and *oppA* mRNA. (**a**) The predicted interactions between the CU-rich regions of Rli22 and the SD region of *oppA* mRNA. Substitutions were introduced within the UCCC motifs (boxed) in Rli22 and the complementary GGGA sequence (boxed) in *oppA* mRNA. The sequences of Rli22_mut_loopA, Rli22_mut_ss, and *oppA*_mut are shown. The mutant variant Rli22_double_mut contains the substitutions of both loop A and the single-stranded stretch. (**b**) Secondary structure of Rli22. The UCCC motifs are located within loop A and the single-stranded stretch (ss). The UCCC motifs are depicted in red. The substitutions introduced within the UCCC motifs are shown (reproduced from ref. [5] with permission from Taylor & Francis)

the secondary structure of the sRNA and were predicted to restore base paring with the mutant *oppA* target mRNA (Fig. 1a).

3.2 Design and Preparation of DNA Template

1. DNA templates for in vitro transcription of the sRNA and target mRNA, respectively, are prepared by using overlapping primer sets (forward and reverse primers, respectively). The DNA templates are designed to encode the entire sRNA and 100–200 nt of the target mRNA containing the predicted site of interaction. From the 5′-end, a forward primer contains: (1) four G's; (2) the T7 promoter sequence as described by the manufacturer for use of the MEGAscript T7 transcription kit;

and (3) DNA sequence encoding the 5′-part of the desired RNA. The reverse primer contains sequence information corresponding to the 3′-part of the desired RNA. Importantly, the forward and reverse primers are designed to contain overlapping sequences of 20–25 nt at their 3′-ends. For preparation of DNA templates encoding mutant derivatives of sRNA Rli22 and target mRNA *oppA*, the primer sets were designed to contain the substitutions described in **step 2** of Subheading 3.1.

2. For each template, prepare a 50 μL DNA template synthesis mix: 1 μL dNTP mix, 5 μL forward primer, 5 μL reverse primer, 10 μL 5× Phusion buffer, 21.5 μL dH_2O, and 0.5 μL of Phusion polymerase (2 U/μL).
3. Run the DNA template synthesis reaction in a thermocycler. To extend the overlapping primers, use the following program: 98 °C for 1 min, then 10 cycles of 98 °C for 15 s, T_M + 3 °C for 15 s, and 72 °C for 15 s (*see* **Note 10**).
4. Check the quality and length of the DNA product by loading 2 μL on a 2% agarose gel.
5. Purify DNA from the residual 48 μL DNA template synthesis reaction using GFX kit. Quantify DNA using NanoDrop or equivalent and adjust the concentration of DNA template to 20 ng/μL.

3.3 In Vitro Transcription and Purification of RNAs

1. For in vitro transcription of the DNA template, use MEGAscript T7 kit. 30 μL reactions are prepared as follows: 3 μL ATP solution, 3 μL CTP solution, 3 μL GTP solution, 3 μL UTP solution, 3 μL 10× reaction buffer, 12 μL of template DNA (20 ng/μL), 3 μL of enzyme mix. Incubate the in vitro transcription reaction ON at 37 °C in an incubator (*see* **Note 11**).
2. After ON incubation, add 1.5 μL of Turbo DNase and digest for 15 min at 37 °C (*see* **Note 12**).
3. Prepare 35 mL of 6% denaturing gel mix with 350 μL APS (10%) and 35 μL TEMED in a Falcon tube and cast the gel.
4. Mix the in vitro transcribed RNA sample with equal amounts of formamide loading dye, then heat for 2 min at 75 °C. Load the sample into a single well (*see* **Note 13**). Separate RNA and template DNA by running the gel at 300 V for approx. 45 min.
5. Separate the glass plates and carefully transfer the gel to vita-wrap. Cover the gel by a single layer of foil. Place the gel on a fluor-coated TLC plate. Visualize the RNA transcript by UV shadowing and mark the band with a marker.
6. Slice out the gel piece and transfer to a 2 mL safe-lock tube. Add 500 μL of 2 M ammonium acetate pH 5.5 and allow

diffusion for 2 h at 17 °C, slowly shaking. Add 500 µL of RNA phenol and mix gently. Leave ON at 17 °C, slowly shaking.

7. Briefly centrifuge the tube containing the gel piece and transfer the liquid to a clean 2 mL safe-lock tube. Centrifuge for 5 min at 13,000 × *g* and transfer the aqueous phase to a clean tube containing 500 µL RNA phenol. Mix by shaking.
8. Centrifuge for 5 min at 13,000 × *g* and transfer the aqueous phase to a clean tube containing 500 µL chloroform. Vortex for 5 s and repeat the centrifugation step.
9. Transfer the aqueous phase to a clean tube and add 0.1 volumes of 3 M sodium acetate pH 4.5 and 3.5 volumes of 96% ethanol to precipitate RNA. Leave at −20 °C for at least 1 h.
10. Centrifuge the samples at 4 °C for 30 min at 21,000 × *g*. Remove the supernatant without disturbing the RNA pellet and add 1 mL of ice-cold 70% ethanol. Re-centrifuge for 15 min. Carefully remove the supernatant. Spin briefly in tabletop centrifuge and remove the remaining supernatant. Allow pellet to air dry at RT for approx. 10 min.
11. Resuspend the RNA pellet in 50 µL of RNase-free H_2O. Quantify and check the integrity of the transcript by NanoDrop and gel electrophoresis (*see* **steps 4** and **5** in Subheading 3.2), respectively. Adjust to a final concentration of 8 pmol/µL RNA and store samples at −80 °C (*see* **Note 14**).

3.4 5′-End Labeling of RNAs

1. Dephosphorylate RNA fragments using SAP. For a 10 µL reaction, mix 1 µL of 10× PNK buffer, 1 µL of SAP, 1 µL of purified RNA (8 pmol/µL), and 7 µL RNase-free H_2O.
2. Incubate at 37 °C for 2 h and heat-inactivate at 65 °C for 10 min. Place on ice and proceed immediately to the next step.
3. Mix 3 µL of dephosphorylated RNA (2.4 pmol) with 0.7 µL 10× PNK buffer, 2 µL ^{32}P γ-ATP, and 3.3 µL RNase-free H_2O to a final volume of 10 µL. Add 1 µL PNK enzyme.
4. Incubate at 37 °C for 30 min, add 20 µL of 7.5 mM EDTA and incubate at 75 °C for 15 min. Place on ice for 2 min. Proceed immediately to the next step (*see* **Note 15**).
5. Purify the labeled RNA transcripts by using Nucleospin miRNA cleanup kit. Mix labeled RNA sample with 200 µL MX binding buffer, load on column (*see* **Note 16**), incubate 1 min at RT, centrifuge for 1 min at 13,000 × *g*. Discard flow through by transferring the column to a clean tube.
6. Add 500 µL MW1 wash buffer, incubate for 1 min at RT. Repeat centrifugation and discard flow through (*see* **Note 17**).
7. Add 500 µL MW2 wash buffer, incubate for 1 min at RT. Re-centrifuge and discard flow through.

8. Elute RNA by adding 30 µL RNase-free H_2O to the column; wait for 1 min; centrifuge for 1 min at 13,000 × *g*.
9. Repeat elution step with another 30 µL RNase-free H_2O. Final RNA concentration: 0.04 pmol/µL (*see* **Note 18**).
10. Store 5′-end labeled RNA transcripts at −20 °C until use.

3.5 EMSA

Set up the EMSA experiment as exemplified for Rli22 and *oppA* RNA, *see* Table 1. The binding between wild-type Rli22 and wild-type or mutant *oppA* RNA is tested by preparing 9 samples containing 5′-end labeled Rli22_wt and increasing amounts of unlabeled *oppA*_wt or *oppA*_mut RNA. The result of this EMSA experiment is presented in Fig. 2a.

1. Mix the samples as described in Table 1. Start by adding the water, binding buffer and tRNA to the tubes (*see* **Note 19**).

Table 1
Overview of samples prepared for a single EMSA experiment, testing the binding between 5′-end labeled Rli22_wt and unlabeled *oppA*_wt or *oppA*_mut

Sample no.	1	2	3	4	5	6	7	8	9
Labeled Rli22_wt (0.04 pmol/µL) Final conc: 4 nM	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
tRNA (10 mg/mL)	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
Unlabeled *oppA*_wt (0.02 pmol/µL) Final conc: 4 nM	–	2 µL	–	–	–	–	–	–	–
Unlabeled *oppA*_wt (0.1 pmol/µL) Final conc: 20 nM	–	–	2 µL	–	–	–	–	–	–
Unlabeled *oppA*_wt (0.5 pmol/µL) Final conc: 100 nM	–	–	–	2 µL	–	–	–	–	–
Unlabeled *oppA*_wt (2.5 pmol/µL) Final conc: 500 nM	–	–	–	–	2 µL	–	–	–	–
Unlabeled *oppA*_mut (0.02 pmol/µL) Final conc: 4 nM	–	–	–	–	–	2 µL	–	–	–
Unlabeled *oppA*_mut (0.1 pmol/µL) Final conc: 20 nM	–	–	–	–	–	–	2 µL	–	–
Unlabeled *oppA*_mut (0.5 pmol/µL) Final conc: 100 nM	–	–	–	–	–	–	–	2 µL	–
Unlabeled *oppA*_mut (2.5 pmol/µL) Final conc: 500 nM	–	–	–	–	–	–	–	–	2 µL
5× binding buffer	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL
RNase-free H_2O	6 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL	4 µL

The result of the binding experiment is shown in Fig. 2a

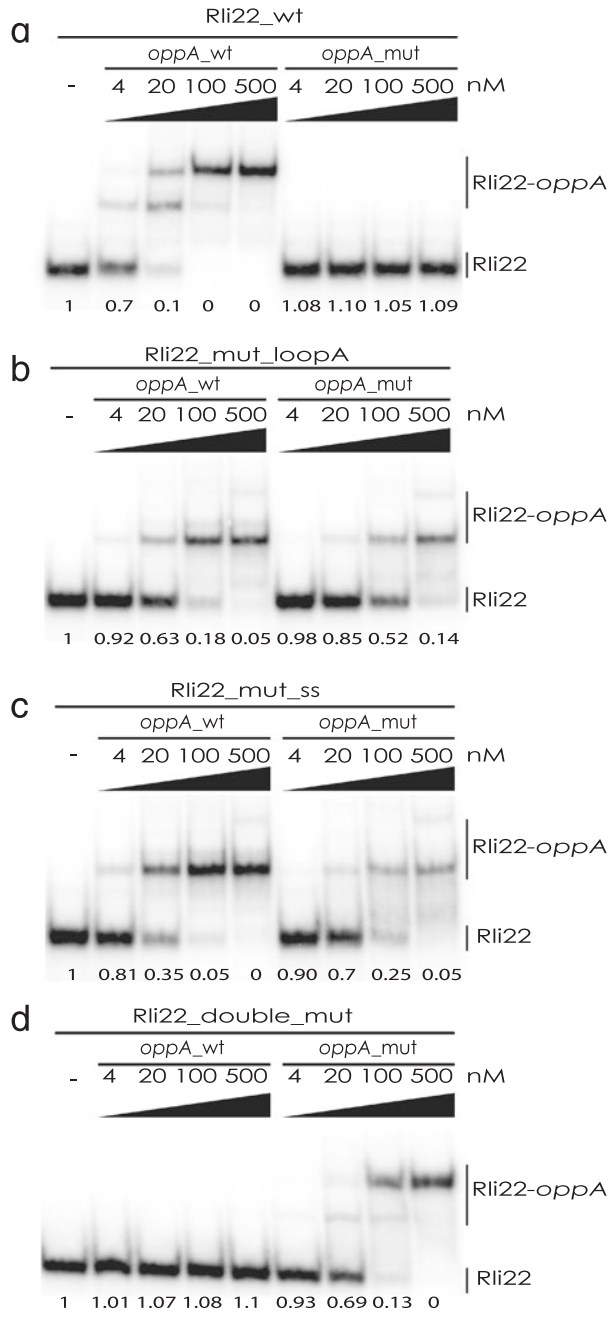


Fig. 2 Base pairing between Rli22 and *oppA* mRNA investigated by EMSA. 5′-end labeled Rli22_wt (**a**), Rli22_mut_loopA (**b**), Rli22_mut_ss (**c**), or Rli22_double_mut (**d**) was incubated with increasing concentrations of unlabeled *oppA*_wt or *oppA*_mut. The fraction of unbound sRNA is shown below the lanes. The Rli22_wt bound *oppA*_wt very effectively leading to two shifted bands (**a**); mutation of *oppA* fully disrupted complex formation with Rli22_wt, indicating that a single Rli22 molecule interacts with two *oppA* mRNA molecules. For Rli22_mut_loopA (**b**) and Rli22_mut_ss (**c**), addition of *oppA*_wt resulted in the formation of single complexes, whereas no binding was observed for Rli22_double_mut and *oppA*_wt (**d**). When incubated with *oppA*_mut, complex formation with Rli22_double_mut was fully restored (**d**) (reproduced from ref. [5] with permission from Taylor & Francis)

2. Incubate the samples for 1 h at 37 °C and place on ice for 10 min. In the meanwhile, prepare the EMSA gel (next step).
3. Cast a 5% native acrylamide gel; solidify with 1% APS and 0.1% TEMED. Pre-run gel in cold room (4 °C) at 110 V for 30 min. Use ½× TBE for running buffer.
4. Add 5 μL of 50% glycerol to the tubes immediately before loading, mix by pipetting once and load 5 μL while the gel is running. Load a control sample consisting of 50% glycerol and BPB to get an estimate of the migration of the samples through the gel.
5. Run the gel for 1–1½ h in cold room (4 °C).
6. Separate glass plates and transfer gel to Whatman grade 2 filter paper. Cover with vita-wrap. Vacuum dry in gel dryer for approx. 1 h. Radioactive bands are visualized and quantified by autoradiography (*see* **Note 20**).
7. For binding studies of the mutant versions of Rli22, simply repeat **steps 1–6** using labeled Rli22_mut_loopA (Fig. 2b), Rli22_mut_ss (Fig. 2c), or Rli22_double_mut (Fig. 2d).

4 Notes

1. Information on how to design overlapping DNA primers is provided in Subheading 3.2.
2. Any other purification kit can be used, as long as there is an option to elute in RNase-free H_2O.
3. Information on how to generate purified DNA templates is provided in Subheading 3.2.
4. Information on how to generate the RNA fragments is provided in Subheading 3.3.
5. This solution can be stored for 1 week at RT, but we recommend preparing a fresh batch each time.
6. Mix and aliquot into 1 mL portions. Store at −20 °C.
7. Based on our experience, using plastic wrap of a different brand or thickness may disrupt the UV shadowing. Vita-wrap is made of 8 μm thick polyethylene.
8. Dissolving phenol may take a long time. Leave with magnetic stirrer for 24 h. Allow to separate into organic and aqueous phase for 24 h. If not dissolved, repeat. Wear proper protective gloves and use flow bench to avoid physical contact with phenol!
9. Information on how to prepare unlabeled and labeled RNA is provided in Subheadings 3.3 and 3.4, respectively.
10. T_M is the melting temperature calculated for the 20–25 nt overlap of the forward and reverse primers.

11. Based on our experience, using an incubator instead of a heating block will minimize evaporation.
12. Turbo DNase is included in the MegaScript kit.
13. The sample can be loaded into two or more smaller wells, but we find that using a single well, large enough to accommodate the entire sample, minimizes the loss of transcript in downstream handling.
14. The RNA is now ready for use as "unlabeled RNA" in EMSA experiments. Otherwise, the RNA may be 5′-end labeled as described in Subheading 3.4.
15. Alternatively, store samples at −20 °C ON.
16. Use the blue or green columns provided with the Nucleospin miRNA cleanup kit.
17. Handle with care. The flow through contains non-incorporated ^{32}P γ-ATP.
18. The recovery yield of the Nucleospin miRNA cleanup kit is >95%.
19. When handling many samples, it is recommended to prepare a pre-mix of the 5× binding buffer, tRNA, and H_2O.
20. We use a Typhoon FLA 9500 (GE Healthcare) to visualize the bands and IQTL 8.0 quantification software (GE Healthcare) to quantify the bands.

Acknowledgments

This work was supported by The Danish Council for Independent Research | Natural Sciences (grant number 12-124735), VILLUM FONDEN, The Lundbeck Foundation, and Novo Nordisk Foundation. We thank Maria Storm Mollerup for discussions and comments on the manuscript.

References

1. Waters LS, Storz G (2009) Regulatory RNAs in bacteria. Cell 136(4):615–628. https://doi.org/10.1016/j.cell.2009.01.043
2. Pennings S (1997) Nucleoprotein gel electrophoresis for the analysis of nucleosomes and their positioning and mobility on DNA. Methods 12(1):20–27. https://doi.org/10.1006/meth.1997.0443
3. Sievers S, Sternkopf Lillebaek EM, Jacobsen K, Lund A, Mollerup MS, Nielsen PK, Kallipolitis BH (2014) A multicopy sRNA of *Listeria monocytogenes* regulates expression of the virulence adhesin LapB. Nucleic Acids Res 42(14):9383–9398. https://doi.org/10.1093/nar/gku630
4. Sievers S, Lund A, Menendez-Gil P, Nielsen A, Storm Mollerup M, Lambert Nielsen S, Buch Larsson P, Borch-Jensen J, Johansson J, Kallipolitis BH (2015) The multicopy sRNA LhrC controls expression of the oligopeptide-binding protein OppA in *Listeria monocytogenes*. RNA Biol 12(9):985–997. https://doi.org/10.1080/15476286.2015.1071011
5. Mollerup MS, Ross JA, Helfer AC, Meistrup K, Romby P, Kallipolitis BH (2016) Two novel members of the LhrC family of small RNAs in Listeria monocytogenes with overlapping regulatory functions but distinctive expression profiles. RNA Biol 13(9):895–915. https://doi.org/10.1080/15476286.2016.1208332

6. Busch A, Richter AS, Backofen R (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics 24(24):2849–2856. https://doi.org/10.1093/bioinformatics/btn544
7. Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, Kleinkauf R, Hess WR, Backofen R (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. Nucleic Acids Res 42(Web Server issue):W119–W123. https://doi.org/10.1093/nar/gku359
8. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31(13):3406–3415

Chapter 11

An Integrated Cell-Free Assay to Study Translation Regulation by Small Bacterial Noncoding RNAs

Erich Michel, Olivier Duss, and Frédéric H.-T. Allain

Abstract

Posttranscriptional regulation of gene expression by small noncoding RNAs (sRNAs) is an important control mechanism that modulates bacterial metabolism, motility, and pathogenesis. Using the bacterial carbon storage regulator/regulator of secondary metabolism (Csr/Rsm) system, we here describe an *E. coli*-based cell-free translation assay that allows a quantitative analysis of translation regulation by ncRNAs and their corresponding translation repressor proteins. The assay quantifies the translation of chloramphenicol acetyltransferase in cell-free expression reactions that contain defined amounts of ncRNA and repressor protein. We demonstrate our protocol with a comparative translation activation analysis of the RsmX, RsmY, and RsmZ sRNAs from *Pseudomonas protegens*, which reveals a superior efficacy of RsmZ over RsmX and RsmY.

Key words Posttranscriptional regulation, Noncoding RNAs, Translation regulation, Cell-free expression, Carbon storage regulator, Regulator of secondary metabolism, Chloramphenicol acetyl transferase, Bacterial virulence, CsrA, RsmA

1 Introduction

Small noncoding RNAs (sRNAs) are important posttranscriptional regulators of gene expression in both prokaryotes and eukaryotes [1]. The bacterial carbon storage regulator/regulator of secondary metabolism (Csr/Rsm) is an important global regulatory system that affects bacterial metabolism and virulence in ~75% of all bacterial species [2, 3]. Translation of target genes in this regulatory system is repressed by the CsrA/RsmA proteins, which occupy the ribosome-binding site of their target messenger RNAs and thereby prevent ribosome entry and translation initiation (Fig. 1) [1, 3, 4]. Changes in environmental conditions can trigger the transcription of sRNAs which contain GGA motifs that sequester the repressor proteins from the mRNA and thus derepress translation (Fig. 1). These sRNAs range from 100 to 500 nucleotides in size and possess various numbers of GGA motifs

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_11,

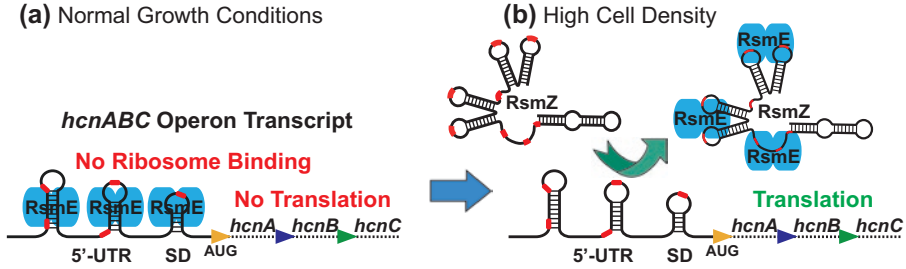


Fig. 1 Schematic overview of posttranscriptional Csr/Rsm-regulated gene expression by the translation repressor protein RsmE and the translation-activating small noncoding RNA (sRNA) RsmZ from *Pseudomonas protegens*. Under normal growth conditions, RsmE binds to GGA motifs (red segments) within the 5′-untranslated region (5′-UTR) of the *hcnABC* operon transcript, which prevents the ribosome to bind to the Shine-Dalgarno (SD) sequence and thereby represses translation initiation. Certain environmental conditions, e.g., a high cell density, trigger the transcription of the sRNA RsmZ, which contains several GGA motifs that sequester RsmE from the 5′-UTR to liberate the SD for ribosome binding and translation of the *hcnABC* transcript

[3, 4], which have affinities for the CsrA/RsmA protein varying by five orders of magnitude [5] and different relative arrangements on the sRNAs to cooperatively bind the CsrA/RsmA protein homodimers [6, 7]. This diversity results in various efficacies of the sRNAs to de-repress translation and provides a means to fine-tune translation activation.

Here we describe an in vitro approach to quantify the translation activation potential of sRNAs of interest. Our assay uses an *E. coli*-based cell-free system for coupled transcription and translation of chloramphenicol acetyltransferase (CAT) in the presence of defined amounts of sRNA and repressor protein. The precise control of the utilized amounts of sRNA and repressor protein provides a substantial advantage over in vivo-based translation assays, where the amounts of sRNA and repressor protein vary significantly over time. We demonstrate our in vitro assay with the *Pseudomonas protegens* (ex-*fluorescens*) regulator of secondary metabolism (Rsm) system, which comprises the RsmE repressor protein and the RsmX, RsmY, and RsmZ sRNAs (Fig. 1). The cell-free expression vector pCFX100 (Fig. 2) combines the 5′-UTR of the *hcnA* mRNA of *P. protegens*, which contains RsmE-regulated GGA motifs within the ribosome-binding site (RBS), with the CAT coding sequence and thereby couples CsrA/RsmE-regulated translation to the quantifiable expression of the reporter enzyme CAT. Analysis of alternative sRNA-regulated gene expression systems can be achieved by simply exchanging the 5′-UTR of the *hcnA* mRNA in pCFX100 with the 5′-UTR of the gene of interest and by usage of the corresponding sRNAs and repressor proteins. An important modification of the cell-free expression setup is the use of a CsrA-depleted *E. coli* cell extract (Fig. 3) because endogenous CsrA in untreated *E. coli* cell extracts significantly

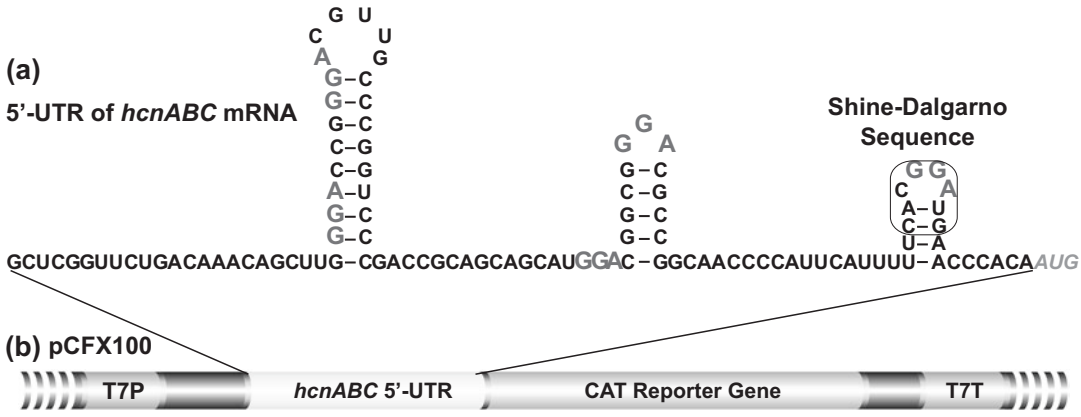


Fig. 2 The 5′-UTR of the *hcnABC* mRNA (**a**) provides RsmE/RsmZ-mediated translation regulation of the quantifiable reporter enzyme chloramphenicol acetyltransferase (CAT) from the pCFX100 vector (**b**). The GGA motifs recognized by RsmE within the 5′-UTR are indicated with gray bold letters and the AUG translation start codon is indicated with gray italic letters. T7P, T7 promoter; T7T, T7 terminator

represses translation and thereby falsifies the interpretation of the assay (Fig. 4).

The translation assay is initiated by mixing all reagents for coupled transcription/translation of CAT with defined amounts of repressor protein and sRNA (Fig. 5). The translation activation potential of sRNAs is readily analyzed by setting up parallel cell-free reactions containing fixed amounts of repressor protein and increasing amounts of the sRNA of interest. After 3 h at 30 °C, the reaction mixture is cleared by centrifugation and the supernatant is mixed with acetyl-CoA, chloramphenicol, and dithionitrobenzoic acid (DTNB). The expressed CAT enzyme catalyzes transfer of the acetyl group from acetyl-CoA to chloramphenicol and thereby liberates CoA, which subsequently reacts via its sulfhydryl group with DTNB to generate the yellow thionitrobenzoate (TNB) chromophore with its characteristic absorbance at 412 nm (Fig. 5). The sRNA-induced translation activation is directly derived from the CAT activity in the reaction supernatant, which is measured by the increase in absorbance at 412 nm over time (Fig. 6). The small reaction volume enables the facile analysis of multiple reactions in parallel. In summary, the here-described protocol allows efficient analysis and quantification of sRNA-mediated translation activation under defined conditions of repressor protein and activating RNA. We demonstrate our protocol with a previously undescribed comparative analysis of the translation activation of the RsmX, RsmY, and RsmZ sRNAs from *P. protegens*, which reveals the superior translation activator potential of RsmZ over RsmX and RsmY (Fig. 7).

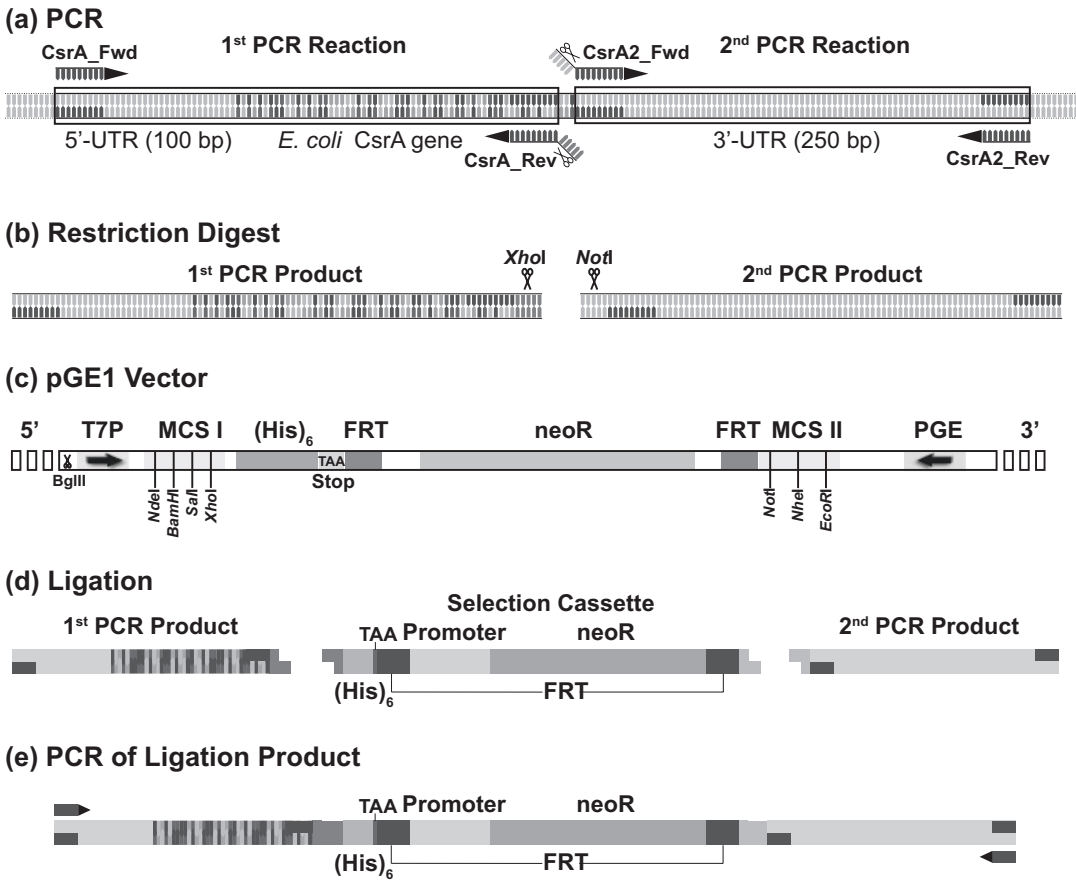


Fig. 3 Preparation of DNA for the genetic engineering of an *E. coli* BL21 (DE3) Star *csrA::(His)*$_6$ strain. (**a**) The entire *csrA* gene including ~ 100 base pairs (bp) of its 5′-UTR and ~ 250 bp of the 3′-UTR are separately PCR-amplified from *E. coli* genomic DNA. (**b**) The two obtained PCR products are digested with the *Xho*I and *Not*I restriction enzymes, respectively, (**c**) which are also used in parallel to excise a DNA cassette from the pGE1 vector that encodes a C-terminal hexahistidine tag (His)$_6$ with a stop codon (TAA) and the kanamycine resistance gene *neoR* flanked by FLP recognition target (FRT) sites. (**d**) The two PCR products and the selection cassette are subsequently ligated, (**e**) the ligation product is PCR-amplified and purified for transformation into *E. coli* BL21 (DE) Star cells harboring the Red recombinase-encoding pKD46 plasmid. MCS, multiple cloning site; PGE, sequencing primer annealing site

2 Materials

All solutions are prepared using reagents of highest available purity and with ultrapure de-ionized water. Unless stated otherwise, the solutions for preparation of *E. coli* cell extract and for setting up the cell-free translation assay are prepared using diethylpyrocarbonate (DEPC)-treated ultrapure water. Particular care should always be taken to avoid RNase contamination during preparation of reagents that are required for the cell-free translation reaction (*see* **Note 1**). All reagents required for cell-free expression are stored at −20 °C, unless indicated otherwise (*see* **Note 2**).

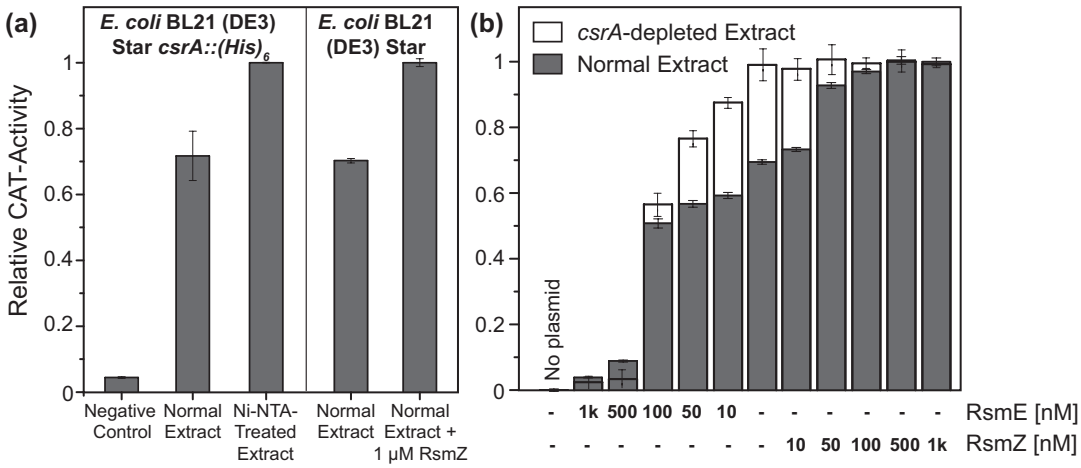


Fig. 4 Comparison of cell-free translation reactions with pCFX100 conducted using either standard or CsrA-depleted *E. coli* cell extract. (**a**) Untreated cell extracts prepared from *E. coli* BL21 (DE3) Star and *E. coli* BL21 (DE3) Star *csrA::(His)*$_6$ strains result in comparable CAT activities. However, Ni-NTA treatment of the cell extract from *E. coli* BL21 (DE3) Star *csrA::(His)*$_6$ significantly increases CAT activity due to removal of the endogenous repressor protein CsrA. The addition of large amounts of RsmZ sRNA to standard cell extracts sequesters CsrA from the 5′-UTR and results in CAT activities that compare to the CsrA-depleted extract. (**b**) Titration of RsmE or RsmZ in the translation assay reveals a biphasic transition with the CsrA-containing standard cell extract (filled bars) and a monophasic transition with the CsrA-depleted extract (empty bars). The second phase of the biphasic transition is due to translation activation by RsmZ which sequesters endogenous CsrA from the 5′-UTR of the mRNA while the CsrA-depleted cell extract shows a monophasic transition because the maximum CAT activity is already obtained without the addition of RsmZ

2.1 Preparation of DNA for Generation of the E. coli BL21 (DE3) Star csrA::(His)$_6$ Strain

1. Oligonucleotide primers: Prepare separate 10 µM solutions of each of the following oligonucleotide primers: CsrA_Fwd (5′-GGAATTC CAT ATG TAA TGT GTT TGT CAT TGC TTA C), CsrA_Rev (5′-CCG CTC GAG GTA ACT GGA CTG CTG GGA TTT TTC), CsrA2_Fwd (5′-AAG ATAAGAAT GCGGCCGC T TTC CGC GTC TCA TCT TTA TCG), and CsrA2_Rev (5′-CCG GAA TTC TGA GAC TTA TAA GTC AAA C).
2. Genomic DNA of the *E. coli* BL21 (DE3) Star strain prepared following standard molecular biology protocols [8] (*see* **Note 3**).
3. Thermal cycler and a high-fidelity DNA polymerase kit.
4. pKD46 [9] and pGE1 vectors (*see* **Notes 4, 5**).
5. *Xho*I and *Not*I restriction enzymes with the supplied digestion buffer.
6. Agarose gel electrophoresis unit including agarose gel casting trays.
7. Tris–acetate-EDTA buffer (TAE; 50×): Dissolve 242 g Tris base, 57.1 mL glacial acetic acid, and 100 mL of 0.5 M EDTA

Cell-Free Translation Reaction

Prepare small-scale (25–50 µL) cell-free reactions by mixing the following reagents:

- *E. coli* S30 extract (CsrA-depleted)
- Amino acids, NTPs, tRNAs, buffer and salts
- pCFX100 plasmid with *hcnABC*-CAT fusion
- T7 RNA Polymerase
- Creatine phosphate and creatine kinase
- Defined amounts of RsmE and RsmZ

(1) Incubate reaction mixtures for 3.5 h at 30 °C
(2) Cool reaction mixtures for 10 min on ice
(3) Spin down mixtures for 10 min at 10000 × *g*

Chloramphenicol Acetyltransferase (CAT) Assay

(1) Dilute 2.5–10 µL of the reaction supernatant to 500 µL with dilution buffer DB [100 mM Tris–HCl at pH 7.8, 1 mg/mL BSA]

(2) Mix 10 µL of the diluted sample with 990 µL reaction buffer RB [100 mM Tris–HCl at pH 7.8, 1 mg/mL BSA, 0.5 mM DTNB, 50 µM acetyl-CoA, 50 µM chloramphenicol]

(3) Follow absorbance at 412 nm for 3–20 min

Acetyl-CoA + Chloramphenicol $\xrightarrow{\text{CAT}}$ CoA + Chloramphenicol-3-Acetate

CoA + Dithionitrobenzoic acid (DTNB) $\longrightarrow$ Thionitrobenzoate (TNB) + CoA derivative

Fig. 5 Workflow for the cell-free translation activation reaction and the CAT assay

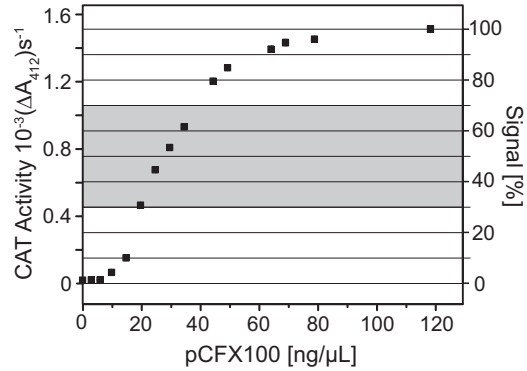


Fig. 6 Analysis of optimal pCFX100 plasmid concentrations for the translation activation assay. Cell-free expression reactions are initiated in parallel using increasing concentrations of the pCFX100 plasmid. The obtained CAT activity in the reaction supernatant is determined by measuring the increase in absorbance at 412 nm (A_{412}) per time. The optimal concentration of plasmid DNA should yield an activity that corresponds to 30–70% of the maximally attainable signal, i.e., between 20 and 40 ng/µL pCFX100 for the depicted case

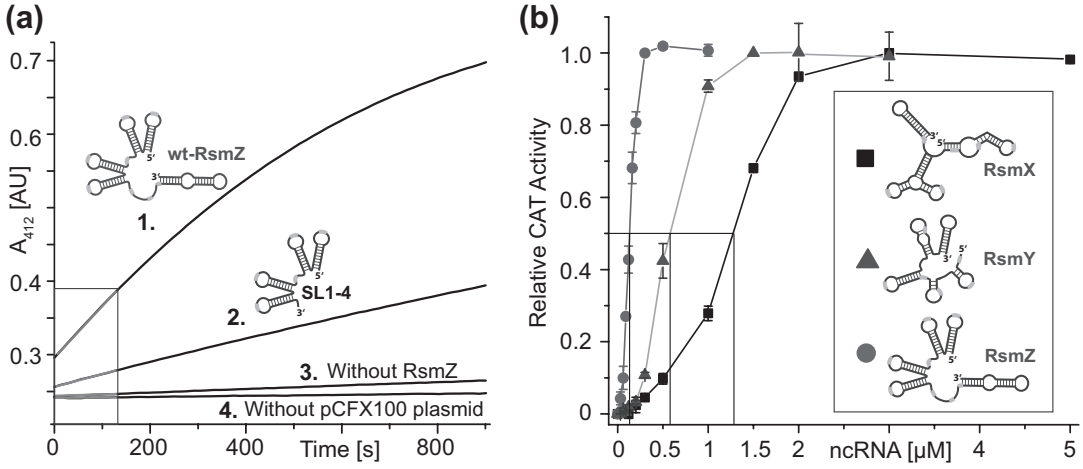


Fig. 7 Examples of translation activation assays. (**a**) Cell-free translation reactions of CAT are conducted in the presence of 100 nM RsmE dimer and either 150 nM RsmZ (curve 1), 150 nM of an RsmZ variant comprising only stem-loops 1–4 (curve 2, SL1–4), no RsmZ (curve 3), or no pCFX100 template plasmid (curve 4). The CAT activity in the reaction supernatant is then determined in the CAT assay by recording the absorbance at 412 nm against time. The slope of the initial rates (boxed gray section of curves) is then determined for each reaction. The curve corresponding to 150 nM RsmZ shows an unfavorable high CAT activity which causes significant acetyl-CoA depletion with a subsequent rate reduction that leads to a nonlinear curve progression. A good curve progression is represented by the curve corresponding to the SL1–4 reaction. (**b**) Individual concentration series of the sRNAs RsmX, RsmY, and RsmZ were conducted in the presence of 250 nM RsmE dimer. The maximum CAT activity in the reactions of RsmX, RsmY, and RsmZ is obtained in the presence of ~2.5 μM, 1 μM, and 0.3 μM of the sRNA, respectively. Alternatively, 50% translation activation is obtained with ~1.3 μM, 0.6 μM, and 0.14 μM of the RsmX, RsmY, and RsmZ sRNAs, respectively. This demonstrates the superior translation activation potential of RsmZ over RsmX and RsmY and reveals the relatively weak translation activation of RsmX

(pH 8.0) in 600 mL de-ionized water and adjust volume to 1 L with de-ionized water.

8. 1% (w/v) TAE-agarose gel: Add 1 g of low electroendosmosis (EEO) agarose to 100 mL 1 × TAE buffer in a 250 mL Erlenmeyer bottle and completely dissolve the agarose by boiling in a microwave (*see* **Notes 6, 7**). Let the solution cool down for 3–5 min, immediately add 5 μL ethidium bromide (10 mg/mL), and pour the solution into a gel casting tray with the comb in place (*see* **Note 8**). Wait until the agarose gel is completely solidified (usually within 1–2 h) before starting the gel electrophoresis.
9. Loading buffer (6×) and molecular weight marker (e.g., 0.1–10 kbp) for DNA agarose gel electrophoresis.
10. UV lamp and a clean scalpel.
11. DNA extraction kit.

12. T4 DNA Ligase kit.
13. 50% (v/v) glycerol: Mix 250 mL glycerol with 250 mL de-ionized water and sterilize by autoclaving.

2.2 CsrA-Depleted *E. coli* Cell Extract

1. Shaking incubator, sterile 1 L, and 5 L Erlenmeyer flasks.
2. *E. coli* BL21 (DE3) Star *csrA::(His)$_6$* strain [7].
3. 0.2 μm sterile filter, laboratory vacuum pump, and a sterile 1 L glass bottle.
4. PYG medium: Dissolve 5.6 g KH_2PO_4, 28.9 g K_2HPO_4, and 10 g yeast extract in 1 L of de-ionized water and sterilize by autoclaving. Prepare a solution of 40% (w/v) glucose in de-ionized water and sterilize by filtration through a 0.2 μm sterile filter. Add 25 mL of the sterile-filtered glucose solution per liter of autoclaved medium to obtain the final PYG medium containing 1% (w/v) glucose.
5. Spectrophotometer to measure the absorbance at 600 nm.
6. Centrifuge with temperature control that allows cooling to 4 °C.
7. SS34 and 1 L centrifugation bottles and corresponding rotors (*see* **Note 9**).
8. DEPC-treated ultrapure water (DEPC-H_2O): Use the fume hood to carefully add 4 mL of DEPC to a 4 L bottle of ultrapure water (*see* **Notes 10**, **11**). Stir overnight with a magnetic bar under the fume hood with the lid loosely covering the bottle. Autoclave the DEPC-treated water and store at 4 °C.
9. S30 buffer: 10 mM Tris–acetate (pH 8.2), 60 mM potassium acetate, 14 mM magnesium acetate, 7.15 mM 2-mercaptoethanol, 1 mM DTT (*see* **Note 12**).
10. French pressure cell press (*see* **Note 13**).
11. 50 mL conical centrifuge tube.
12. Water bath with temperature control.
13. 10 mL of a 50% (w/v) Ni-NTA agarose resin suspension.
14. Gravity purification column (50 mL capacity) with frit.
15. 20 mM amino acid stock solution (AA): Mix 0.14 g Gln, 0.15 g Asn, 0.17 g Arg, 0.2 g Trp, 0.18 g Lys-HCl, 0.21 g His-HCl, 0.16 g Phe, 0.13 g Ile, 0.15 g Glu, 0.12 g Pro, 0.13 g Asp, 0.08 g Gly, 0.12 g Val, 0.11 g Ser, 0.09 g Ala, 0.12 g Thr, 0.12 g Cys, 0.15 g Met, 0.13 g Leu, 0.18 g Tyr, and 0.08 g DTT into 40 mL DEPC-H_2O while stirring with a magnetic bar, adjust the pH to 7.0 with 5 N KOH and add DEPC-H_2O to a final volume of 50 mL (*see* **Note 14**).
16. Preincubation buffer: 293.3 mM Tris–acetate (pH 8.2), 84 mM phosphoenolpyruvate–KOH (pH 7.0), 13.17 mM ATP–KOH (pH 7.0), 9.24 mM magnesium acetate, 4.4 mM

DTT, 40 μM of each proteinogenic amino acid, 6.67 U/mL of pyruvate kinase (*see* **Notes 14**, **15**).

17. 12–14 kDa molecular weight cutoff (MWCO) dialysis membrane with appropriate dialysis tubing clips.
18. 2 L glass beaker.
19. Magnetic stirrer with magnetic bar.
20. 15 mL centrifugal filter device with 10 kDa MWCO.
21. 1.5 mL snap-lock tubes.

2.3 Cell-Free Translation Activation Assay

1. Cell-free expression vector pCFX100 (1 mg/mL), prepared using a plasmid DNA maxiprep kit (*see* **Notes 16**, **17**). Dissolve the purified plasmid DNA pellet in DEPC-H_2O.
2. Tabletop incubator with temperature control and shaking.
3. CsrA-depleted *E. coli* cell extract.
4. RsmX, RsmY, and RsmZ. The sRNAs are prepared by in vitro runoff transcription of linearized plasmid DNA with T7 RNA polymerase as previously described [10] (*see* **Notes 18**, **19).**
5. RsmE repressor protein prepared as previously described [11] (*see* **Note 20).**
6. Energy-regeneration, cofactor, nucleotide (ERCN) solution [12]: 155 μL of 1 M HEPES–KOH (pH 7.5), 271 μL of 40% (w/v) PEG-8000, 89 μL of 6 M potassium acetate, 8.7 μL of 0.55 M DTT, 8.1 μL of 0.4 M ATP–KOH (pH 7.0), 5.8 μL of 0.4 M GTP–KOH (pH 7.0), 5.8 μL of 0.4 M CTP–KOH (pH 7.0), 5.8 μL of 0.4 M UTP–KOH (pH 7.0), 34.5 μL of 5.27 mM folinic acid, 17.3 μL of 100 mM cAMP, 33 μL of 2.2 M ammonium acetate, 216 μL of 1 M creatine phosphate–KOH (pH 7.0), 150 μL DEPC-H_2O to obtain a final volume of 1 mL (*see* **Note 21**).
7. 65 μM T7 RNA polymerase, prepared as described [13] or purchased from commercial providers.
8. *E. coli* tRNA solution (tRNA): Dissolve 17.5 mg (18 U/mg) of *E. coli* tRNA in 1 mL of DEPC-H_2O.
9. Creatine kinase: Dissolve 37.5 mg of creatine kinase (Roche) in 10 mL DEPC-H_2O and store 500 μL aliquots at −20 °C.
10. 0.77 M sodium azide (*see* **Note 22**).
11. 1.6 M magnesium acetate.
12. 1.5 mL microcentrifuge tubes.

2.4 Quantification of CAT Activity

1. Tabletop microcentrifuge.
2. 1.5 mL microcentrifuge tubes.
3. Dilution buffer (DB): 100 mM Tris–HCl (pH 7.8), 1 mg/mL BSA (*see* **Note 23**).

4. CAT reaction buffer (RB): 100 mM Tris–HCl (pH 7.8), 0.5 mM DTNB, 50 μM acetyl-CoA, 50 μM chloramphenicol, 1 mg/mL BSA (*see* **Notes 23, 24**).
5. UV spectrophotometer with cuvettes to measure the absorbance at 412 nm (*see* **Note 25**).

3 Methods

3.1 Preparation of DNA for Generation of the E. coli BL21 (DE3) Star csrA::(His)$_6$ Strain

The preparation of an *E. coli* cell extract devoid of the endogenous CsrA repressor protein is essential to carry out quantitative translation activation analysis. To get a "handle" on CsrA for its subsequent removal from the cell extract, we genomically introduce DNA between the CsrA-coding region and the 3′-UTR (Fig. 3). The introduced DNA sequence encodes a C-terminal $(His)_6$-tag, which allows to subsequently remove the translated endogenous CsrA protein by Ni-NTA affinity chromatography.

1. PCR-amplify the *csrA* gene with its 5′-UTR using the CsrA_Fwd and CsrA_Rev oligonucleotide primers and amplify the 3′-UTR of the *csrA* gene using the CsrA2_Fwd and CsrA2_Rev oligonucleotide primers, respectively (Fig. 3). Prepare 50 μL PCR reactions according to the protocol of the DNA polymerase kit used in your lab and use 100 ng *E. coli* BL21 (DE3) Star genomic DNA as a template and 300 nM of each oligonucleotide primer.
2. Digest 5 μg of the pGE1 vector with the *Xho*I and *Not*I restriction enzymes according to the supplier's protocol using a total reaction volume of 50 μL.
3. Add 10 μL of the 6× DNA loading buffer to each of the PCR reactions and the pGE1 restriction enzyme digest. Completely apply each mixture on individual lanes of the 1% TAE-agarose gel and perform electrophoresis at 80–100 V and using 1 × TAE as a running buffer. Stop electrophoresis when the loading buffer dye is approaching the end of the gel (typically ca. 1–1.5 h).
4. Visualize the separated DNA on the agarose gel using the UV lamp and excise the bands corresponding to the desired PCR products and the pGE1 insert with a clean scalpel (*see* **Notes 26**, **27**).
5. Extract the DNA from the excised pieces of agarose gel using the DNA extraction kit.
6. Digest the purified 5′-UTR-csrA PCR product with the *Xho*I and the purified 3′-UTR PCR product with the *Not*I restriction enzyme, respectively.

7. Inactivate the restriction enzymes after digestion by incubating the reaction mixture for 20 min at 65 °C.
8. Set up a 20 μL DNA ligation reaction according to the supplier's protocol containing 100 ng of the pGE1 insert, 20 ng of the *Xho*I-digested 5′-UTR-csrA PCR product, and 20 ng of the *Not*I-digested 3′-UTR PCR product.
9. Amplify the ligation product in a 50 μL PCR reaction using 5 μL of the ligation reaction as template DNA and 300 nM of each CsrA_Fwd and CsrA2_Rev oligonucleotide primer (*see* **Note 28**).
10. Separate and purify the PCR product as described in **steps 3** to **5** (*see* **Notes 29**, **30**).
11. Use the purified DNA for transformation and genetic engineering of electro-competent *E. coli* BL21 (DE3) Star cells harboring the pKD46 plasmid according to the procedure described by Datsenko and Wanner (*see* **Note 31**).
12. Prepare 1 mL aliquots of the resulting *E. coli* BL21 (DE3) Star *csrA::(His)$_6$* strain by thoroughly mixing 0.5 mL of a log phase cell culture with 0.5 mL of 50% (v/v) glycerol. Freeze the cell suspension in snap-lock tubes in liquid nitrogen and store at −80 °C until needed.

3.2 Preparation of CsrA-Depleted *E. coli* Cell Extract

All steps after cell harvest are performed on ice, unless indicated otherwise. Precool rotors and centrifuges timely before the intended use. Rinse all bottles and equipment that gets in contact with the cell extract sequentially with RNase-AWAY and DEPC-H_2O (*see* **Note 32**).

1. Inoculate 200 mL of sterile PYG medium with a 1 mL aliquot of *E. coli* BL21 (DE3) Star *csrA::(His)$_6$* cell suspension and grow overnight at 37 °C in a 1 L Erlenmeyer bottle with shaking (*see* **Note 33**).
2. Transfer 20 mL of the overnight pre-culture to each of the three sterile 5 L Erlenmeyer flasks containing 1.8 L pre-warmed PYG medium and grow the cell culture at 37 °C with shaking to an OD_{600} of 0.8–1.0 (*see* **Note 34**). Prepare 4 L of S30 buffer while the cells are growing and chill the buffer in an ice-water bath (*see* **Note 12**).
3. Harvest the cells by 10 min centrifugation at 5000 × *g* and 4 °C in 1 L centrifugation bottles (*see* **Note 9**).
4. Re-suspend each cell pellet on ice with 25 mL pre-chilled S30 buffer and transfer the suspension to a pre-weighted 1 L centrifugation bottle (*see* **Note 35**). Centrifuge cells for 10 min at 5000 × *g* and 4 °C and re-suspend the cell pellet in 250 mL pre-chilled S30 buffer. Repeat this centrifugation/re-suspension step two more times.

5. Weigh the cell pellet and re-suspend the cells in 1.8 volumes S30 buffer on ice. Sequentially wash a pre-cooled French pressure cell press with 70% (v/v) ethanol, RNase-AWAY spray, DEPC-H_2O, and S30 buffer. Lyse cells by a single passage through the cleaned and pre-cooled French pressure cell press at 4 °C.
6. Centrifuge the disrupted cells for 30 min at 30,000 × *g* and 4 °C in a pre-chilled SS-34 rotor. In the meantime you should prepare the preincubation buffer corresponding to ¼ of the volume of the cell lysate (*see* **Note 15**).
7. Carefully remove the supernatant by pipetting without disturbing the pellet. Transfer supernatant into a 50 mL tube, gently mix with the preincubation buffer, and incubate for 40 min in a water bath set to 30 °C.
8. Apply 10 mL of a 50% (w/v) Ni-NTA agarose resin suspension into a gravity purification column and equilibrate the column with 50 mL ice-cold S30 buffer at 4 °C. Chill the preincubated cell extract for 10 min in an ice-water bath and centrifuge for 10 min at 4000 × *g* and 4 °C (*see* **Note 36**). Carefully pass the supernatant twice over the Ni-NTA column by gravity flow at 4 °C to remove the $(His)_6$-tagged CsrA from the cell extract (*see* **Note 37**).
9. Dialyze the CsrA-depleted cell extract for 2 h in a 12–14 kDa MWCO dialysis membrane against 1.5 L pre-cooled S30 buffer with stirring at 4 °C, then exchange with 1.5 L fresh S30 buffer and continue dialysis overnight at 4 °C.
10. Transfer the dialyzed cell extract into a 50 mL tube, centrifuge for 10 min at 5000 × *g* and 4 °C and collect the supernatant. Adjust the absorbance at 260 nm either by dilution with pre-chilled S30 buffer or by concentration in a 10 kDa centrifugal filter device at 4000 × *g* and 4 °C to an A_{260} of ca. 250 (*see* **Note 38**).
11. Centrifuge the cell extract in a 50 mL tube for 10 min at 5000 × *g* and 4 °C. Transfer 0.5 mL aliquots of the supernatant into 1.5 mL snap-lock tubes on ice. Immerse the aliquots for ~ 2 min into liquid nitrogen and store the tubes at −80 °C until needed (*see* **Note 39**).

3.3 Performing the Cell-Free Translation Activation Assay

The analysis of cell-free translation activation requires an initial optimization of the pCFX100 plasmid concentration (Fig. 6; *see* **Note 40**). Good sensitivity in the assay is obtained with a plasmid concentration that yields ca. 30–70% of the maximally attainable CAT expression level (Fig.6; *see* **Note 41**). The concentration of repressor protein in the translation reaction should be adjusted to

a concentration that yields ca. 50% translation repression (Fig.4b; *see* **Notes 42**, **43**).

1. Thaw the following components on ice: ERCN solution, pCFX100, CsrA-depleted *E. coli* cell extract, sRNAs, RsmE protein, 20 mM amino acid mixture (AA), T7RNAP, tRNA, CK, NaN_3, and $Mg(OAc)_2$.
2. For 1 mL of cell-free reaction solution, prepare a master mix of the following reagents on ice: 373 µL ERCN solution, 10 µL tRNA, 10 µL NaN_3, 4.3 µL $Mg(OAc)_2$, 75 µL AA (*see* **Note 14**), 66.7 µL CK, 10 µL T7RNAP, and 300 µL CsrA-depleted *E. coli* cell extract. Add the pCFX100 vector and the RsmE repressor protein according to the initially determined optimal concentration (*see* **Notes 40**, **43**, **44**).
3. Always prepare a 50 µL negative control reaction devoid of the pCFX100 plasmid.
4. For each envisaged translation activation reaction, transfer 1/20 of the master mix into separate 1.5 mL tubes on ice. Add the desired amounts of sRNA, including a reaction with no sRNA, to each reaction and adjust the final volume to 50 µL with DEPC-H_2O (*see* **Note 45**).
5. Carefully mix the reaction tubes by pipetting and incubate them for 3.5 h with gentle agitation at 30 °C (*see* **Note 46**).
6. Place the reaction tubes for 10 min on ice. Centrifuge for 10 min at 12,000 × *g* and transfer the supernatant to a clean tube.

3.4 Quantification of CAT Activity

1. Thoroughly mix 5 µL of the reaction supernatant with 495 µL DB on ice.
2. Thoroughly mix 10 µL of the diluted reaction supernatant with 990 µL RB preincubated to the measurement temperature (*see* **Note 25**). Work speedy when preparing several reactions for parallel measurements in a multi-cell spectrophotometer (*see* **Note 47**).
3. Immediately start to record A_{412} against time and proceed with recording until you have accumulated enough data points for a linear fit of the initial rate (Fig. 7a; *see* **Note 47**).
4. Determine the slope of A_{412} against time for each reaction.
5. Subtract the value of the negative control reaction from the value of each reaction (*see* **Note 48**).
6. Normalize each independent sRNA concentration series by division with the largest measured slope value that was obtained after full translation activation (*see* **Notes 45**, **49**).
7. Plot the normalized values against the corresponding concentrations of sRNA (Fig. 7b).

4 Notes

1. The introduction of undesired RNase activity into the cell-free translation mixture can significantly reduce translation efficiencies and thereby distort the interpretation of the resulting data. This RNase sensitivity stems from the strict dependence of translation efficiency on intact tRNAs, mRNAs, and ribosomes, which are all potential targets of common RNases. Also, RNases can degrade the sRNAs used to activate translation initiation, which results in a higher apparent activation threshold for the sRNAs of interest.
2. Storage at −20 °C significantly prolongs the lifetimes of hydrolysis-sensitive reagents such as NTPs. We usually do not observe any significant deterioration in the activity of stored reagents within 1 year at −20 °C. The *E. coli* S30 cell extract is more sensitive and should be stored at −80 °C.
3. Any other *E. coli* strain that contains the *csrA* gene sequence of the *E. coli* K12 strain is equally suitable for this step.
4. The pGE1 plasmid was prepared by insertion of a coding sequence for a $(His)_6$-tag upstream of the FLP recognition target (FRT) site into the pKD4 vector. Furthermore, we introduced restriction enzyme recognition sites (MCSI and MCSII) to be able to excise the cassette comprising the encoded $(His)_6$-tag, the FRT sites and the kanamycin resistance gene of the vector pKD4 [9] (Fig. 3).
5. As an alternative to the pGE1 vector template, you can PCR-amplify the cassette comprising the FRT-sites that flank the kanamycin resistance gene in pKD4 [9] using oligonucleotide primers that encode the $(His)_6$-tag and the *Xho*I and *Not*I restriction sites of MCSI and MCSII, respectively.
6. Pay extreme caution when handling the Erlenmeyer bottle because the solution may spontaneously spill over due to superheating. Always wear heat-proof gloves, a lab coat, and safety goggles and try to use a fume hood when handling hot agarose solutions.
7. Avoid excessive boiling of the agarose solution after it is completely dissolved, as this causes evaporation and will increase the final concentration of agarose in your gel.
8. Ethidium bromide is toxic and a known mutagen. Always wear gloves and a lab coat and use a fume hood when working with ethidium bromide. Dispose ethidium bromide waste following the regulations of your institution.
9. Alternative centrifuge bottle sizes can also be applied. However, we find that 1 L centrifugation bottles are very con-

venient for cell harvest and minimize the time to harvest and wash the cell pellets.

10. DEPC is toxic and should always be handled under a fume hood while wearing gloves, safety goggles, and a lab coat. Freshly opened DEPC bottles may spontaneously spill over due to release of excess pressure.
11. You can alternatively distribute the 4 L of S30 buffer into bottles of smaller volume.
12. Prepare S30 buffer immediately before the intended application and use it up within 1–2 days. Store S30 buffer at 4 °C with a tightly closed cap to minimize air oxidation.
13. You can alternatively use a cell cracker for cell lysis; however, this procedure usually dilutes the cell lysate and necessitates increased efforts to concentrate the cell extract at the end of the preparation protocol. Cell disruption by sonication usually reduces the activity of the final cell extract and should therefore be avoided.
14. Certain amino acids, e.g., tyrosine, have a low solubility in water and will not completely dissolve in the stock solution/suspension. You should therefore always quickly vortex the stock suspension before use.
15. Add pyruvate kinase as final component immediately before mixing the preincubation buffer with the cell extract.
16. For the analysis of alternative sRNA-regulated systems, simply transfer the sRNA-responsive 5′-UTR of the system of interest directly upstream to the CAT coding region in the pIVEX1.3-CAT vector.
17. Diligently follow the DNA plasmid purification procedure to eliminate all RNase activities that may be dragged along from the RNase-containing re-suspension buffer.
18. The concentration of the sRNA stock solution should be in the range of 100 μM. The working solution usually contains 1–20 μM sRNA in DEPC-H_2O.
19. We commonly use the pTx1 vector for plasmid-based runoff transcription of target RNAs (Fig. 8).
20. The RsmE protein can be stored at 100 μM in S30 buffer at −20 °C.
21. Refer to [12] for a comprehensive list of reagent suppliers.
22. Sodium azide is extremely toxic and must be handled in a fume hood while wearing gloves, safety goggles, and a lab coat. Dispose sodium azide-containing waste following the regulations of your institute.
23. It is extremely important to include BSA in both DB and RB buffers! In the absence of BSA we observe nonspecific adsorp-

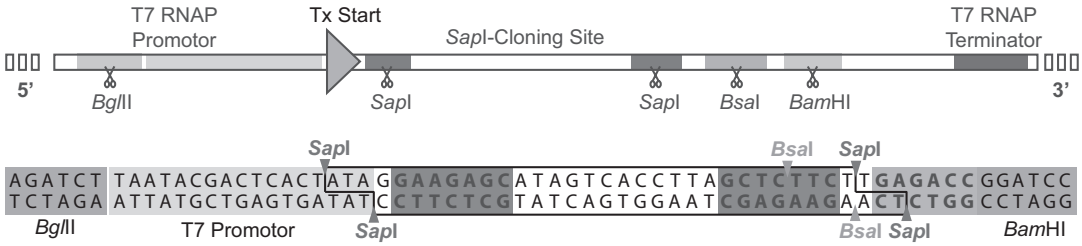


Fig. 8 Template vector pTx1 for runoff transcription of RNA. The DNA for transcription of the sRNA of interest is seamlessly inserted into the pTx1 vector using the SapI restriction sites. The obtained plasmid encoding the sRNA of interest is then amplified and linearized with the BsaI restriction site for runoff transcription by T7 RNA polymerase (T7 RNAP)

tion of CAT to surfaces which results in erroneous data interpretation.

24. Acetyl-CoA in the RB is susceptible to hydrolysis and the formed product will immediately react with DTNB. The RB should ideally be prepared freshly before each usage. However, RB can also be stored at −20 °C up to 2 weeks.
25. To ensure reliable data, you should use a spectrophotometer equipped with temperature control. Furthermore, a spectrophotometer that supports the simultaneous measurement of multiple cells dramatically speeds up the required measurement time. We routinely use a Cary 300 Bio spectrophotometer set to 20 °C to simultaneously follow the A_{412} in 12 cells.
26. Always wear gloves, safety goggles, and a lab coat when working with UV light. To avoid UV-induced DNA damage, you should work speedy to minimize the exposure time of your DNA to the UV light.
27. Expected size of PCR products and the pGE1 insert: 5′-UTR-csrA (301 bp), 3′-UTR (268 bp), pGE1 insert (1341 bp).
28. Depending on the efficiency of the DNA ligation reaction, you may have to use more of the template DNA in the PCR reaction to obtain sufficient amounts of the PCR product. To avoid excessive dilution of the PCR reaction mixture, you can use ethanol precipitation as described [14] to concentrate the DNA in the ligation reaction mixture.
29. The expected size of the PCR product comprising the ligated 5′-UTR-csrA/pGE1 insert/3′-UTR is 1889 bp.
30. Use ultrapure water to elute the purified PCR product from the gel extraction column.
31. The kanamycin resistance cassette can optionally remain integrated in the *E. coli* genome without affecting the activity of the cell extract.

32. Common laboratory consumables such as 1.5 mL or 50 mL tubes and pipette tips should be purchased as certified RNase-free materials and therefore do not require treatment with RNase-AWAY and DEPC-H_2O.
33. Pay attention to maintain a sterile working environment.
34. The main cell culture usually requires ca. 3–4 h to reach an OD_{600} of 0.8–1.0.
35. Use an electric pipette controller or a magnet stirrer with a magnet bar for fast and convenient re-suspension of the cell pellet.
36. The preincubation reaction usually generates significant amounts of precipitate. This is normal and will not negatively affect the activity of the resulting cell extract.
37. You can accelerate the passage of the cell extract over the Ni-NTA resin by gentle application of pressure on top of the column, e.g., by using the pump of a pipette controller.
38. Concentration of cell extract with centrifugal filter devices can take up to several hours. You should occasionally clear the membranes of the centrifugal filter device by gentle re-suspension of the cell extract with a pipette.
39. The cell extract can be stored at −80 °C for at least 6–12 months without a significant reduction of activity.
40. The optimal plasmid concentration can be determined by incrementing the plasmid concentration, e.g., by proceeding in 5 ng/μL steps from 0 to 120 ng plasmid per μL reaction mixture (Fig. 6).
41. Modification of the plasmid concentration directly affects the amounts of transcribed mRNA in the translation assay. A plasmid concentration that yields ca. 50% of the maximally attainable CAT expression level is a compromise between CAT activity, which increases with its expression level, and avoiding saturation of the ribosomes with the transcribed mRNA.
42. The translation activation assay is most sensitive at 50% translation repression due to the steep slope of the sigmoidal dose-response curve at that point.
43. Use the previously determined optimal plasmid concentration (*see* **Note 41**) to set up translation reactions containing various repressor protein concentrations ranging from, e.g., 10 nM to 1 μM and analyze the effect on CAT translation repression (Fig. 4).
44. Although it is also possible to distribute the pCFX100 plasmid and the repressor protein individually to each reaction tube, we strongly discourage this procedure due to the potential error introduced by pipetting small volumes.

45. A typical concentration series of sRNA in the translation reaction includes, e.g., 0, 30, 60, 90, 120, 160, 200, 300, 500, 1000, and 2000 nM of the sRNA. Depending on the translation activation potential of the investigated sRNA, you may have to reduce or expand the analyzed concentration range to obtain complete coverage of the transition from repressed to fully activated translation.
46. The cell-free translation reaction is typically finished after ca. 2 h due to consumption of dNTPs, amino acids, and energy equivalents and due to the accumulation of inhibitory by-products. However, prolonging the incubation time to 3.5 h does not negatively affect the subsequent CAT assay and ensures that all reactions come to completion.
47. The ideal slope for a linear fit is obviously linear. However, enzymatic rate assays typically start with an initial rate which is approximately linear and progressively slow down as the substrate is consumed (Fig. 7). It is therefore of critical importance that your reactions proceed slow enough to catch the initial rate phase and fast enough to generate significant reaction product to increase A_{412}. Reactions that proceed too fast should be stopped and repeated using less reaction supernatant. Reactions with almost no detectable increase in A_{412} should also be stopped and repeated using more reaction supernatant. We usually adjust the amounts of reaction supernatant to obtain curves that increase by 0.1–0.2 arbitrary A_{412} units in 10 min.
48. The reducing agents present in the translation mixture react with DTNB and cause a small increase in A_{412}. In addition, the slightly basic pH of the RB promotes hydrolysis of acetyl-CoA and additionally contributes to a small background increase in A_{412}.
49. The maximal value is obtained when the sigmoidal dose-response curve approaches the upper horizontal asymptote. It is therefore very important to use a concentration series of sRNA in the translation assay that completely covers the transition from repressed to fully activated translation (*see* **Note 45**).

References

1. Waters LS, Storz G (2009) Regulatory RNAs in bacteria. Cell 136(4):615–628. https://doi.org/10.1016/j.cell.2009.01.043
2. Gripenland J, Netterling S, Loh E, Tiensuu T, Toledo-Arana A, Johansson J (2010) RNAs: regulators of bacterial virulence. Nat Rev Microbiol 8(12):857–866. https://doi.org/10.1038/nrmicro2457
3. Romeo T, Vakulskas CA, Babitzke P (2013) Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. Environ Microbiol 15(2):313–324. https://doi.org/10.1111/j.1462–2920.2012.02794.x
4. Lapouge K, Schubert M, Allain FH, Haas D (2008) Gac/Rsm signal transduction pathway of gamma-proteobacteria: from RNA

recognition to regulation of social behaviour. Mol Microbiol 67(2):241–253. https://doi.org/10.1111/j.1365-2958.2007.06042.x

5. Duss O, Michel E, Diarra Dit Konte N, Schubert M, Allain FH (2014) Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. Nucleic Acids Res 42(8):5332–5346. https://doi.org/10.1093/nar/gku141
6. Mercante J, Edwards AN, Dubey AK, Babitzke P, Romeo T (2009) Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression. J Mol Biol 392(2):511–528. https://doi.org/10.1016/j.jmb.2009.07.034
7. Duss O, Michel E, Yulikov M, Schubert M, Jeschke G, Allain FH (2014) Structural basis of the non-coding RNA RsmZ acting as a protein sponge. Nature 509(7502):588–592. https://doi.org/10.1038/nature13271
8. Chen WP, Kuo TT (1993) A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. Nucleic Acids Res 21(9):2260
9. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97(12):6640–6645. https://doi.org/10.1073/pnas.120163297
10. Duss O, Maris C, von Schroetter C, Allain FH (2010) A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. Nucleic Acids Res 38(20):e188. https://doi.org/10.1093/nar/gkq756
11. Schubert M, Lapouge K, Duss O, Oberstrass FC, Jelesarov I, Haas D, Allain FH (2007) Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. Nat Struct Mol Biol 14(9):807–813. https://doi.org/10.1038/nsmb1285
12. Michel E, Wüthrich K (2012) High-yield *Escherichia coli*-based cell-free expression of human proteins. J Biomol NMR 53(1):43–51. https://doi.org/10.1007/s10858-012-9619-4
13. Etezady-Esfarjani T, Hiller S, Villalba C, Wüthrich K (2007) Cell-free protein synthesis of perdeuterated proteins for NMR studies. J Biomol NMR 39(3):229–238. https://doi.org/10.1007/s10858-007-9188-0
14. Sambrook J, Russell DW (2006) Standard ethanol precipitation of DNA in microcentrifuge tubes. CSHarbor Protoc 2006(1). https://doi.org/10.1101/pdb.prot4456

Part III

Quantitation and Subcellular Localisation of sRNA

Chapter 12

Quantitative Super-Resolution Imaging of Small RNAs in Bacterial Cells

Seongjin Park, Magda Bujnowska, Eric L. McLean, and Jingyi Fei

Abstract

We present a method for the quantification of small regulatory RNAs (sRNAs) in bacteria, by combining single-molecule fluorescence in situ hybridization (smFISH), super-resolved single-fluorophore microscopy, and clustering analysis. Compared to smFISH imaging with diffraction-limited fluorescence microscopy, our method provides better quantification for short and abundant RNA (such as sRNAs) in a small volume of bacterial cells. Our method can also be directly used for the quantification of messenger RNAs (mRNAs).

Key words Fluorescence in situ hybridization, Super-resolution microscopy, Clustering analysis, Small regulatory RNA

1 Introduction

Bacterial small regulatory RNAs (sRNAs) regulate gene expression and provide growth benefit for bacterial cells under stress conditions. They are typically 50–300 nucleotides in length, and are often associated with Hfq for their functions [1, 2]. In most common scenarios, sRNAs regulate gene expression posttranscriptionally by base pairing with target messenger RNAs (mRNAs) and thereby affect translation efficiency and/or mRNA stability [1, 2]. Single-molecule fluorescence in situ hybridization (smFISH) is widely used to count mRNAs at the single-cell level [3, 4], and has been applied to study gene expression, especially in eukaryotic systems. In smFISH, many, often dozens, fluorescently labeled DNA oligos are hybridized to different portions of the same mRNA to generate bright signal as diffraction-limited spots with conventional microscopy [4, 5]. However, direct quantification by counting spots from smFISH signal cannot be applied to bacterial sRNAs due to their short length that limits the number of hybridized DNA oligos, and therefore the signal brightness, and the usually high copy number of sRNAs in the small volume of a bacterial cell (Fig. 1).

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_12,

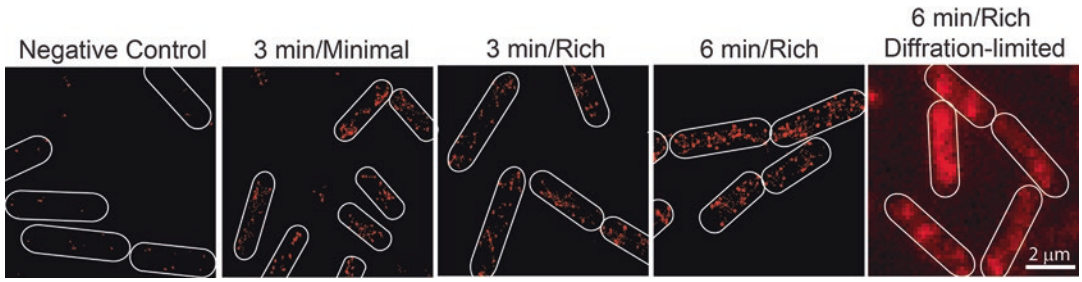


Fig. 1 Presentative images of SgrS. From the left to right are: negative control for probe nonspecific binding with Δ*sgrS* strain by SR imaging, 3 min post-induction by αMG in MOPS minimal medium by SR imaging, 3 min post-induction by αMG in MOPS rich medium by SR imaging, 6 min post-induction by αMG in MOPS rich medium by SR imaging, 6 min post-induction by αMG in MOPS rich medium by diffraction-limited microscopy. Cell boundaries are shown with white lines

Here, we provide a protocol of combining smFISH with super-resolved single-fluorophore microscopy (such as stochastic optical reconstruction microscopy (STORM), photoactivated localization microscopy (PALM), and fluorescence photoactivation localization microscopy (FPALM), and hereafter referred as SR microscopy) [6–9], and determination of the copy number of sRNA in individual bacterial cells with clustering analysis [10]. In SR imaging, photo-switchable fluorophores are used to label the sample of interest such that at any given time, only a subset and sparsely localized fluorophores are activated into the bright state and the centroids of these fluorophores are pinpointed with nanometer accuracy [7]. Repetitive activation and imaging allow reconstruction of final images with about a 10-fold enhancement in spatial resolution compared to the diffraction-limited microscopy. A density-based clustering analysis algorithm, DBSCAN [10], is then applied to 3D reconstructed SR images to segregate individual spots into clusters, which allows further determination of RNA copy number.

We use an sRNA, SgrS, which is involved in glucose-phosphate stress in *E. coli* [11–13], as an example to illustrate the method. The same method can be used to quantify mRNAs [13]. We will include bacterial sample preparation, SR imaging, and copy number calculation in the protocol.

2 Materials

2.1 Bacterial Culture

1. MOPS rich defined medium: Solution is made using TEKnova® MOPS EZ Rich Medium Kit. For 500 mL of medium, mix 50 mL of 10× MOPS mixture, 5 mL of K_2HPO_4, 50 mL of 10× ACGU, 100 mL of 5× Supplement EZ, and 295 mL of water. Medium is sterilized using 0.2 μm filter. Store in the dark at room temperature.

2. MOPS minimal medium: For 500 mL of medium, mix 50 mL of 10× MOPS mixture and 5 mL of K_2HPO_4 and 445 mL of water. Medium is sterilized using 0.2 μm filter. Store in the dark at room temperature.

Carbon sources are added depending on the experimental needs. In our example case, 0.2% glucose and 0.2% fructose are mixed with MOPS rich or minimal media.

2.2 Labeling of FISH Probe

1. 1 M $NaHCO_3$, pH 8.6: Dissolve 4.2 g of $NaHCO_3$ in 40 mL of water. Adjust the pH with 5 M NaOH, and adjust the final volume to 50 mL with water.
2. Absolute ethanol.
3. 3 M NaOAc, pH 5: Dissolve 12.3 g of NaOAc in 40 mL of water. Adjust the pH with glacial acetic acid, and adjust the final volume to 50 mL with water.
4. Silicone-based purification columns, such as Bio-Spin P-6 column (Bio-Rad™) or G25 column (GE Healthcare Life Sciences™).

2.3 Fixation and Hybridization

1. 37% formaldehyde.
2. Deionized formamide.
3. Phosphorus-buffered saline (1×), pH 7.4, PBS.
4. Saline-sodium citrate (20×), SSC.
5. FISH wash buffer: 10% formamide in 2× SSC. Add 5 mL of 20× SSC and 5 mL of deionized formamide to 40 mL of water. Store at 4 °C.
6. FISH hybridization buffer: 10% dextran sulfate and 10% formamide in 2× SSC. Weigh out 1 g of dextran sulfate (average molecular weight >500,000) and add to a 15 mL conical tube. Add 1 mL of 20× SSC, 1 mL of formamide, and 7 mL of water to the tube. Incubate at room temperature with shaking or vortexing until the dextran sulfate dissolves. Adjust the final volume to 10 mL with water. Divide the solution into aliquots and store at −20 °C.

2.4 Microscope

SR microscopy setup can be configured in many ways, and here we describe our example (Fig. 2). An inverted optical microscope (Nikon Ti-E with 100× NA 1.49 CFI HP TIRF oil immersion objective) can be combined with laser illumination, such as a red laser (647 nm, 120 mW, Cobolt MLD) and a violet laser (405 nm, 25 mW, CrystaLaser), through fiber coupling. Laser excitation lights are reflected by a dichroic mirror (Chroma zt405/488/561/647/752rpc-UF3) for near-TIRF excitation. The emission signal is collected by the objective lens, filtered by the emission filter (Chroma ET700/75 m), and imaged by a

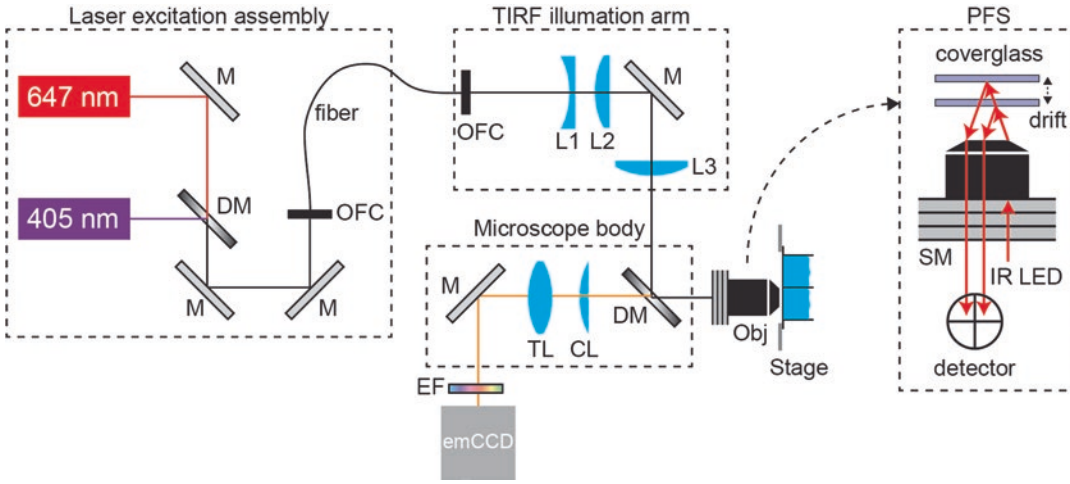


Fig. 2 Schematic illustration of the microscope setup. *M* mirror, *DM* dichroic mirror, *OFC* optical fiber coupling, *L1 & L2* lenses for expanding the beam, *L3* lens at the back focal plane, *Obj* objective lens, *CL* cylindrical lens, *TL* tube lens, *EF* emission filter, *SM* step motor. *PFS* Perfect focus system from Nikon, utilizing near-infrared 870-nanometer LED and CCD line sensor, is used to maintain the *z* stability

1024 × 1024 pixels EMCCD camera (Andor iXon Ultra 888). For 3D imaging, a cylindrical lens with 10 m focal length (CVI RCX-25.4-50.8-5000.0-C-415-700) is inserted in the emission path to cause the astigmatism effect of fluorophores. Single molecule samples (labeled DNAs or proteins immobilized on a coverslip) can be imaged in different *z*-planes to correlate the astigmatism effect with the corresponding *z* values. Then by interpolation and fitting, a calibration curve is created to obtain *z* values of single fluorophores in actual imaging [7]. To maintain the *z*-focus while imaging, Nikon Perfect Focus System (PFS) is used. The laser power density on the sample is about 2000 W cm^{-2} for the red laser, and the maximum power density for the violet laser is about 130 W cm^{-2}.

2.5 Imaging

1. 2-Metacaptoethanol (BME, Sigma-Aldrich M6250).
2. Catalase (EMD Millipore 219001): Make aliquots of 454.5 kU/mL =33.6 mg/mL in 50% glycerol/1× PBS and keep them in −20 °C.
3. Glucose oxidase (Sigma-Aldrich G2133): Make aliquots of 100 mg/mL (>10 KU/mL) in 50% glycerol/1× PBS and keep them in −20 °C.
4. The imaging buffer: 10 mM NaCl, 50 mM Tris and 10% glucose in 1× PBS (for immunostained samples) or 4× SSC (for FISH samples), with pH adjusted to 8.
5. 8-Well chambered cover glass (Nunc Lab-Tek 155409 or Cellvis C8-1.5H-N).
6. Poly-L-lysine (Sigma Aldrich P8920).
7. Nikon NIS-Elements software for image acquisition.

2.6 Image Processing and Data Analysis

1. IDL for reconstruction of 3D SR images.
2. MATLAB for copy number analysis.

3 Methods

3.1 FISH Probe Design, Labeling, and Purification

1. FISH probes are designed using Stellaris® Probe Designer 4.2 (*see* **Note 1**). Probes contain an amine modification for fluorophore conjugation.
2. The following protocol for probe labeling is for a total reaction volume of 15 μL. Add 1.5 μL of 1 M $NaHCO_3$ (pH 8.6) to 13.0 μL 100 μM amine-modified DNA oligo and mix. In a separate tube, dissolve 0.025 mg of fluorophore in 0.5 μL DMSO. Mix the dye solution with the oligo solution and incubate at 37 °C overnight in the dark (*see* **Note 2**).
3. On the next day, purify the labeled DNA oligos from the free dyes with ethanol precipitation. Add 1/9 reaction volumes of 3 M NaOAc (pH 5) and 2.5 volumes (of total volume including reaction and NaOAc solution) of 100% ethanol to the labeling reaction. Incubate for at least 6 h at −20 °C in the dark. Centrifuge the probe solution at full speed (~21,000 × *g*) for 30 min and then resuspend the pellet in 30–40 μL of water.
4. Further purify the labeled oligo from the residual free dyes with silica-based purification column, according to the manufacturer's instruction.
5. Take the UV-Vis absorption spectrum of the sample to calculate the concentrations of the DNA oligo, the fluorophore, and the labeling efficiency (defined by the ratio of the concentration of the fluorophore to the concentration of the oligo).

3.2 Culture Preparation, Fixation, and Hybridization

SgrS is used as an example here for the demonstration of the imaging and copy number calculation. Along with the wild-type strain, a *ΔsgrS* strain is also prepared as a negative control. SgrS is induced with methyl α-D-glucoside (αMG). A sample containing a relatively low-copy number and well-separated SgrS is prepared for RNA copy number calculation (referred as "single-RNA calibration sample," and *see* Subheading 3.5). In this case, we use SgrS induced for 3 min in MOPS minimal medium as the single-RNA calibration sample. Samples with SgrS induced for 3 min and 6 min in MOPS rich medium are prepared and analyzed as additional examples for comparison (Fig. 1).

1. Grow bacteria strains overnight in MOPS EZ rich medium (Subheading 2.1) supplemented with 0.2% glucose and 0.2% fructose at 37 °C at 250 rpm in a shaker. On the next day, dilute the overnight cultures 100-fold using fresh medium:

10 mL culture of *ΔsgrS* strain in MOPS EZ rich medium, 10 mL culture of wild-type strain in MOPS minimal medium, and 20 mL culture of wild-type strain in MOPS EZ rich medium. In each case supply the carbon source as 0.2% glucose and 0.2% fructose. Cells are grown again at 37 °C at 250 rpm until OD_{600} reaches 0.2–0.3 for SgrS induction.

2. Add 50% αMG to the cultures to a final concentration of 0.5% to induce SgrS. Harvest the cells (10 mL each) at different time points post induction by directly adding 1.2 mL 37% formaldehyde to the culture (final concentration of 4%) to fix the cells.
3. Cells mixed with formaldehyde are incubated for 30 min on a nutator at room temperature.
4. Collect the cells by centrifuging at 4000 × *g* for 10 min at room temperature.
5. Wash cells twice in the reaction volume of 1× PBS by resuspending, centrifuging at 600 × *g* for 3.5 min, and removing the supernatant. The reaction volume is calculated such that the density of the cells is ~5×10^9 per mL (*see* **Note 3**).

$$\text{Reaction volume}(\text{mL}) = (\text{Number of cells from the culture}) / 5 \times 10^9$$

6. Permeabilize cells by resuspending the pellet in the reaction volume of 70% ethanol. To avoid clumping of cells, first resuspend cells in water and then mix with 100% ethanol to a final concentration of 70% ethanol. Incubate cells in 70% ethanol for at least 1 h at room temperature on a nutator (*see* **Note 4**).
7. For hybridization, take 60 μL of cells in 70% ethanol. Centrifuge for 7 min at 600 × *g* and remove supernatant. Resuspend with 100 μL of 10% FISH wash solution and leave for a few minutes.
8. Prepare the hybridization mix in a fresh tube. Mix the labeled oligos with the appropriate amount of FISH hybridization buffer. The final concentration is 50 nM per probe for sRNA and 15 nM per probe for mRNA.
9. Centrifuge cells for 7 min at 600 × *g* and remove supernatant. Add 15 μL of hybridization mix to each cell sample and mix, avoiding air bubbles. Incubate at 30 °C overnight.
10. On the next day, add 200 μL of FISH wash buffer to each hybridization reaction, mix, centrifuge at 600 × *g* for 7 min, and remove supernatant.
11. Resuspend the pellet with 200 μL of FISH wash buffer and incubate for 30 min at 30 °C. Afterward, centrifuge and remove supernatant. Repeat the wash step two more times. Resuspend the final cell pellet in 20 μL of 4 × SSC (*see* **Notes 5** and **6**).

12. To immobilize the cells, an 8-well chambered cover glass is coated with poly-L-lysine for 45 min, then washed with water three times and air-dried. Per each well, 5 μL of the cell resuspension from **Step 11** are diluted with 125 μL of 4× SSC and incubated in each well for >1 h in 4 °C for immobilization.

3.3 Imaging

1. Imaging buffer is mixed with oxygen scavenger system right before imaging. Add 5 μL of BME, 3 μL of Glucose oxidase, and 1 μL of catalase to 491 μL of the imaging buffer, and transfer the mixed buffer to the imaging well. Incubate the sample in the imaging buffer with oxygen scavenger and BME for a few minutes (*see* **Note 7**).
2. A DIC image is taken for an area of interest. Then count the number of cells in the imaging area (*see* **Note 8**).
3. Imaging acquisition is done with repetitive cycles with the sequence of violet (405 nm, 1 frame) → red (647 nm, 3 frames) as a single cycle. The imaging acquisition starts with 0 $W \cdot cm^{-2}$ violet laser power. The violet laser power is modulated so that the number of "blinking-on spots" is kept above 50% of the number of cells in the field of view. When the number of "blinking-on" spots reaches less than this, even with the maximum violet laser power (130 $W \cdot cm^{-2}$), the acquisition is terminated. The controlled and automatic acquisition is coded in the Jobs module in NIS-Element.

3.4 Imaging Reconstruction

The data analysis is based on a previously published algorithm [7, 14]. An analysis package coded in IDL is used for reconstruction of 3D SR images.

1. Peak Identification and Fitting: To remove the noise, the image is blurred by Gaussian convolution of 9 × 9 pixels.
   (a) A user-defined intensity threshold (usually ~4) is introduced. Then the software finds all the pixels whose values are greater than "threshold" × "standard deviation of all the pixel values of the entire frame."
   (b) The software finds the local maximum intensity pixels whose pixel values are greater than their 24 surrounding pixels.
   (c) The software removes overlapping spots or bright junks by applying sharpness (ranging from 0.4 to 3) filter. Sharpness is defined as the intensity ratio between the peak and the background.
   (d) For peaks filtered so far, the area of 19 × 19 pixels surrounding local maximum intensity pixel is fitted with an Elliptical Gaussian function,

$$G(x,y) = h\exp\left(-2\frac{(x-x_0)^2}{w_x^2} - 2\frac{(y-y_0)^2}{w_y^2}\right) + b$$

where (x_0, y_0) is the position of the peak, w_x and w_y are its widths, and h and b are remaining fitting constants.

2. Horizontal Drift Correction: Drift Correction by fast Fourier transformation. The dataset is divided into an equal number of frames, and sub-SR images appearing similar to each other are obtained, except for the effect of drift. Then the fast Fourier transformation (FFT) is applied to each sub-SR image. By comparing the center images from each transformed image, the software determines the relative drift among sub-SR images. By linear interpolation, the drift is corrected through all the frames (*see* **Note 9**).

3.5 Copy Number Analysis

Copy number calculation is performed with analysis package coded in MATLAB. The package is free to distribute upon request.

1. Clustering analysis: Detecting clusters from the raw data is necessary for cleaning the nonspecific background of the raw data, as well as defining each cluster resulting from single or multiple RNAs. The detected clusters are then used for RNA copy number calculation. We use DBSCAN [10], a density-based algorithm, for cluster detection (Fig. 3). DBSCAN requires two input parameters as Npts and Eps (*see* **Note 10**). A "core spot" is defined if there are more than Npts spots (including itself) around it within Eps distance. As an example, Fig. 3A (upper penal) shows a case in which Npts = 7, hereby starting with two core spots, colored as red and blue. Any spot within the Eps distance of a core spot becomes a new core spot if more than Npts spots can be found within Eps distance around it, and becomes a part of the initial cluster. In this way, clusters can expand (Fig. 3A, the center panel). On the other hand, any spot within Eps distance from a core point becomes a "border spot," if it is surrounded by fewer spots than Npts within Eps distance. In this way, the cluster stops expansion (Fig. 3A, the lower panel). Spots not belonging to any cluster are considered background noise and are not considered further for the following analysis (Fig. 3A, the black spots). After the cluster analysis, we obtain the following information: (1) the total number of clusters in each cell, (2) the number of spots in each identified cluster, (3) the total number of spots in clusters in each cell, and (4) the average "radius" per each cluster. The radius of a cluster is defined as the average distance between the cluster center (the average of all the coordinates of spots in the cluster) and all the spots in the cluster.
2. Baseline correction: DBSCAN analysis gives the total number of "clustered" spots in each cell, but not all those spots come from the real RNAs. Even with the sequence-specific probes to label the RNAs of interest, nonspecific binding of probes yields

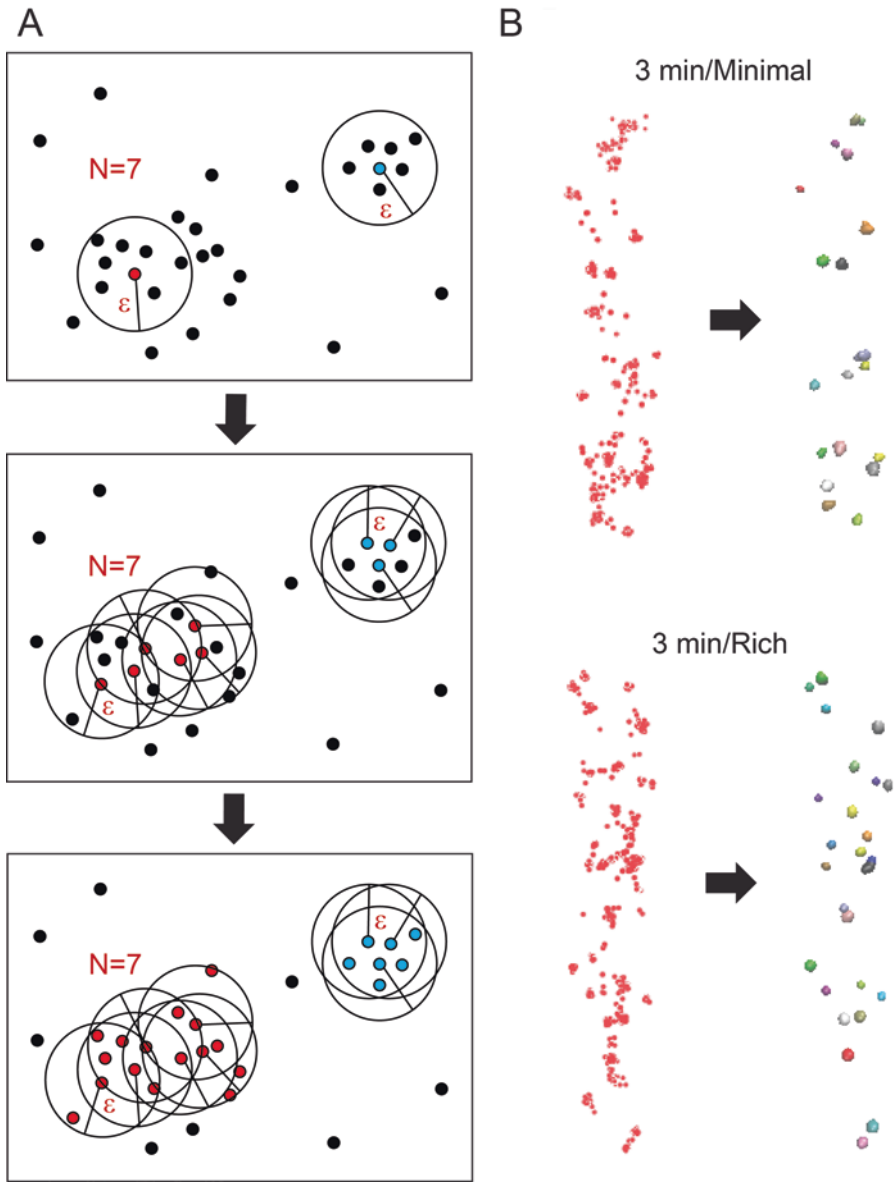


Fig. 3 Clustering analysis. (**A**) Illustration of DBSCAN algorithm. (**B**) Examples of DBSCAN on individual cells. Left: raw data from SR imaging; right: clustered data from DBSCAN, with each color representing each cluster

the false positive signal (the background signal) in each cell. Therefore, the tested samples all need to be corrected for the contributions from the background signal.

(a) Building a baseline for background signal: This background varies in different cells, as shown in the case of Δ*sgrS* cells labeled by SgrS FISH probes (Fig. 1, left most

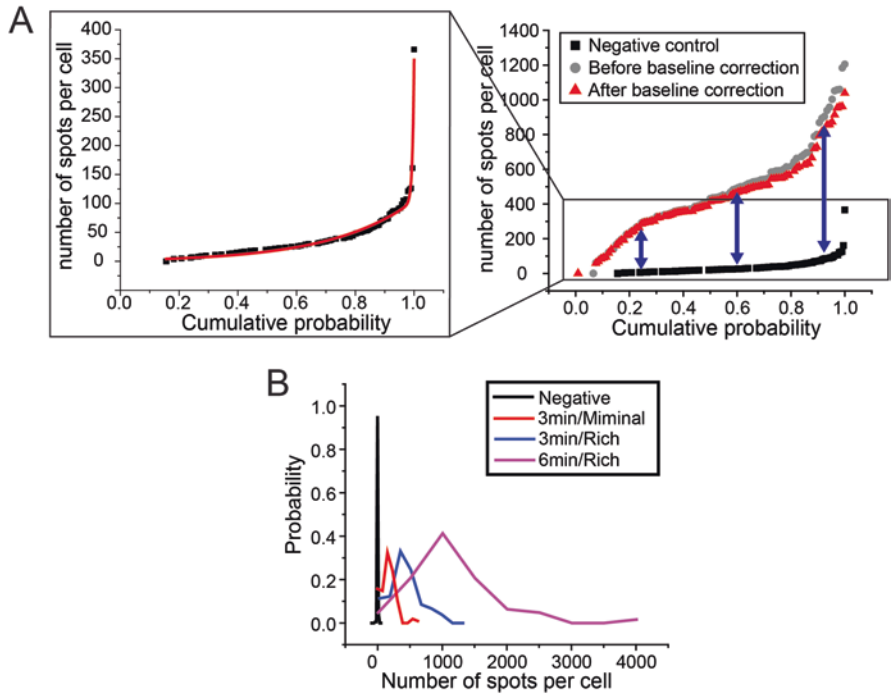


Fig. 4 Baseline subtraction for nonspecific binding. (**A**) *x–y* Inversed cumulative probability distribution of number of spots per cell in negative control is fit with double-exponential growth, which serves as a baseline for subtracting signal from probe nonspecific binding. (**B**) Distribution of the number of spots per cell after baseline subtraction

panel). We plot the cumulative probability distribution of background signal in individual cells by sorting the number of spots per cell from least to most (Fig. 4A, black points), and determine a baseline function by fitting this cumulative probability curve with a double exponential function (Fig. 4A, the red curve).

$$N_B = B + A_1 \exp\left(E_1 x - C_1\right) + A_2 \exp\left(E_2 x - C_2\right)$$

where x is the cumulative probability and N_B is the number of spots per cell. This function needs to be updated based on one's own negative strain collected with the SR microscopy setup.

(b) Baseline subtraction: The number of spots in each cell of the tested samples are sorted the same way, and subtracted by the baseline level of N_B calculated from the fitting curve obtained from the **step** (**a**) (Fig. 4B). Then the baseline-corrected number of spots per cells are used for the downstream analysis (Fig. 4C).

3. Single RNA characterization: This step requires single-RNA calibration sample, i.e., cells with low copy number and well-separated (not granule or transcription site forming) of

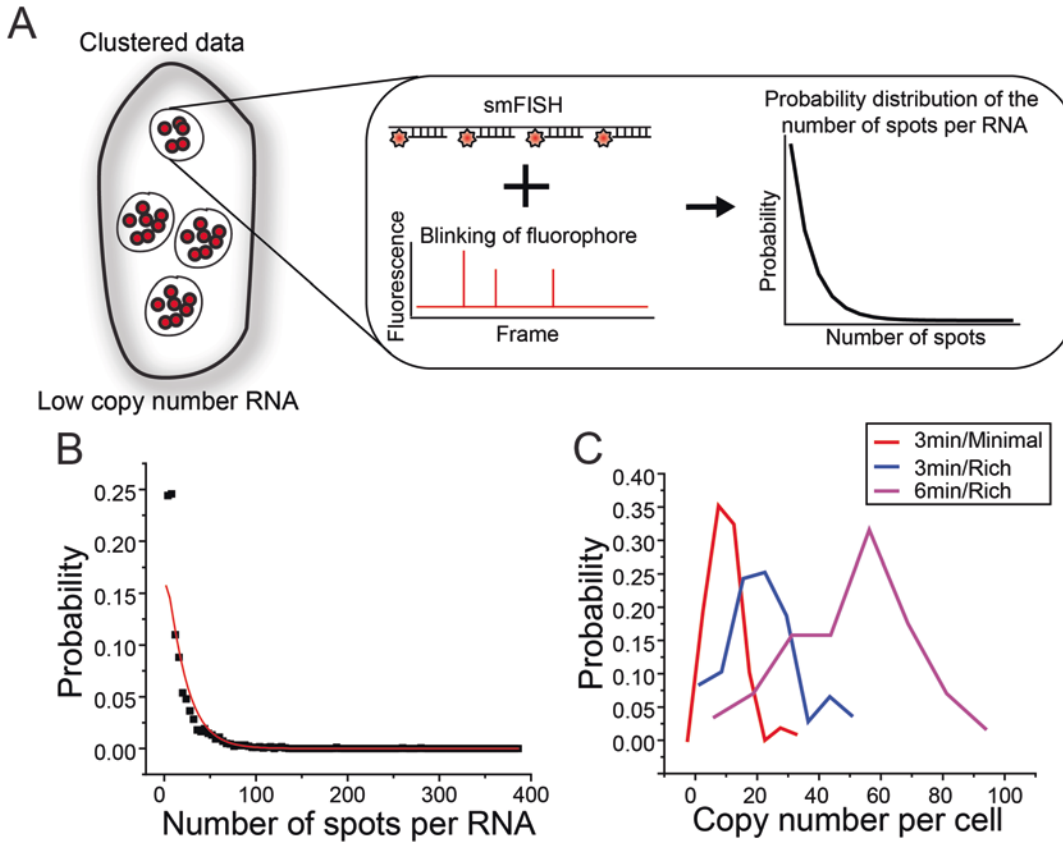


Fig. 5 Copy number calculation. (**A**) Illustration of copy number calculation. Single-RNA calibration sample is used to estimate number of spots per RNA, in which individual clusters are approximated as individual RNAs. The heterogeneity in the number of spots per RNA (cluster) is contributed by both the heterogeneities of number of hybridized probes per RNA and the number of spots (blinking events) generated per fluorophore. The resulting distribution of the number of spots per RNA (cluster) is empirically described by a negative binomial distribution. (**B**) Probability distribution of the number of spots per RNA from single-RNA calibration sample. (**C**) RNA copy number distributions in the three cases in Fig. 1

RNAs (Fig. 5A), such as cells fixed a short time after sRNA induction (Figs. 1 and 3 min/Minimal). In this case, we assume that DBSCAN analysis generates each distinct cluster corresponding to each RNA, and thus the number of spots per cluster represents the number of spots per RNA (N_0). We empirically fit the probability distribution of N_0 with a negative binomial distribution. This fitting is based on the assumption that each blinking event from each RNA is a Bernoulli trial, with the "ON" probability p, and we conduct imaging until r number of "OFF" states occur. We conduct imaging for a sufficiently long time so that the final event is always "OFF," as photobleaching. Then the number of "ON" states, i.e., the number of recorded spots per RNA (N_0) is fitted by the negative binomial distribution (Fig. 5A, B):

$$P(N_0;r,p)=\binom{N_0+r-1}{N_0}(1-p)^r\,p^r$$

These N_0 values are the observables which we extract from DBSCAN analysis for multiple clusters and from multiple cells, and by fitting this data we obtain p and r values (Fig. 5B).

4. Generation of $P(N|C)$ matrix: To obtain the number of RNAs in the cases of high-expression level, or granule forming, we utilize the convolution of negative binomial distribution. When there are RNAs with a copy number of C in a cell, given the distribution of N_0 per RNA, we can expect the distribution of a total number of spots (N) given the copy number of C to be:

$$P(N|C)=P(N;r^*C,p)=\binom{N+Cr-1}{N}(1-p)^{Cr}\,p^r$$

Based on this, we simulate a matrix of $P(N|C)$:

$$\tilde{P}(n|c)=\begin{bmatrix} P(N'=1|C'=1) & \cdots & P(N'=1|C'=c) \\ \vdots & \ddots & \vdots \\ P(N'=n|C'=1) & \cdots & P(N'=n|C'=c) \end{bmatrix}$$

5. Copy number calculation: According to Bayes' theorem and the law of total probability, probability of cells with N spots having RNA copy number of C can be calculated by:

$$P(C|N)=\frac{P(N|C)P(C)}{P(N)}=\frac{P(N|C)}{\sum_{C'}P(N|C')P(C')}P(C)$$

Assume $P(C')$ is uniform for all possible copy numbers, i.e., P(C′) = P(C),

$$P(C|N)=\frac{P(N|C)}{\sum_{C'}P(N|C')}$$

Then the expectation value of the RNA copy number C under the condition of observing N spots per cell is

$$E(C|N)=\sum_{C'}C'P(C'|N')=\frac{\sum_{C'}P(N|C')C'}{\sum_{C'}P(N|C')}$$

here $P(N|C')$ is obtainable from the constructed probability mass function matrix, $\tilde{P}(n|c)$ from **step 4**. Fig. 5C shows the distribution of RNA copy number per cell for multiple cells in different conditions in SgrS induction (also refer to Fig. 1).

4 Notes

1. Stellaris® Probe Designer 4.2 is a free program that designs probes based on the nucleotide sequence provided by the user. Maximizing the number of probes against a target increases the signal-to-noise ratio during imaging. Make sure that probes are not targeted to sequences involved in stable secondary structures.
2. The reaction can be scaled up when needed. In general, the amount of fluorophore used should be 25–30-fold in excess of the molar concentration of oligos. The exposure of fluorophore to room temperature should be minimized. Wrapping the tube with aluminum foil helps minimize its exposure to light.
3. The number of cells is estimated from OD_{600}. OD_{600} of 1 corresponds to a cell density of 3×10^8–1×10^9 depending on specific spectrometers.
4. Cells in 70% ethanol can be stored at 4 °C for a few weeks.
5. It is optional to perform postfixation; however, it can help preserve the hybridization of probes for longer. For postfixation, wash the cell sample with 200 μL of 1× PBS. Resuspend the pellet in 100 μL of 4% formaldehyde in 1× PBS and incubate for 10 min at room temperature on a nutator. Afterward, centrifuge, remove supernatant, and perform one more wash with 200 μL of 1× PBS.
6. It is highly recommended to image the cells when they are fresh to prevent signal loss due to probe dissociation.
7. Since pH will change over time due to the activity of glucose oxidase, imaging has to be done within ~ 1 h.
8. Cell counting can be done either by eye or by commercial software, such as Nikon NIS-Element's Analysis Explorer → "Cell count" feature.
9. An alternative way to correct the drift is to track the position of a fiducial marker, such as a nano-diamond, during data acquisition.
10. The choices of Npts and Eps are empirical depending on the specific microscopic setup, number of probes per RNA, fluorophore, and imaging conditions. It is recommended to try different combinations of Npts and Eps, and visually validate the clustered data compared to raw data. In addition, experimental validation, such as quantitative PCR, is recommended to compare with the copy number calculation from imaging approach.

Acknowledgments

The data presented as examples were collected during the 2016 Center for the Physics of Living Cells (CPLC) Summer School at University of Illinois at Urbana-Champaign. We therefore thank National Science Foundation grant PHY-1430124 (Physics Frontiers Center for the Physics of Living Cells) for funding for the CPLC Summer School, Zan Luthey-Schulten and Joseph R. Peterson for co-teaching the scientific theme of "Quantitative Imaging and Cell Simulation of Small Regulatory RNA," and the summer school students: Bijoy Desai (Columbia University), Dennis Fernandes (University of Toronto, Mississauga), Jialei Tang (University of Central Florida), and Kailun Zhang (Texas A&M University) for participation in data collection and analysis.

References

1. Wagner EGH, Romby P (2015) Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. Adv Genet 90:133–208
2. Storz G, Vogel J, Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 43:880–891
3. Femino AM, Fay FS, Fogarty K, Singer RH (1998) Visualization of single RNA transcripts in situ. Science 280:585–590
4. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods 5:877–879
5. So L-H, Ghosh A, Zong C, Sepúlveda LA, Segev R, Golding I (2011) General properties of transcriptional time series in Escherichia coli. Nat Genet 43:554–560
6. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF (2006) Imaging intracellular fluorescent proteins at nanometer resolution. Science 313:1642–1645
7. Huang B, Wang W, Bates M, Zhuang X (2008) Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. Science 319:810–813
8. Rust MJ, Bates M, Zhuang X (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods 3:793–795
9. Hess ST, Girirajan TPK, Mason MD (2006) Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophys J 91:4258–4272
10. Daszykowski M, Walczak B, Massart D (2001) Looking for natural patterns in data. Chemom Intell Lab Syst 56:83–92
11. Vanderpool CK, Gottesman S (2004) Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. Mol Microbiol 54:1076–1089
12. Rice JB, Vanderpool CK (2011) The small RNA SgrS controls sugar-phosphate accumulation by regulating multiple PTS genes. Nucleic Acids Res 39:3806–3819
13. Fei J, Singh D, Zhang Q, Park S, Balasubramanian D, Golding I, Vanderpool CK, Ha T (2015) RNA biochemistry. Determination of in vivo target search kinetics of regulatory noncoding RNA. Science 347:1371–1374
14. Bates M, Blosser TR, Zhuang X (2005) Short-range spectroscopic ruler based on a single-molecule optical switch. Phys Rev Lett 94:108101

Chapter 13

Extraction and Analysis of RNA Isolated from Pure Bacteria-Derived Outer Membrane Vesicles

Janine Habier, Patrick May, Anna Heintz-Buschart, Anubrata Ghosal, Anke K. Wienecke-Baldacchino, Esther N.M. Nolte-'t Hoen, Paul Wilmes, and Joëlle V. Fritz

Abstract

Outer membrane vesicles (OMVs) are released by commensal as well as pathogenic Gram-negative bacteria. These vesicles contain numerous bacterial components, such as proteins, peptidoglycans, lipopolysaccharides, DNA, and RNA. To examine if OMV-associated RNA molecules are bacterial degradation products and/or are functionally active, it is necessary to extract RNA from pure OMVs for subsequent analysis. Therefore, we describe here an isolation method of ultrapure OMVs and the subsequent extraction of RNA and basic steps of RNA-Seq analysis. Bacterial culture, extracellular supernatant concentration, OMV purification, and the subsequent RNA extraction out of OMVs are described. Specific pitfalls within the protocol and RNA contamination sources are highlighted.

Key words Bacteria, RNA, Outer membrane vesicle (OMV), Gram-negative, Sequencing, Analysis, Extraction, Ultracentrifugation, Ultrafiltration, Density gradient

1 Introduction

The secretion of outer membrane vesicles (OMVs) is a common phenomenon apparent in many Gram-negative bacteria [1]. OMVs are spherical nanovesicles with an average diameter of 20–200 nm and are composed of LPS, protein, lipids, and nucleic acid [1]. The amount of secreted-OMVs as well as their content is dependent on the bacterial growth environment [2], and therefore OMVs can play a function in intra- and inter-species [3, 4] and even in inter-kingdom communication [5–8]. Interestingly, it has recently been shown that an OMV-associated small RNA molecule disseminated from *Pseudomonas aeruginosa* is able to trigger a functional response in its host target cell [9]. Thus, analogous to observations in eukaryotes it seems possible that OMVs not only protect bacterial extracellular RNA but also represent means of transfer of

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_13,

RNA molecules to recipient cells where they could be functionally active. Whether, this mechanism of inter-kingdom communication is restricted to *Pseudomonas aeruginosa* or if it is a common phenomenon in Gram-negative bacteria still remains unclear.

Bacterial extracellular RNA has recently been described for different bacterial species and environments [10–13]. Bacteria secrete products into the extracellular environment either through their secretory systems via continuous or discontinuous passages across the bacterial membrane or by the release of OMVs [1, 10, 14]. Another origin of extracellular bacterial RNA is the release of such molecules from decomposing or disrupted cells. As the extracellular RNA can thus originate from different sources, it is likely that the different fractions of RNA molecules also present distinct functions. To investigate this theory, it is necessary to analyze the different RNA subpopulations individually.

Therefore, we describe here a protocol on how to isolate non-OMV-associated and OMV-associated RNA. The protocol allows the extraction of total RNA, large RNA, or only small regulatory bacterial RNA molecules. We also highly recommend to control the percentage of dead bacteria in the bacterial culture of interest. Furthermore, it is important to note that OMV-associated RNA is only present in the extracellular environment in very low amounts. Consequently, extracellular RNA-Seq is prone to be enriched in RNA contamination sequences. Therefore, we highly recommend to include controls for contamination, such as extraction blanks.

2 Materials

It is worth noting that all the materials and methods described within this chapter have been applied and optimized for the culturing of *Salmonella enterica* subsp. *enterica* strain LT2 (henceforth referred to *Salmonella*), and thus culturing conditions and biomolecular extraction methods might need to be optimized for other Gram-negative species. Furthermore, we focus within this protocol on the extraction of small RNA out of OMVs, but in principle, this protocol can also be used to extract total RNA from OMVs or from the OMV-free extracellular environment.

All solutions should be prepared using ultrapure water (sensitivity of 18 MΩ-cm at 25 °C) and analytical grade reagents. Prepare and store all reagents at room temperature (unless indicated otherwise). All working areas should be cleaned as much as possible with an RNase decontamination solution, and certified RNase-free lab material should be used during the whole protocol execution. Precisely execute biosafety and standard operation procedures in order to follow all waste disposal regulations when disposing waste materials.

2.1 Minimal Bacterial Culture Media M9 (see Note 1)

1. M9 Salts 5×: 800 mL H_2O and add 64 g Na_2HPO_4-$7H_2O$, 15 g KH_2PO_4, 2.5 g NaCl, 5.0 g NH_4Cl. Stir until dissolved. Adjust to 1000 mL with distilled H_2O. Sterilize by autoclaving. 5× media should be kept at 4 °C until use.
2. 3000 mL of distilled autoclaved H_2O.
3. Autoclaved 50 mL 1 M $MgSO_4$: add 6.0183 g of $MgSO_4$ into 50 mL of distilled H_2O, dissolve, and autoclave (place at 4 °C until use).
4. Autoclaved 50 mL 1 M $CaCl_2$: add 5.549 g of $CaCl_2$ into 50 mL of distilled H_2O, dissolve, and autoclave (place at 4 °C until use).
5. M9 media 1×: add 600 mL of M9 salts 5×, 6 mL of 1 M $MgSO_4$, 300 µL of $CaCl_2$, 12 g of glucose in powder into an autoclaved 5 L Schott bottle. Adjust to 3000 mL with distilled autoclaved H_2O. Filter sterilization should be done using a steritop filter (0.2 µm). 1× M9 media should be prepared on the day of use.

2.2 Bacterial Live/Dead Staining (see Note 2 and Fig. 1)

1. LIVE/DEAD® BacLight™ Bacterial Viability Kit: SYTO 9 dye, 3.34 mM (Component A), 300 µL solution in DMSO, Propidium iodide, 20 mM (Component B), 300 µL solution in DMSO, BacLight mounting oil (Component C).

2.3 Ultrafiltration (see Note 3)

1. Tangential flow filtration device (Quixstand Benchtop System) using 100 kDa hollow fiber membrane and all the following solutions to wash the tangential flow device.
2. Phosphate-buffered saline, PBS 1×: heat to 50 °C.
3. Sterile distilled water, 5 L.

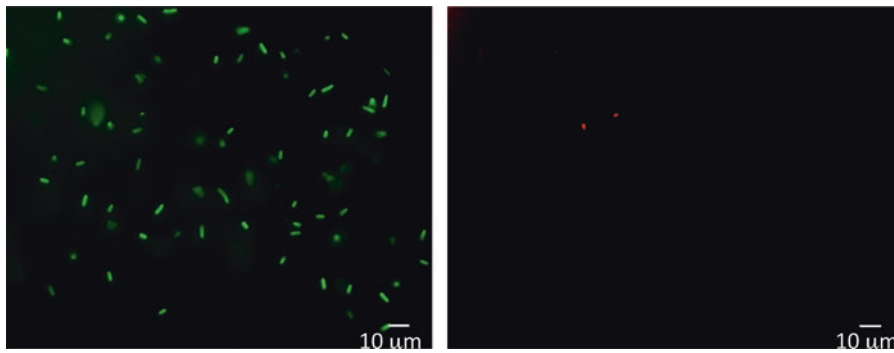


Fig. 1 Example of a Live/Dead stain for *Salmonella* grown in M9 minimal media (OD_{600}: 1). 1 mL of the *Salmonella* culture has been stained with a commercially available Live/Dead stain following the supplier's recommendations, whereby all *Salmonella* cells are stained in green (syto 9) and only the dead cells are stained in red (propidium iodide). The left image shows all the *Salmonella* cells present on the imaged area of the slide and the right image shows the dead cells within the same imaged area. Scale bar: 10 µm

4. NaOH–NaOCl solution (0.5 M–300 ppm): 10 g NaOH and 0.03% NaOCl in 500 mL sterile distilled water; heat to 50 °C.

2.4 Ultra-centrifugation (see Note 4 and Fig. 2)

1. Precooled Beckman ultracentrifuge with a fixed angle rotor (90 Ti) and precooled thickwall, polypropylene, (10mL, 16 × 76mm) ultracentrifugation tubes.
2. Optiprep diluent buffer, ODB: 50 mM Hepes, 1.19 g into 500 mL of distilled water, and 150 mM NaCl, 4.24 g into 500 mL of distilled water, pH 6.8. Filter sterilization should be done using a steritop filter (0.2 μm) and the ODB should be stored at 4 °C until use.

2.5 Iodixanol Gradient Separation Protocol (see Fig. 3)

1. Optiprep diluent buffer (ODB): *see* Subheading 2.4.
2. OptiPrep™ Density Gradient Medium: Iodixanol solution.
3. Precooled Beckman ultracentrifuge with swinging bucket rotor (SW40 Ti) and precooled thinwall, polypropylene, (14 mL, 14 × 95 mm) ultracentrifugation tubes.

2.6 SDS-PAGE and Immunoblotting

1. 12% Bis-Tris selfcasted or precasted gels.
2. SDS-PAGE running buffer, 1×: 50 mM MOPS, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.7: To prepare 500 mL of 20× MOPS SDS Running Buffer, dissolve 104.6 g of MOPS, 60.6 g of Tris Base, 10 g of SDS, and 3.72 g of EDTA in 400 mL ultrapure water; mix well and adjust the volume to

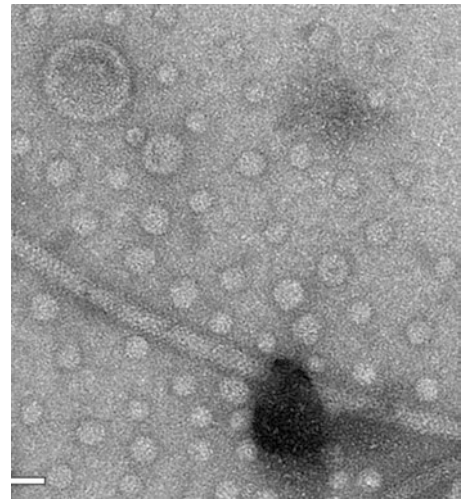


Fig. 2 Example of crude OMVs isolated from *Salmonella* grown in M9 media. An electron microscopy micrograph showing OMVs of sizes ranging from 20 to around 100 nm. The scale bar represents 50 nm. The long tubular structure in the middle of the micrograph represents flagellum, pilus, or fimbrium. The presence of these structures shows that crude OMVs are not pure and are still containing large protein complexes

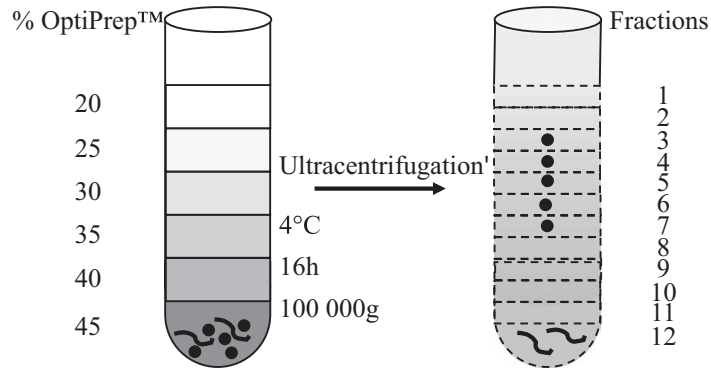


Fig. 3 Illustration of the bottom-up density gradient ultracentrifugation. Before ultracentrifugation, visible separations between the different density fractions can be observed (intact lines), whereas no clear visible layers are observable after ultracentrifugation (dashed lines). Crude OMVs are loaded in the bottom fraction and during ultracentrifugation vesicles migrate into the lower density fractions, while non-vesicular structures remain in the high density fraction. 1 mL fractions are taken from the top to the bottom and can then be further analyzed for the presence of vesicles

500 mL with ultrapure water; the buffer is stable for 6 months when stored at 4 °C. For electrophoresis, dilute this buffer to 1× with water. The pH of the 1× solution is 7.7. Do not use acid or base to adjust the pH.

3. Precision Plus Protein All Blue Standards (or any other protein ladder for polyacrylamide gels). For the lysis, denaturation and protease protection of the OMV samples: RIPA buffer, mini complete protease cocktail and Laemmli buffer with 10% β-mercaptoethanol.
4. Transfer Buffer: Mix 1 vol of 10× Tris/Glycine/SDS with 2 volume of methanol and 7 volume of ultrapure water. For 1 L: mix 100 mL of 10× Tris/Glycine/SDS, 200 mL methanol and 700 mL water. Store at 4 °C.
5. Blocking buffer: For 50 mL, mix 5 mL of 10× PBS and 1.5 g BSA and fill up to 50 mL with ultrapure water. Store at 4 °C.
6. Washing buffer: For 500 mL, mix 50 mL of 10× PBS and 2.5 mL of Tween-20. Store at 4 °C.
7. Polyvinylidene difluoride (PVDF) membrane and western blotting filter paper.
8. First and second antibody dilution buffer: For 10 mL, mix 1 mL 10× PBS, 0.3 g BSA, and 0.05 mL Tween-20. Prepare fresh for each SDS-PAGE.
9. First and secondary antibody. As an example, we use in the protocol as a first antibody a rabbit anti-*Salmonella typhi* outer

protein A and as a second antibody a goat anti-rabbit coupled to HRP (*see* **Note 5**).

2.7 Trichloroacetic Acid (TCA) Precipitation (see Note 6)

1. 20% of w/v TCA: for 100 mL: 20 g of TCA filled up to 100 mL with ultrapure water.
2. 80% ice-cold acetone: 40 mL of 100% acetone and 10 mL of sterile distilled water.

2.8 Small RNA Extraction (see Note 7)

1. Any commercially available kits allowing the extraction of small RNA.
2. 10 μg/mL lysozyme solution: add 1 mg of lysozyme to 1 mL of ultrapure water (1 mg/mL solution), mix, and dilute 100× to obtain a 10 μg/mL solution. The lysozyme solutions should be prepared just before the small RNA extraction and should not be stored for further use.
3. RNA clean & concentrator-5 kit (supplied with DNase I).
4. Agilent 2100 Bioanalyzer and Agilent Small RNA kit.

2.9 Small RNA Sequencing and Basic Analysis

1. To prepare a small RNA sequencing library different commercially available kits (NEB, Trueseq Illumina; or Perkin Elmer) can be used and the library can be sequenced using a single-end sequencing strategy on an Illumina Genome Analyzer, MiSeq, or NextSeq, where the maximum read length is set to 50 nt.
2. The data analysis can be performed on any computer having a command-line available (Linux/Mac) and Perl (http://www.perl.org) installed.
3. To process the data, the following software is needed: FastQC: for sequence data quality control [15], FASTX-Toolkit: for trimming and processing of sequence data [16], NovoAlign/NovoIndex: for mapping sequence data onto a reference genome [17, 18], HTSeq: Python tool to count RNA-Seq sequencing reads for genomic features [19], DESeq: R package for differential expression analysis [20, 21].

3 Methods

Carry out all experimental procedures at room temperature unless otherwise specified.

3.1 Bacterial Culturing

1. Inoculate *Salmonella* from a pure glycerol stock into 3 mL of M9 defined media (*see* **Note 1**) and grow this pre-culture for 6–8 h a day at 37 °C with a rotation of 200 rpm (*see* **Note 8**).
2. When the pre-culture reached an OD_{600} of around 1, inoculate 2 × 1.5 L of M9 media with the pre-culture by splitting it into

two. Incubate these cultures at 37 °C, overnight, with a rotation of 160 rpm. When the OD_{600} is between 0.5 and 1 (*see* **Note 9**), pursue with Subheading 3.2.

3.2 Live/Dead Staining and Crude OMV Isolation via Ultrafiltration and Ultracentrifugation

1. It is highly recommended to include a Live/Dead staining of the bacterial culture used for the isolation of outer membrane vesicle (OMV)-derived RNA (*see* **Notes 2** and **9**). For this staining, any commercially available kit can be used. If over 95% of the bacterial culture stain as viable cells (Fig. 1), it is recommended to pursue with the OMV-derived RNA isolation procedure.
2. Spin down the 3 L of bacterial culture for 30 min at 4700 × *g* at 4 °C. Take off the supernatant and place it into an autoclaved clean flask (the supernatant contains the OMVs). After having taken the desired amount of the bacterial culture for bacterial-derived RNA extraction (*see* **Note 10**), discard the remaining of the bacterial pellet as recommended by the biosafety rules appropriated for the bacterial type used. Filter the supernatant containing the OMVs through a 0.22 μm or 0.45 μm filter in order to eliminate the remaining bacteria that are still present in the supernatant. Use this filtered supernatant and concentrate it by using ultrafiltration.
3. Use a Quixstand Benchtop System (or any other tangential flow device) equipped with a 100 kDa hollow fiber membrane to concentrate the bacterial supernatant 50-fold following the supplier's recommendations (*see* **Note 11**). All supernatant that is not placed immediately into the tangential flow device should be kept at 4 °C until use. The concentrate (60 mL) should be filtered again through a 0.22 or 0.45 μm in order to be sure that no bacterial cells remain in the supernatant. Store the concentrate overnight at 4 °C or go on with ultracentrifugation.
4. Place the concentrate into precooled 10 mL ultracentrifuge tubes uniformly. Place the balanced tubes into a Beckman ultracentrifuge with a rotor 90 Ti (fixed angle rotor; the rotor and centrifuge should be precooled at 4 °C before use) and run with 150,000 × *g* for 2 h at 4 °C. After ultracentrifugation, the tubes should be turned upside down and the last drops should be absorbed on a tissue paper. The OMV pellets should be dissolved and pooled in 500 μL of ODB (depending on the amount of bacterial culture and the type of bacterial media used the crude OMV pellet will be visible or not; *see* **Notes 1** and **11**). The crude OMVs resuspended in buffer can be stored for 10 days at 4 °C. It is important to note that the pellet isolated at this step is highly enriched in OMVs but still contains larger protein complexes which may associate with RNA (*see* **Note 4** and Fig. 2) [10]. Therefore, we recommend to pro-

ceed with an iodixanol gradient separation in order to obtain pure OMVs. If, however, only a crude OMV isolation is desired, then the protocol can be stopped here and the pellet can be resuspended in 1× PBS or ODB. RNA extraction can be performed from crude OMVs if the absolute proof that a given RNA is associated with OMVs for follow-up experiments is not required. In that scenario, please skip Subheadings 3.3 and 3.4 and continue with Subheading 3.5.

3.3 Isolation of Pure OMVs via Iodixanol Gradient Separation (Fig. 3)

All the steps should be performed on ice.

1. Prepare the different OptiPrep™ density solutions by following the pipetting scheme shown in Table 1.
2. Follow a bottom-up approach to build up the density gradient within a 14 mL ultracentrifugation tube. For this, pipette the 2 mL 45% OptiPrep™-sample solution within the bottom of the tube and overlay carefully with the 2 mL 40% OptiPrep™-ODB solution. Continue to overlay with each of the 2 mL solutions prepared in **step 1** using the 35–20% OptiPrep™-ODB solution until obtaining a 12 mL density gradient where the different layers are visibly separated.
3. Place the ultracentrifuge tube containing the density gradient into a precooled bucket and place it into a precooled swinging bucket rotor (SW 40 TI). In the opposing bucket, place a balancing tube. Run the ultracentrifuge with 100,000 × *g*, for 16 h at 4 °C and importantly switch off the break of the ultracentrifuge so that the different density fractions that will form during ultracentrifugation will not be disturbed while the ultracentrifuge breaks. When running the ultracentrifuge at 100,000 × *g* without break it takes about 40 min to stop once the run has finished.

Table 1
Iodixanol gradient pipetting scheme for one tube

	OptiPrep™ (mL)	Sample (mL)
45% OptiPrep™-sample	1.5	0.5
	OptiPrep™ (mL)	ODB (mL)
40% OptiPrep™-ODB	1.3	0.7
35% OptiPrep™-ODB	1.2	0.8
30% OptiPrep™-ODB	1.0	1.0
25% OptiPrep™-ODB	0.8	1.2
20% OptiPrep™-ODB	0.7	1.3

4. Once the ultracentrifugation is finished, collect 1 mL fractions from top to the bottom of the gradient. Keep them separated and store on ice. From each fraction take 12.5 μL to check for the OMV presence using SDS-PAGE (*see* Subheading 3.4). Store the remaining 987.5 μL at 4 °C until use.

3.4 SDS-PAGE and Immunoblotting

To identify the iodixanol density fraction that contains pure OMVs, SDS-PAGE and immunoblotting can be used (*see* **Notes 5**, **12** and Fig. 4). In a pure OMV fraction, protein components of *Salmonella* flagella should not be detectable.

1. To prepare the SDS-PAGE sample, add to each 12.5 μL aliquot from the 12 individual iodixanol fractions, 5 μL of RIPA buffer and 2.8 μL of 7× complete mini protease. Incubate the samples on ice for 5 min. Then, add 7.5 μL of β-mercaptoethanol-containing 4× Laemmli buffer to each sample. Heat the samples for 5 min at 95 °C. Samples can now be stored at −20 °C or used for SDS-PAGE immediately.
2. Fill the gel running chamber with 1× SDS-PAGE running buffer and load the samples prepared in **step 1**, as well as an appropriated protein ladder on a 12% Bis-Tris gel. Run the SDS-PAGE with 200 V for 40 min. Stop migration, remove the gel, and proceed with western blotting.
3. Pre-soak PVDF membrane in methanol, then equilibrate in transfer buffer. Also, equilibrate 2 extra thick filter papers in transfer buffer. Prepare the western-blot sandwich as follows within a semidry blotter: filter paper–PVDF membrane–Bis-Tris gel–filter paper. Transfer with 25 V, 380 mA for an appropriated amount of time allowing the transfer of proteins with a molecular weight in the range of the protein of interest.

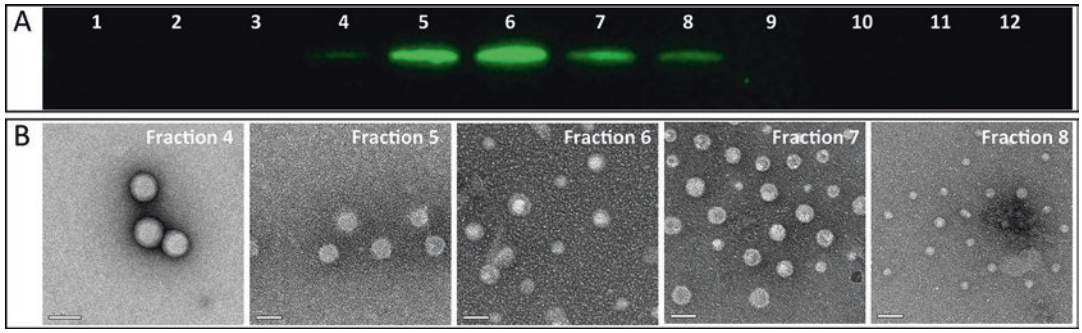


Fig. 4 Combination of electron microscopy and SDS-PAGE to identify iodixanol gradient fractions containing pure *Salmonella*-derived OMVs. (**a**) 10 μL out of the 1 mL fractions of all the individual 12 iodixanol gradient fractions had been analyzed for the presence of *Salmonella typhi* outer membrane protein A (OmpA) by SDS-PAGE. Fraction 1 being the lowest density fraction and fraction 12 being the highest density fraction. OmpA, an OMV-associated protein, is detected in fractions 4–8. (**b**) Electron micrographs showing that in fractions 4–8 spherical structures are present (OMVs have an average diameter of 20–200 nm and are bilayered). Scale bars are all 50 nm, except the scale bar in fraction 4 which represent 100 nm

4. Block the membrane with blocking buffer for 1 h. Wash the membrane 3 times for 5 min using the washing buffer. Incubate the membrane for 1 h with a first antibody directed against the protein of interest (we used antibody directed against *Salmonella typhi* outer membrane protein A: *see* **Notes 5** and **12**). Alternatively, the membrane can also be incubated with the primary antibody overnight at 4 °C. Wash the membrane 3 times for 5 min and incubate the membrane with an appropriated HRP-conjugated secondary antibody. Reveal and visualize the protein of interest by addition of peroxidase substrate and subsequent light detection by a CCD camera or photographic film. Identify the iodixanol fractions that contain OMVs and proceed with Subheading 3.5 with the 987.5 μL of these fractions set aside in Subheading 3.3 (**Note 13**).

3.5 Trichloroacetic Acid (TCA) Precipitation

TCA precipitation should be avoided if intact OMVs are required for subsequent analyses (*see* **Note 6**).

1. To each of the 987.5 μL OMV-containing fractions obtained in Subheading 3.3 and identified in Subheading 3.4, add 4 mL of ODB and 5 mL of 20% w/v TCA to achieve a final concentration of the TCA solution of 10% w/v. Vortex and incubate on ice for 30 min. Centrifuge for 30 min at 17136 × *g* at 4 °C. Discard the supernatant carefully and resuspend the pellet in 450 μL ice-cold 80% acetone. Centrifuge for 30 min at 10,000 rpm at 4 °C. Discard the supernatant carefully and dry the pellets (either let the pellets air-dry or use a vacuum pump). It is recommended to do for each OMV-containing fraction an individual TCA precipitation in order to dilute as much as possible the OptiPrep™ remaining in each fraction. When the pellets are dry, resuspend and pool them in 100 μL ODB. Keep the resuspended biomolecules at 4 °C and continue with Subheading 3.6.

3.6 Small RNA Extraction from Pure OMVs

1. Before isolation of RNA molecules from OMVs, the vesicles need to be lysed in order to release the RNA molecules. Add 10 μL of 10 μg/mL lysozyme solution, freshly prepared, to the precipitated OMV sample. Incubate for 5 min at room temperature.
2. Proceed with the extraction of RNA from the biomolecular fractions precipitated in Subheading 3.5 using either spin-column chromatography or phenol-chloroform extraction (*see* **Note 7**).
3. Isolated RNA can then further be concentrated using any commercial available clean-up and concentrator spin-columns that allow concentration of small molecules (<200 nt). Some concentrator columns can also be used to separate small and large RNA into individual fractions.

4. Finally, to digest any residual contaminant DNA within the isolated RNA sample, it is recommended to treat the extracted RNA with a DNAse.

3.7 Small RNA-Seq

1. All small RNA extracted out of OMVs (3 L *Salmonella* culture in M9 media) should be used to prepare a sequencing library (*see* **Note 1** and Fig. 5). Small RNA-Seq libraries can be prepared using commercially available kits and the library can be sequenced using a single-end sequencing strategy on an Illumina Genome Analyzer, MiSeq, or NextSeq, where the maximum read length is set to 50 nt. As OMV-derived small RNA samples are rather low yield samples, they are prone to contamination, and therefore it is highly recommended to include in the sequencing library a blank sample that has been processed exactly as the "real" samples, just that for this sample no bacteria had been inoculated in the growth media (*see* **Note 7**).
2. Optimal coverage and read depth are dependent on the experimental setup. A guide to determine recommended coverage and

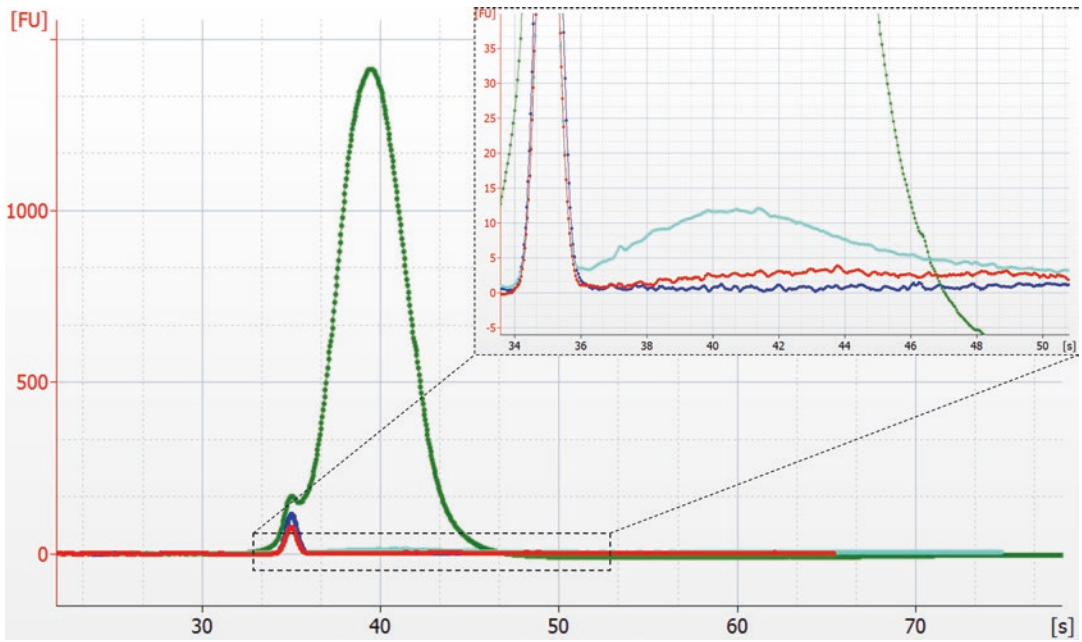


Fig. 5 Enriched broth media contains small RNA and the amount of small RNA released into OMVs varies drastically according to the growth environment of the OMV-secreting bacteria. Electropherograms of small RNA fractions isolated from density-gradient purified vesicles out of a rich broth media (light blue), of a *Salmonella* culture grown in a rich broth media (dark green) or of a *Salmonella* culture grown in a minimal chemically defined media (red). The electropherogram represented in dark blue shows the migration time and fluorescent unit of a water control sample. The OMV-secreting bacteria have been cultured to a similar OD_{600} within the chemically defined minimal media or the broth media and the supernatant volume used to isolate the small RNA has been identical for the three samples shown here

read depth is available in the following link: https://genohub.com/recommended-sequencing-coverage-by-application/.

3.8 Small RNA-Seq Analysis

1. For reference genome data acquisition, download bacterial genome sequences as well as known plasmid sequences in Fasta format and the genome annotation in Genbank (.gb), GTF (.gtf), or GFF3(.gff) format from the NCBI or EnsemblBacteria webpages. For *Salmonella*, the genome and plasmid sequences (in Fasta format) and annotations (in Genbank format) are available from NCBI, the genome accession AE006468 (https://www.ncbi.nlm.nih.gov/nuccore/AE006468) and the plasmid accession AE006471.1 (https://www.ncbi.nlm.nih.gov/nuccore/AE006471.1) or EnsemblBacteria (http://bacteria.ensembl.org/Salmonella_enterica_subsp_enterica_serovar_typhimurium_str_sl1344/Info/Index) (*see* **Note 14**).
2. For data preprocessing, FastQC [15], including visual inspection of QC reports, should be used to perform an initial quality check (QC) of Fastq (.fq) format files per lane. Fastq files should then be concatenated per condition and a second round of QC using FastQC, including visual inspection of the QC reports, should be performed and possible contamination/adapter sequences from the QC report (listed as enriched sequences) should be removed with the FastX-Toolkit [16]. The QC should be performed until FastQC detects no enriched adapter/contamination sequences anymore. Sequences are trimmed based on a Phred quality score threshold of 25 (nucleotides with lower quality are trimmed from the end of the sequence, and the Phred quality score of >25 has to apply for 100% of the bases in a sequence). Within small RNA *Salmonella* datasets, sequences with a length lower than 17 bp are prone for cross-mapping and therefore it is recommended to fix the minimum sequence length to 17 bp and to discard reads with a length < 17 bp. The remaining sequences are collapsed, so that identical sequences are represented by one sequence, but the information of the original read count is maintained in the read identifier.
3. For read mapping, a genome sequence index should be constructed with NovoIndex [17]. For example, for a *Salmonella* dataset, a *Salmonella* "genome" consists of the *Salmonella* genomic and plasmid sequence. Additional sequence indices should be built for genomes that are used for contamination filtration (e.g., human and/or yeast genomes). The mapping of trimmed sequence reads is performed with the following parameters using Novoalign [18] onto the *Salmonella* genome: the maximum acceptable mapping score of 60, which corresponds to 2 mismatches—no homopolymer filtering—a minimum length of 6 bp need to align—all mapping locations

(multiple) are reported, and the number of recorded mappings is limited to 51, which applies to the number of mappings with a score equal to the best alignment. When this limit is getting reached no further mappings are recorded and the search for this read is stopped.

4. Finally for annotation, successfully mapped reads are annotated for the genes of interest with a weighted score accounting for cross-mapping correction [22]. For example, annotation is done based on the *Salmonella* genome (AE006468) and plasmid (AE006471.1) NCBI Genbank, GTF or GFF3 files. The annotation for genetic features should comprise the assignment of different RNA types, which are "gene," "CDS," "tRNA," "rRNA," "ncRNA," "tmRNA," "unknown," and, if available, some literature annotations for RNAs that could be available for other bacterial strains of interest. "CDS" refers here to the coding region of protein-coding genes. Reads are counted per genetic feature using the HTSeq tool [19] for each sample using any annotation file as input. If desired, differential expression analysis can then performed based on the count tables derived from HTseq using the R package DESeq [20, 21]. A detailed description can be found here: http://bioconductor.org/packages/release/bioc/vignettes/DESeq/inst/doc/DESeq.pdf.

4 Notes

1. For the bacterial culture, alternatively to chemically defined minimal media, enriched bacterial culture media, such as Lysogeny Broth (LB), can also be used. In that case, 1 L of culture is enough to obtain bioanalyzer-detectable amounts of OMV-associated RNA. However, it is important to note that non-bacteria-derived small RNAs are present in any broth media (Fig. 5) and appropriate controls should be taken into account. For example, for functional assays, small RNA extracted from only LB should also be tested for their functionality in comparison to the bacterial OMV-derived small RNA. For small RNA sequencing not only OMV-derived small RNA should be sequenced but also LB-derived RNA.

   It is also important to note that the amount of OMVs produced by bacteria [2, 23] and the amount of small RNA released into OMVs by bacteria does vary drastically depending on their growth environment (Fig. 5).

2. Any other bacterial viability measurement can be used. The Live/Dead staining of the cultured bacteria is only to assure that the extracellular small RNA is indeed extracted from OMVs that were actively released by intact bacteria and is not due to a contamination by small RNA released by dead bacteria.

3. Very low yield samples are always prone to contamination, and therefore it becomes essential to clean thoroughly the hollow fiber membrane and tubing from the ultrafiltration device. To get rid of nucleic acid contaminants, it is recommended to use NaOH–NaOCl (0.5 M–300 ppm) as a cleaning agent. To perform a more effective cleaning, the cleaning solution should be preheated to 50 °C before use. Clean ultrafiltration device in a closed loop for 1 h and rinse several times with water before concentration of the sample.
4. After ultrafiltration and ultracentrifugation of the bacteria-derived supernatant the recovered pellet is enriched in OMVs, but it does not contain pure OMVs (Fig. 2). If no pure OMV sample is required, the protocol can be stopped here and the OMVs can be resuspended in PBS 1×.
5. If no *Salmonella* strain is used to isolate bacteria-derived OMVs, then an antibody directed against an outer membrane protein of that given species might be used instead. It is important that electron microscopy analysis or any other technique allowing the detection of vesicles within a given sample correlates with the results of SDS-PAGE (Fig. 4 and also *see* **Note 12**).
6. If intact OMVs are required, then the TCA precipitation should be avoided and pure OMVs can alternatively be gained by following the subsequent method. Each 1 mL fraction collected after the density gradient centrifugation does still contain OptiPrep™, and to remove the latter, samples need to be diluted in ODB. Therefore, each 1 mL fraction should be transferred into a 10 mL ultracentrifugation tube and 8 mL of ODB should be added. The samples should then be ultracentrifuged using a fixed angle rotor (90 Ti) (150,000 × *g* for 3 h at 4 °C). After ultracentrifugation the supernatant should be decanted carefully in order to not perturb the pellet (not visible if minimal growth media has been used), containing pure OMVs. The pellets from the OMV-containing fractions (identified in Subheading 3.4) should be pooled and taken up in either ODB or PBS 1×.
7. RNA can be isolated from OMVs using phenol-chloroform extraction or silica-column based approaches. Phenol-chloroform extractions are more economic, and in combination with hot phenol provide higher lysis efficiency and therefore higher yields [3]. On the other hand, phenolic residues can interfere with downstream analyses, such as preparation of sequencing libraries, and the reagents are harmful. Silica-columns for the isolation of small RNAs are commercially available from a range of manufacturers and are applicable for the isolation of bacterial small RNAs. As with all reagents, care

has to be taken to control that kits are free of environmental small RNAs, like rRNA fragments, which can skew analyses, in particular when the number of RNA molecules in the samples is low. In contrast to mRNAs, which are hardly ever stable enough to contaminate lab-ware or reagents, small RNA can form stable structures, which may be further stabilized on silica matrices. Columns can be controlled by adding a defined spike—in mix of small RNAs with appropriate lysis/binding buffers, followed by washing and elution as usual. Eluates can be sequenced or analyzed by qPCR. We have observed several hundred thousands of contaminant RNA molecules per μl eluate in one line of commercial columns. In our experience microRNA enrichment columns of Norgen Biotek and Machery Nagel were free of contaminating RNAs.

8. Always prepare a non-inoculated media control tube (to compare the turbidity: bacterial growth in the inoculated tube, to have a blank for OD_{600} measurement and to see if the media per se has not been contaminated with another microorganism during the inoculation procedure).
9. An OD_{600} of 0.5–0.6 usually corresponds to the log or exponential growth phase during bacterial culture. During the log phase, bacteria are generally the most reproductive, and beyond these OD_{600} values, bacterial population stabilize and then decline. We observed that *Salmonella* culture grown in M9 media with an OD_{600} of 0.7–1 allows to isolate more OMVs compared to a culture with a lower OD_{600} value. It is however important to note that if OMV-derived RNAs are isolated from the supernatant of a bacterial culture with an OD_{600} above the exponential phase, it is crucial to check the viability of the culture before extraction of the RNA in order to exclude dead bacteria-derived RNA contamination in the OMV-resulting RNA preparation.
10. If RNA extraction is required from the bacterial pellet, it is recommended to snap-freeze a part of the bacterial pellet. To avoid overloading of the extraction columns, only roughly 10^9 bacteria should be loaded (for *Salmonella* an OD_{600} of 0.58 correspond approximately to 123×10^6 bacteria/mL). An appropriated bacterial pellet should be snap-frozen and stored at −80 °C until extractions will be done.
11. The yield of OMVs in the extracellular environment is very low, and thus to obtain a visible pellet after ultracentrifugation, a total volume of 5–7 L of bacterial culture should be used. However, depending on the amount of OMV needed, the volume of bacterial culture can be reduced [24]. It is also important to note that the amount of OMVs produced by bacteria is dependent on the environmental conditions that the bacteria encounter (*see* **Note 1** and Fig. 5).

12. In this protocol we suggest the use of SDS-PAGE to examine the presence of OMVs within the individual iodixanol gradient fractions. However, it is important to note that SDS-PAGE can only be used if specific antibodies directed against OMV-associated proteins are available. Moreover, SDS-PAGE should be combined at least once with another technique allowing for the detection of vesicles in order to correlate the presence of an OMV-associated protein within a given fraction with the presence of intact vesicles within that same fraction. Techniques that could be used are electron microscopy, single-particle tracking, atom force microscopy, and high-resolution flow cytometry [25–27]. We combined our SDS-PAGE experiment with an electron microscopy analysis (Fig. 4). Therefore, we collected 1 mL fractions from top to bottom of the iodixanol gradient and added in each fraction 8 mL of ODB buffer in order to dilute the iodixanol. We ultracentrifuged the 12 individual fractions for 2 h at 4 °C at 150,000 × *g*. After ultracentrifugation we discarded the supernatant and resuspended the pellet in 100 μL of ODB. For SDS-PAGE analysis we used 10 μL of each fractions and the remaining 90 μL had been fixed with 4% paraformaldehyde and analyzed by electron microscopy (EM). We detected by SDS-PAGE the *Salmonella typhi* outer membrane protein A (OmpA) in fractions 4–8 and by EM we observed vesicular structures in the exact same fractions. Interestingly, we could not observe any vesicles by EM in the high density fractions (fractions 10–12) and identically we could not detect any OmpA by SDS-PAGE in these same fractions indicating that no or almost no vesicles-free OmpA is present in our sample. Interestingly, we were also unable to observe any long tubular structure representing a flagella or pilus or fimbria in the fractions 4–8, indicating the purity of the vesicle preparation.
13. From the SDS-PAGE analysis not only OMV-containing iodixanol fractions can be visualized but also iodixanol fractions that are OMV-free (usually fractions 10–12).
14. As OMV-derived RNA samples are low yield samples, it is important to control contamination sources, and thus sequenced reads should be mapped against possible contamination genomes and sequences, e.g., the human (current version of the human genome (hg38) from the UCSC webpage (https://genome.ucsc.edu/)) or yeast genomes (SGD database; http://www.yeastgenome.org/) or the PhiX phage sequence (NCBI accession number: NC_001422.1; https://www.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/phix-control-v3.html) that is often used as a spike—in control for Illumina sequencing runs. Unmapped "bacterial" reads from the analyzed dataset are

mapped optionally against the human hg38 and yeast genome to check for sample contamination using NovoAlign with the same parameter as used for the mapping onto the genome of interest.

Acknowledgments

This work was supported by a CORE programme grant (CORE/14/BM/8066232) to J.V.F and by a Proof-of-Concept grant (PoC/13/02) to P.W, all funded by the Luxembourg National Research Fund (FNR). We are grateful to Dr. Jean-François Ménétret from the Department of Structural Biology and Genomics Institute of Genetics and of Molecular and Cellular Biology (IGBMC; France) for the acquisition of the electron microscopy images presented in this protocol. We also thank Alton Etheridge and David Galas (Pacific Northwest Diabetes Research Institute, Seattle, Washington 98122) for RNA-Seq. Bioinformatics analyses presented in this book chapter were carried out in part using the HPC facilities of the University of Luxembourg (http://hpc.uni.lu).

References

1. Mashburn-Warren LM, Whiteley M (2006) Special delivery: vesicle trafficking in prokaryotes. Mol Microbiol 61:839–846. https://doi.org/10.1111/j.1365-2958.2006.05272.x
2. Bonnington KE, Kuehn MJ (2016) Outer membrane vesicle production facilitates LPS remodeling and outer membrane maintenance in salmonella during environmental transitions. MBio 7:e01532-16. https://doi.org/10.1128/MBIO.01532-16
3. Kadurugamuwa JL, Beveridge TJ (1995) Virulence factors are released from Pseudomonas aeruginosa in association with membrane vesicles during normal growth and exposure to gentamicin: a novel mechanism of enzyme secretion. J Bacteriol 177:3998–4008
4. Mashburn LM, Whiteley M (2005) Membrane vesicles traffic signals and facilitate group activities in a prokaryote. Nature 437:422–425. https://doi.org/10.1038/nature03925
5. Kuehn MJ, Kesty NC (2005) Bacterial outer membrane vesicles and the host-pathogen interaction. Genes Dev 19:2645–2655. https://doi.org/10.1101/gad.1299905
6. Shen Y, Giardino Torchia ML, Lawson GW et al (2012) Outer membrane vesicles of a human commensal mediate immune regulation and disease protection. Cell Host Microbe 12:509–520. https://doi.org/10.1016/j.chom.2012.08.004
7. Park K-S, Choi K-H, Kim Y-S et al (2010) Outer membrane vesicles derived from Escherichia coli induce systemic inflammatory response syndrome. PLoS One 5:e11334. https://doi.org/10.1371/journal.pone.0011334
8. Bomberger JM, Maceachran DP, B a C et al (2009) Long-distance delivery of bacterial virulence factors by Pseudomonas aeruginosa outer membrane vesicles. PLoS Pathog 5:e1000382. https://doi.org/10.1371/journal.ppat.1000382
9. Koeppen K, Hampton TH, Jarek M et al (2016) A novel mechanism of host-pathogen interaction through sRNA in bacterial outer membrane vesicles. PLoS Pathog 12:1–22. https://doi.org/10.1371/journal.ppat.1005672
10. Ghosal A, Upadhyaya BB, Fritz JV et al (2015) The extracellular RNA complement of *Escherichia coli*. Microbiologyopen 4:252–266. https://doi.org/10.1002/mbo3.235
11. Biller SJ, Schubotz F, Roggensack SE et al (2014) Bacterial vesicles in marine ecosystems. Science 343:183–186
12. Sjöström AE, Sandblad L, Uhlin BE, Wai SN (2015) Membrane vesicle-mediated release of

bacterial RNA. Sci Rep 5:15329. https://doi.org/10.1038/srep15329

13. Blenkiron C, Simonov D, Muthukaruppan A et al (2016) Uropathogenic Escherichia coli releases extracellular vesicles that are associated with RNA. PLoS One. https://doi.org/10.1371/journal.pone.0160440
14. Tseng T-T, Tyler BM, Setubal JC (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. BMC Microbiol 9(Suppl 1):S2. https://doi.org/10.1186/1471-2180-9-S1-S2
15. Bioinformatics B (2014) FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
16. Lab H FASTX-Toolkit: FASTQ/A short-reads pre-processing tools. http://hannonlab.cshl.edu/fastx_toolkit/
17. NovoCraft (2008) NovoIndex. http://www.novocraft.com/documentation/novoindex/
18. NovoCraft (2009) NovoAlign. http://www.novocraft.com/products/novoalign/
19. Anders S, Pyl PT, Huber W (2015) Genome analysis HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169. https://doi.org/10.1093/bioinformatics/btu638
20. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106. https://doi.org/10.1186/gb-2010-11-10-r106
21. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/PREACCEPT-8897612761307401
22. de Hoon MJ, Taft RJ, Hashimoto T, Kanamori-Katayama M, Kawaji H, Kawano M, Kishima M, Lassmann T, Faulkner GJ, Mattick JS, Daub CO, Carninci P, Kawai J, Suzuki HHY (2010) Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries. Genome Res 20:257–264
23. Klimentová J, Stulík J (2015) Methods of isolation and purification of outer membrane vesicles from gram-negative bacteria. Microbiol Res 170:1–9. https://doi.org/10.1016/j.micres.2014.09.006
24. Youn Kim O, Sil Hong B, Park K-S, et al (2013) Preparation of outer membrane vesicle from Escherichia coli. 3. http://www.bio-protocol.org/e995
25. Gardiner C, Di Vizio D, Sahoo S et al (2016) Techniques used for the isolation and characterization of extracellular vesicles: results of a worldwide survey. J Extracell Vesicles 5:32945. https://doi.org/10.3402/jev.v5.32945
26. Maas SLN, de Vrij J, van der Vlist EJ et al (2015) Possibilities and limitations of current technologies for quantification of biological extracellular vesicles and synthetic mimics. J Control Release 200:87–96. https://doi.org/10.1016/j.jconrel.2014.12.041
27. Lötvall J, Hill AF, Hochberg F et al (2014) Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. J Extracell Vesicles 3:26913. https://doi.org/10.3402/jev.v3.26913

Chapter 14

Absolute Regulatory Small Noncoding RNA Concentration and Decay Rates Measurements in *Escherichia coli*

Florent Busi, Véronique Arluison, and Philippe Régnier

Abstract

Regulation of RNA turnover is of utmost importance for controlling the concentration of transcripts and consequently cellular protein levels. Among the processes controlling RNA decay, small noncoding regulatory RNAs (sRNAs) have recently emerged as major new players. In this chapter, we describe and discuss protocols that can be used to measure sRNA concentration in vivo and to assess sRNA decay rates in Gram-negative bacteria. Precisely, we focus our analyses on the *Escherichia coli* Gram-negative bacterium as a model. The information described in this chapter provides a guideline to help develop a protocol in order to assess these important parameters and to identify RNA-processing enzymes involved in sRNA degradation processes.

Key words Small noncoding RNA, Post-transcriptional control, Nucleic acid quantification, RNA degradation

Abbreviations

ncRNA	Noncoding RNA
nt	Nucleotide
NTD/CTD	N-terminal/C-terminal domain
PAGE	PolyAcrylamide gel electrophoresis
PAP	Poly(A) polymerase
PNPase	Polynucleotide phosphorylase
RNAP	RNA polymerase
RNAse	Ribonuclease
sRNA	Small RNA
ss/ds	Single/double stranded
T_m	Melting temperature

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_14,

1 Introduction

The degradation of RNAs in bacteria has been analyzed in detail for years, and it is now well established that RNA stability depends on numerous factors, which are mainly the presence of secondary structures within the single-stranded RNA or mRNA translation efficiency (for a recent review, *see* [1]). The whole RNA degradation process involves the organized and successive actions of a series of ribonucleases, which may cleave within the RNA (*i.e.*, endonuclease) or remove one nucleotide from the extremities of the nucleic acid (*i.e.*, exonuclease) (Fig. 1). As indicated below, these nucleases sometimes need the assistance of cofactors in order to modify the ends of the RNA or to unwind double-stranded sequences (for a review *see* [1]) (Fig. 1). In Gram-negative bacteria, more precisely in its best-characterized model *E. coli*, the first major step that triggers decay is usually the cleavage of mRNA within single-stranded (ss) regions by the endonuclease RNAse E. Although alternative enzymes such as the RNAse III endonuclease can also initiate mRNA decay by cleaving in double-stranded RNA regions, this RNAse has a less prominent role in mRNA decay. Usually, RNAse E is found associated with other proteins in a cellular complex called the degradosome [2]. In *E. coli* this complex contains the exoribonuclease polynucleotide phosphorylase (PNPase), an RNA-helicase (RhlB), and the enolase glycolytic enzyme in addition to RNase E (Fig. 2). Following initial RNA cleavage by the endonuclease, the RNA fragments can thus be degraded, in a processive way, by the 3′–5′ exonucleases RNAse II, RNAse R, or PNPase (the latest being a part of the degradosome complex). When the 3′ end of RNA fragments are protected by stable secondary structures, poly(A) extensions synthesized by poly(A) polymerase (PAP I) or melting of annealed nucleotides by the RhlB RNA helicase (also a part of degradosome) can also facilitate the degradation by exonucleases. Finally, the essential oligoribonuclease Orn completes the degradation of the short oligoribonucleotides into individual nucleotides available for fresh rounds of RNA synthesis. While each step of RNA decay is important, the first cleavage by RNAse E is considered to be kinetically limiting. Nevertheless, RNAse E activity is allosterically activated by the presence of 5′ monophosphate ends, which may be generated either by an initial endonuclease cleavage, or by a protein cofactor, the RppH pyrophosphatase. RppH is able to remove pyrophosphates from the triphosphate 5′-end of the RNA. Note that this RNA degradation machinery is organized in a structured network localized in the vicinity of the inner bacterial membrane, an organization that probably helps to temporally and spatially coordinate the different activities [3].

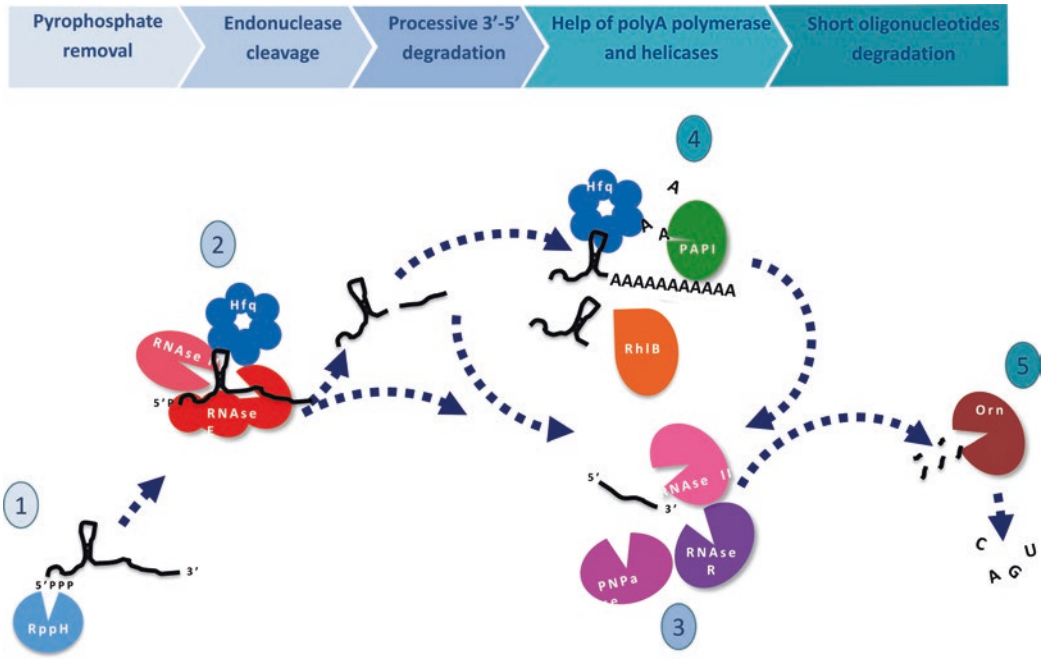


Fig. 1 General scheme for *E. coli* RNA decay. RNA pyrophosphohydrolase RppH trims off the 5′-pyrophosphate of mRNA. RNase E preferentially cleaves internally in 5′-monophosphate RNA substrates. RNase III has a limited role in mRNA processing, but plays a major role in sRNA decay. PNPase, RNase II, and RNase R exonucleases carry out the 3′ → 5′ RNA degradation. RhlB helicase facilitates the degradation by exoribonucleases. Similarly, addition of a poly(A) tail by PAP I facilitates the action of exonucleases. The essential exoribonuclease Orn is responsible for completion of degradation of the short RNA fragments left by the other RNAses. Hfq modulates translation and consequently RNA stability. It also interacts with RNAse E, facilitates the action of PAP I, and binds directly with sRNA

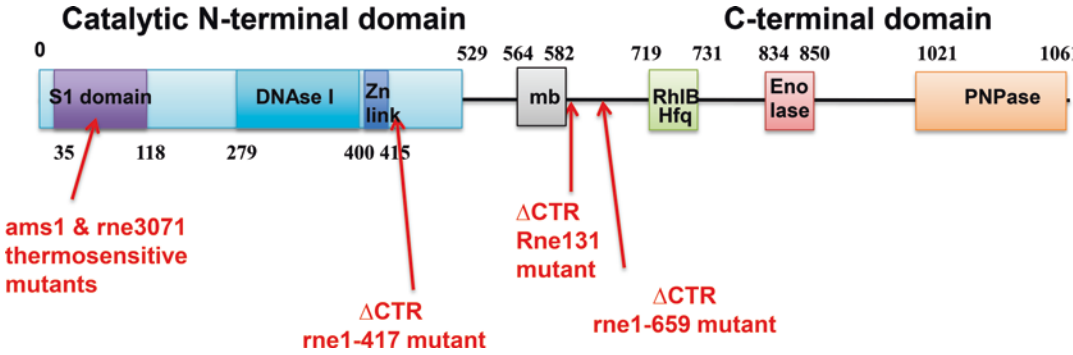


Fig. 2 Structure of RNAse E and mutations used for decay analyses. RNAse E is a 1061 amino acid protein essential for *E. coli* cells. It is made up of an N-terminal catalytic domain and a C-terminal domain, the later containing regions that nucleate the formation of the degradosome complex as well as the inner-membrane interacting region. Two thermosensitive mutations (*rne1*, originally named *ams1*, and *rne1071*) are found in the first part of the catalytic domain. Different *rne* alleles coding for truncated RNAse E without CTR can also be used (*see* Table 1)

Table 1
Strains that can be used for the analysis

Strain name and genotype	Phenotype	Source
BL 322: *thi-1 argH-1 sup44*	[Rnc$^+$]	W. Studier
BL 321: *thi-1 argH-1 supE44 rnc105*	[Rnc$^-$]	W. Studier
IBPC5321: *thi-1, argG6, argG3, his-4, mtl-1, xyl-5, tsx-29, rpsL, ΔlacX74*	[Rne$^+$]	[40]
IBPC642: IBPC5321 *ams1^{ts} zcf229::Tn10*	[Rne^{ts}]	[41]
N3433: *HfrH, lacZ43, λ$^-$, relA1, spoT1, thi-1*	[Rne$^+$]	D.Apirion
N3431: N3433 *rne3071*	[Rne^{ts}]	D.Apirion
IBPC637: N3431, *rnc105 nadB51::Tn10*	[Rne^{ts}, Rnc$^-$],	[33]
IBPC929: N3433 *hfq1*::Ω(kan^R, *BclI*)	[Hfq$^-$]	[42]
IBPC953: N3433 *hfq*Δ22-294 *cycA30::Tn10* (tet^R)	[Hfq$^-$]	[42]
MG1693: LAM-, *thyA715, rph-1, deoC1*	WT control	CGSC
SK5665: MG1693, *rne1*	[Rne^{ts}]	[43]
SK5691: MG1693, *pnp7*	[Pnp$^-$]	[43]
SK5704: MG1693, *rne1, pnp7, rnb500*	[Rne^{ts}, Pnp$^-$, Rnb^{ts}]	[43]
SK5671: MG1693, *rne1, pnp7*	[Rne^{ts}, Pnp$^-$]	[43]
SK5715: MG1693, *rne1, rnb500*	[Rne^{ts}, Rnb^{ts}]	[43]
SK7988: MG1693 *pcnB*::KmR	[PAP$^-$]	[44]
IBPC690: MG1693 *pcnB$^-$*	[PAP$^-$]	[44]
IBPC667: SK5665 *pcnB$^-$*	[PAP$^-$, Rne^{ts}]	[44]
IBPC673: SK5704 *pcnB$^-$*	[PAP$^-$, Rne^{ts}, Pnp$^-$, Rnb^{ts}]	[44]
SK5006: MG1693, pDK39(*rnb500*)	WT control	[34]
SK5003: MG1693, *pnp7, rnb500*, pDK39(*rnb500*)	[Pnp$^-$, Rnb^{ts}]	[34]
IBPC6790: *thi-1, argE3, his-4, lacX74, mtl-1, xyl-5, tsx-29, supE44, rpsL, zce-726::Tn10*	[Rne$^+$]	[45]
IBPC6981: IBPC6790 *rne131*	[RneΔCTD(584–1061)]	[45]
SK9714: MG1693 *rneΔ1018::bla, recA56, srlΔ300::Tn10 Tc^R/pSBK1 [rne+, CmR]*	[Rne$^+$]	S. Kuschner
SK9714: MG1693 *rneΔ1018::bla, recA56, srlΔ300::Tn10 Tc^R/pRne-SG4*	[Rne R169Q 47]	[32]
SK9714: MG1693 *rneΔ1018::bla, recA56, srlΔ300::Tn10 Tc^R/pRne-SG5*	[Rne T170A 89]	[32]

(continued)

Table 1
(continued)

Strain name and genotype	Phenotype	Source
SK9714: MG1693 *rneΔ1018::bla, recA56, srlΔ300::Tn10 Tc^R/pRne-SG31*	[Rne T170A + Δ530–1061]	[32]
IT1568: W3110mlc	[Hfq$^+$]	[29]
TM589: W3110mlc Δ*hfq*	[Hfq$^-$]	[29]
EM1055: MG1655 Δ*lac X174*	[Hfq$^+$]	[46]
EM1265: EM1055 *hfq-1::Ω(kan;Bcl1)*	[Hfq$^-$]	[14]
MC1000: F-, *Δ(araA-leu)7697, [araD139]B/r, Δ(codB-lacI)3, galK16, galE15(GalS), λ-, e14-, relA1, rpsL150(strR), spoT1, mcrB1*	WT control	CGSG
AT8: MC1000 *rne1–417-cat*	[RneΔCTD(417-1061)]	[3]
AT28: MC1000 *rne1–417-kan*	[RneΔCTD(417-1061)]	A. Taghbalout
AT27: MC1000 *rne1–659-kan*	[RneΔCTD(659-1061)]	[3]
AT247: MC1000 Δ*pcnb-kan*	[PAP$^-$]	[3]
AT156: MC1000 Δ*rnc-kan*	[RNAse III$^-$]	A. Taghbalout
FL1: MC1000 Δ*rnb-kan*	[RNAse II$^-$]	[47]
AT165: MC1000 Δ*hfq*	[Hfq$^-$]	[3]
AT160: MC1000 Δ*hfq-kan*	[Hfq$^-$]	A. Taghbalout

Recently, new players in RNA decay appeared in this general scheme. Indeed, small noncoding regulatory RNAs (sRNA) have been proven to induce mRNA degradation in Gram-negative bacteria, although a few sRNAs may act as positive regulators [4]. This destabilizing effect occurs mainly through binding of a specific sRNA to its mRNA target, with the concomitant action of the RNA-processing machinery. Indeed, RNAse III endonuclease plays an important role in the degradation pathway of mRNA targets coupled to cognate sRNAs. Furthermore, sRNA often binds in the vicinity of the ribosome-binding site (RBS), which causes a stop in mRNA translation, which also promotes degradation of the silenced transcript (for a review *see* [5]). As the Hfq protein plays an important role in the mediation of sRNA/mRNA interactions, this protein also influences RNA decay [6]. On one hand, Hfq mediated sRNA-mRNA annealing triggers mRNA decay indirectly while, on the other hand, a small RNA that is not associated to Hfq is rapidly degraded, although the latter mechanism is not fully understood. Furthermore, Hfq is also found to interact with other proteins involved in mRNA decay such as Poly(A) polymerase and RNAse E [7, 8], which may also influence its effect on RNA decay in vivo.

In this chapter, we intend to describe and discuss in detail the different protocols and bacterial strains that can be used to measure sRNA decay and absolute concentration in *E. coli* bacterial cell. Two main methods will be discussed.

2 Materials

2.1 Reagents

To ensure the reproducibility of the experiments described below, we recommend using:

1. Phusion® High-Fidelity DNA Polymerase (from New England Biolabs or Thermo Scientific) to produce DNA template for transcription. Alternatively iProof DNA polymerase (from Biorad) can be used.
2. RNeasy column from Qiagen for RNA purification.
3. SYBR green II is from Thermofisher Scientific.

2.2 Mutant Strains Affected in sRNA Decay

A list of the strains that can be used for the analyses of sRNA decay are summarized in Table 1. These strains are described in detail in Subheading 3.1.1.

2.3 Buffers and Solutions

All solutions have to be RNAse-free.

1. TE buffer: 10 mM Tris–HCl, 1 mM EDTA pH 8.0.
2. TAE buffer: 40 mM Tris-Acetate, 1 mM EDTA pH 8.0.
3. TBE buffer: 40 mM Tris-Borate, 1 mM EDTA pH 8.0.
4. T7 RNAP transcription buffer (supplied with T7 polymerase).
5. Denaturing gel-loading buffer: 60% formamide, 12 mM EDTA, 0.03% bromophenol blue, 0.03% xylene cyanol.
6. RNA extraction Solution 1: 10 mM Tris–HCl pH 7.5, 10 mM KCl, 5 mM $MgCl_2$.
7. RNA extraction Solution 2: 20 mM Tris–HCl pH 8.0, 200 mM NaCl, 40 mM EDTA, 1% SDS.
8. Roti-Hybri-Quick ready-to-use solution for DNA and RNA hybridization buffer (Carl Roth, this buffer is based on the hybridization solution published by Church & Gilbert [9]).
9. Saline-sodium citrate SSC buffer 20× stock: 3 M sodium chloride, 300 mM trisodium citrate adjusted to pH 7.0 with HCl.
10. SYBR green qPCR ready-to-use Master Mix contains all components for qPCR, except primers and templates, and can be obtained from manufacturers such as Roche Diagnostics, Bioline, Thermo Scientific, Agilent, Biorad, and Promega. Minor technical advantages can be observed with the mix from

some manufacturers; this is especially due to components used in the kit: in particular, the polymerase (processivity, proofreading activity or not and most importantly sensitivity to inhibitors) as well as the dye and its physicochemical properties (density of intercalation within the double stranded DNA, fluorescence quantum yield, etc.). But overall one should be able to achieve quantification experiments with any commercial kit regardless of the manufacturer.

11. Qiagen RLT and RPE buffers (composition is confidential).

3 Methods

3.1 Cell Cultures and RNA Extraction

Contribution of RNAses and cofactors to the metabolism of a sRNA and its target mRNA(s) can be determined by looking whether the fate of a particular RNA is affected in mutant strains failing to produce the active proteins. This is usually achieved by comparing the concentration and the time course of decay of the preexisting RNA after RNA polymerase inhibition by rifampicin in strains containing and lacking RNAse(s) or cofactors (Table 1).

3.1.1 Description of Mutant Strains of Particular Interest for Determining the Degradation Pathways of sRNA

Available data indicate that RNAse E and PNPase are prominent actors of processing and degradation of sRNA whose metabolism has been investigated [10–13]. In addition, RNAse E is recruited to get rid of mRNAs silenced by sRNA [14, 5] and RNAse III, which specifically targets double-stranded RNAs, often catalyzes the cuts which trigger decay of sRNA-mRNA duplexes [15]. Beside these main actors, one must also mention RNAse II [11, 16], the Hfq RNA chaperone and the PAPI poly(A) polymerase [17–19, 10], which favor sRNA-mRNA annealing and facilitate exonucleolytic decay of RNAs, respectively.

1. *RNAse E strains.* RNAse E is a polypeptide of 1061 amino acids that is essential for cell viability, made up of a N-terminal catalytic domain (NTD) and a C-terminal domain (CTD), containing polypeptide segments which nucleate formation of complexes with the other degradosome components and the inner membrane [20, 2, 21]. Two thermosensitive mutations referred to as *rne1*, originally named *ams1* [22, 23], and *rne1071* are commonly used (Table 1). Both harbor point mutations mapping at close locations in the NTD catalytic domain (Fig. 2), which cause inactivation of RNA processing at nonpermissive high temperature [24]. The role of the RNAse E CTD can be investigated by means of *rne* alleles encoding truncated polypeptides lacking part or all of this multidomain half of the protein (Fig. 2). One of these alleles, *rne131* [25], generated by a point mutation, gives rise to a 616 amino acid polypeptide terminated by 32 out-of-frame

residues, thus omitting most of the CTD (Fig. 2). The *rne131* strain containing thermosensitive RNAse E and the *rne*$^{+}$ control cells, grown at permissive temperature, must both be shifted to 42 °C at the time of rifampicin addition in order to detect whether RNase E affects RNA stability. Many other strains harboring deletions of one or several domains of the CTH have also been used to investigate the role of the degradosome partners and membrane binding in RNA decay [20, 21, 26–30]. Initially characterized through its implication in processing of rRNA precursors, RNAse E is now known to be involved in maturation and decay of all kinds of RNA. This plurality of substrate recognition capability has been attributed to the presence of a sensor pocket of the NTD able to interact with the 5′ monophosphorylated terminus and structural motifs of the CTD, which bind internal parts of the RNA [31]. Interestingly, strains harboring *rne* alleles affected at one or both of these two sites (Table 1) allow investigation into whether RNA decay follows a 5′ end-dependent or an internally initiated mode of decay [32].

2. *RNAse III strains.* RNAse III is another major player of sRNA mediated regulations and sRNA metabolism. Although the RNAse III deficient cells harboring the *rnc105* allele grow slightly more slowly than the rnc^{+} strain, comparison of RNAs in the rnc105 and *rnc* + strains grown at 37 °C allows to determine whether RNAse III is involved in the metabolism of a RNA. Considering that association of *rne1* and *rnc105* is not viable in all genetic backgrounds, it is necessary to investigate the effects of inactivation of both RNAse E and RNAse III on the time course of RNA decay in a *rne*1071-*rnc*105 double mutant grown at permissive temperature, whose thermosensitive RNAse E has been inactivated at 42 °C at the time of rifampicin addition [33].
3. *PNPase strains.* Cells harboring the *pnp7* allele are deficient for PNPase. This allele can be associated with the *rne1*, *rne1071*, and *rnc105* alleles to investigate RNA metabolism in multiple mutants lacking several ribonucleases. Notably, PNPase deficiency is not compatible with the lack of another ribonuclease, RNAse II, which also degrades the 3′ end of RNAs. Nonviability of *pnp-rnb* (coding RNAse II) double mutants can be overcome by using *pnp200*^{ts} or *rnb500*^{ts} alleles, producing thermosensitive PNPase and RNAse II [34–36].
4. *PAPI strains.* It is also worth mentioning that strains deficient for Poly(A) polymerase (*pcnB*) and the RNA chaperone Hfq (*hfq*) have been used to study the role of these key players of sRNA activity in decay rate of RNAs [17, 37, 5].

A list of the strains that can be used for experiments are summarized in Table 1.

3.1.2 Cell Cultures

Rifampicin is added in growing bacteria reaching the middle of the logarithmic phase. Cultures are then continued and total RNA is extracted from samples withdrawn just before and at several times after inhibition of RNA synthesis (including $t = 0$). Rates of RNA disappearance of specific RNA in different strains are then quantified on Northern blots or with RT-PCR.

Bacteria are usually grown at 37 °C except in the case of thermosensitive strains that are cultivated at 30 °C before being shifted to the nonpermissive temperature, which inactivates the thermosensitive ribonucleases. In the case of RNAse E, the thermosensitive *rne-1*^{ts} or *rne-3071*^{ts} (Table 1) and the *rne*$^{+}$ cells are grown at the permissive temperature (30 °C) and rifampicin is added at the time of the temperature shift to 42 °C, which inactivates the thermosensitive RNAse E. Cells are then cultivated longer at 42 °C and total RNA prepared and analyzed as a function of time as described above. The same protocol is also applied to investigate the role of other thermosensitive ribonucleases such as the *rnb500*^{ts} encoded RNAse II in strains lacking PNPase.

Because cells harboring *rne-1*^{ts} allele still grow at 37 °C due to residual activity of the thermosensitive *Rne1* polypeptide, it is possible to examine RNAse E implication by comparing the amount and the fate of a RNA in *rne-1*^{ts} and *rne*$^{+}$ cells grown at 37 °C [28].

3.1.3 RNA Extraction

1. When the culture of bacterial strains grown in appropriate conditions reaches the appropriate OD_{600}, 10 mL of the culture is added to an equivalent volume of pre-chilled (−20 °C) absolute ethanol and kept on ice before **step 2**.
2. The different samples are then centrifuged at 12,000 × *g* for 5 min at 4 °C.
3. The bacterial pellet is re-suspended in 1.5 mL pre-chilled solution 1 (4 °C), and 1.5 mL of solution 2 (pre-warmed to 95 °C) is added immediately.
4. The mixture is incubated for 2 min at 95 °C. The lysate can be frozen after this step if necessary.

3.1.4 RNA Purification

1. 3 mL of Acid Phenol:Chloroform:Isoamyl alcohol (25:24:1, pH 4.5) is added to the bacterial lysate.
2. After vigorously mixing with a vortex, the mixture is incubated at 65 °C for 15 min under agitation.
3. The tube is centrifuged at 10,000 × *g* for 15 min and the aqueous phase collected in another tube. This allows for the removal of DNA which partitions into the organic phase.
4. This step has to be repeated a second time to maximize DNA removal.
5. 3 mL of chloroform is added and the sample mixed vigorously with a vortex.

6. The mixture is centrifuged at 10,000 × *g* for 15 min.
7. The aqueous phase is collected and RNA precipitated with 0.1 volume of NaCl 5 M and 2.5 volumes of absolute EtOH.
8. The pellet is washed with 70% EtOH and dried.
9. The RNA is dissolved in 100 µL H_2O.
10. Finally, RNAs are treated with 1 unit of RQ1 DNAse per µg of DNA during 60 min at 37 °C and RNAs are further purified with RNeasy column (Qiagen) as follows, in order to purify RNA from RQ1 DNAse.

*3.1.5 RNeasy Purification Adapted for sRNA (see **Note 1**)*

1. Up to 10^9 bacteria are disrupted and homogenized in 350 µL of RLT lysis buffer containing guanidine thiocyanate.
2. 3.5 volumes (1225 µL) of 100% ethanol is added. Vortex to mix and proceed immediately to **step 3** (no centrifuge).
3. Pipet 700 µL of the sample, including precipitate, into an RNeasy spin column introduced in a 2 mL collection tube. Close the lid and centrifuge for 15 s at 8000 × *g*. Discard the flow-through. Repeat **step 3** until the whole sample has been pipetted into the spin column.
4. Place the spin column into a new tube. Add 500 µL of RPE Buffer. Close the lid, and centrifuge for 15 s at 8000 × *g* to wash the membrane. Buffer RPE is supplied as a concentrate. Be sure to add ethanol to Buffer RPE before use.
5. Again add 500 µL RPE Buffer into the column, close the lid, and centrifuge for 15 s at 8000 × *g*.
6. Place the spin column into a new 2 mL collection tube. Centrifuge at 12,000 × *g* for 1 min.
7. Place the spin column into a 1.5 mL collection tube. Add 30–50 µL RNAse-free water to the spin column. Close the lid and centrifuge for 1 min at 8000 × *g* to elute the sRNA (and total RNA if present).
8. Repeat **step 7** with a second volume of RNAse-free water. Elute into the same collection tube.

RNA concentrations are measured by UV spectrophotometry at 260 nm. The isolated RNA can be quantified using a NanoDrop spectrophotometer (NanoDrop Technologies, USA), allowing measurements of 1 µL concentrated samples without the need for dilution.

3.2 Quantitative RT-PCR Analysis of Noncoding sRNAs

RNA extraction is carried out as in Subheading 3.1. In order to correct absolute quantifications with RNA extraction yields, 3 pmol of a transcript from a *kan*^{R} gene fragment (103 nt) produced by T7 in vitro transcription is added to the bacterial lysate prior to RNA extraction.

3.2.1 In Vitro Synthesis of kan^R Fragment Transcript

1. To produce the *kan*^{R} fragment PCR fragment, the T7 promoter (underlined) is added to *kan*^{R} fragment sequence in the forward primer: 5′ TAATACGACTCACTATAGGG-GCGAG TGATTTTGATGACGA 3′

   Reverse primer is: 5′ CATGAGTGACGACTGAATCC 3′

   kan^R fragment sequence is amplified from a plasmid containing *kan*^{R} gene, such as pET28a. It is usually not necessary to add a T7 terminator to the reverse primer if the transcript corresponds to the end of the PCR fragment. We recommended to check the quality of the PCR product on a gel and to remove abortive products if necessary.
2. T7 transcription is performed in 20 µL of transcription buffer (supplied with T7 RNAP) with 1 pmol of the PCR amplified DNA matrix, 2 mM NTP, 12 mM $MgCl_2$, and 100 U of T7 RNAP. The reaction mix is incubated for 3–6 h at 37 °C.
3. DNA is then digested by RNAse-free RQ1 DNAse. The digestion is performed with 1 unit of enzyme per µg of DNA during 15 min at 37 °C.
4. The RNA is extracted by Phenol:Chloroform:Isoamyl alcohol (25:24:1) extraction and ethanol precipitation. The transcript is then re-suspended in 10 µL of TE buffer.

3.2.2 Checking the Result of In Vitro Transcription

1. In order to check the efficiency of transcription, the noncoding RNA transcript is loaded on a denaturing PAGE. This can be achieved with a 15% acrylamide/bis 19:1 gel containing 8 M urea in TBE buffer.
2. The sample is pre-heated at 80 °C in denaturing gel-loading buffer. The electrophoresis is run for 1 h with TBE buffer at ~50 °C (20 V/cm).
3. After migration, the RNA is stained with SYBR green II or Ethidium Bromide (0.5 µg/mL). The visualization is achieved with a transilluminator (254 or 300 nm). Two results can be observed:
   (a) If there are no abortive products, the RNA can be purified directly using the RNeasy Mini kit as described in Subheading 3.1.5.
   (b) If there are some abortive products, the transcript needs to be purified from the PAGE.
4. For this goal, the band corresponding to the RNA has to be cut from the gel.
5. The piece of gel is ground (crushed) and eluted overnight at 37 °C under agitation with 2 volumes of elution buffer.
6. The supernatant is recovered and SDS is removed by phenol-chloroform/chloroform extraction followed by ethanol precipitation.

7. The transcript is re-suspended in 10 µL of TE buffer. RNA concentrations are measured by UV spectrophotometry at 260 nm.

3.2.3 Reverse Transcription

1. 20 U/µL M-MLV reverse transcriptase is used for cDNA synthesis from 1 µg total RNA in 25 µL. This step is performed at 37 °C for 60 min (*see* **Note 2).**
2. Heat at 85 °C for 5 min to inactivate reverse transcriptase.

Because, most of the time, the short length of the sRNA does not permit different oligonucleotides to be used to prime the reverse transcription reaction and the qPCR primers, the primers used for reverse transcription are the same as the reverse primers used for qPCR (see below). Negative controls without reverse transcriptase are assembled in the same manner for each condition to evaluate putative genomic DNA contamination.

3.2.4 RT-qPCR Analyses

The average copy number of sRNA in bacterial cells is measured with a SYBR Green based RT-qPCR method. The quantification is performed with the amplification of the three following sequences: the sRNA of interest and the *kan*^{R} and *rrsB* transcripts. *rrsB* (coding 16S rRNA) is used as a housekeeping gene control to normalize RNA quantities between samples (*see* **Note 3),** while *kan*^{R} is used to take into account the efficiency of RNA extraction.

The quantitation using a hydrolysis probe-based qPCR (such as TaqMan and other probe-based methods) is an alternative. However, this strategy suffers from constraints inherent to sRNA properties. Indeed, the assay primers and probe require a careful design and, in general, fulfilling these objectives both in primers and in the probe is therefore more probable to be unsuccessful than for Sybr Green I assays for which only two regular PCR primers are required. In addition, one should consider the cost of such a strategy as a specific probe must be designed with fluorophore(s) and quencher(s) coupled to oligonucleotides for each single assay. Alternatively, having such an assay available may offer several advantages. Most often the apparent PCR efficiencies calculated from serial dilutions are better than for SYBR green I assays, especially because of shorter PCR product amplification, and more importantly the sensitivity of these assays are much better and allow the detection of fewer copies of starting material in single reaction.

1. The primers are designed using the primer 3 program (http://primer3.ut.ee/) in the following manner:

   One main issue in a qPCR experiment is insufficient qPCR assay efficiency, which is mainly affected by the capability of primers to prime nonspecific polymerizations competing with amplification of the specific product synthesis. Two kinds of unspecific amplifications can occur and affect drastically the

PCR efficiency: (1) primer-dimer amplification because of priming on the high concentration oligonucleotides (used as primers) within the reaction mixture and (2) unspecific priming of oligonucleotides on cDNA harboring short nucleotide sequence(s) highly similar to the target sequence, or when the target gene belongs to a conserved multigene family with several related isoforms. This latter case less frequently occurs in sRNA qPCR assays. These issues can be avoided with a careful design of qPCR oligonucleotides.

Minimization of primer-dimer formation is achieved in the online software when requiring the lowest intra- and inter-primer complementarity ("(TH) Max (both "Self" and "Pair") Complementarity" and "(TH) Max 3′ (both "Self" and "Pair") Complementarity" parameters and should be as small as possible, close to '0' for each primer and for the combination of both (in the old interface), in addition to minimization of "TH Max Primer Hairpin" parameter). The T_m difference between both primers ("Maximum T_m Difference") should be as low as possible and the GC content ("Primer GC%") ideally balanced. Repetition of the same nucleotide ("Max Poly-X") should also be minimized. Two general strategies can be used in parallel with primer 3: (1) either let the software pick the primers according to the quality parameters that have been filled and relax these constraints until obtaining a primer pair, or (2) force the sequence of the primers in the dedicated text boxes to get the primers analyzed. It is then possible to refine the primer sequence by lengthening or shortening this sequence at the 5′-end if the complementarity parameters, in addition to the T_m of the primers, are satisfactory. The latter strategy is best adapted for the design of qPCR primers for the smallest sRNAs as it is possible to tailor the primer at the suited position. Indeed, according to the predicted secondary structures of both sRNAs, one should avoid highly stable secondary structure regions (RNA stems with high local GC content) as well as the formation of intra-primer hairpins. It is often a good idea to check the specificity of the designed primer oligonucleotides by BLASTing them on the same species nucleotide database.

2. RT-qPCR reactions in 10 μL total volume are set up as follows: 5 μL of properly diluted cDNA is added to 5 μL of 2× SYBR green qPCR Master Mix containing the primers of interest. One has to find a compromise between low dilution for the detection of low level RNAs and sufficient dilution to prevent PCR inhibition. Because of carryover of impurities from purification steps in addition to by-products of the RT reaction, in general, a 1/40 dilution offers a good starting condition. For the quantification of RNA found at high abundance (such as

ribosomal RNA), an additional set of cDNA dilutions has to be prepared. In parallel, serial dilutions of a stoichiometric (1:1:1) mix of the 3 targets borne by plasmids (1–3, if several gene sequences are found on the same plasmid) and containing the sRNA, *kan*^{R} and *rrsB* gene target sequences of interest, are used as a standard.

3. The following thermal cycling conditions can be used as a starting protocol for the Roche Light Cycler 480 system (or any qPCR thermalcycler from any other manufacturer): 95 °C 10 min (required in the case of thermoactivation-requiring enzymes) or 2–5 min (if antibodies are used for hotstarting the reaction in the kit), followed by 45 cycles of 95 °C 10 s, 55–62 °C 10 s (according to Tm designed in primer3) and 72 °C 10 s (allow elongation of 25 bases per second).
4. Amplification is followed by a melting curve analysis using the default program of a Roche Light Cycler 480 thermal cycler, and in parallel by a denaturing PAGE electrophoresis to check for the specificity of the reaction.

3.3 sRNA Half-Life Measurements

3.3.1 Northern Blotting Procedure

1. 5 μg of total RNAs preparation (*see* Subheading 3.1) are separated on a 6% polyacrylamide denaturing gel as described in Subheading 3.2.2.
2. After PAGE running, RNAs are electro-transferred to a hybond-N membranes (GE-Healthcare) using TAE buffer. Handle the membrane with tweezers and keep it wet to avoid immediate hybridization.
3. Hybridization is performed using 5′-^{32}P labeled oligonucleotides and Roti-Hybri-Quick (Carl Roth) hybridization buffer at 42 °C for a minimum of 4 h. Membranes are washed twice in 2× SSC with 0.1% SDS (once rapidly at room temperature and once for 10 min at 42 °C) and then 5 times for 2 min in 0.2× SSC buffer with 0.1% SDS at RT.

3.3.2 Design of the Probe

The ^{32}P-labeled probe used to probe the sRNA can consist of a synthetic DNA oligonucleotide which is complementary to the sRNA sequence. Most of the time, an oligonucleotide with a size around 30 nucleotides will yield satisfactory results; however, shorter (from 18 nt) and longer (more than 50 nt) oligonucleotides can be used. A balanced 40–60% GC content should be preferred to avoid stable nonspecific hybridization. It is important to check the probe sequence for the propensity to form secondary structures, as these might alter the hybridization to the target sRNA sequence. This can be achieved using online software such as mfold (use the version for RNA molecule: http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form) or RNA fold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). Primer3 can also be used as described in Subheading 3.2.4 to

choose the probe while taking into account these parameters; in this case one has to fill the form with the target sequence and adjust the parameters cited above for oligonucleotide selection. In addition, Primer3 will facilitate avoiding repeated stretches of the same nucleotide more than 4 times. The oligonucleotide designed at this step needs to be BLASTed against a nucleotide database with the target species to ensure specificity.

Labeling of the oligonucleotide is performed at its 5′-end with [γ-^{32}P]ATP and T4 polynucleotide kinase.

3.3.3 sRNA Stability Assay (see **Note 4)**

1. *E. coli* strains were grown at 37 °C to appropriate growth phase. Rifampicin was added at 200 μg/mL to stop transcription and then samples were taken just before and at several time points after rifampicin addition (time ranging between 0 and 60 min). Each sample was chilled on ice and then total RNAs were extracted as described in Subheading 3.1. The total concentration of RNA was measured as described previously.
2. Reverse transcription was performed as described in Subheading 3.2.3. The relative quantification of transcript can be performed either by real-time RT-qPCR (Subheading 3.2) or northern blotting (Subheading 3.3).
3. Normalization was performed using the amount of total RNA in each RT reaction (250 ng) and by the level of expression of a reference gene (*rrsB*) at time zero point (*i.e.*, just before addition of rifampicin). Relative mRNA level in each sample was normalized to time zero point for each strain.

3.3.4 Half-Life Measurement

The turnover rate or stability of sRNA in vivo is usually reported as the time required for degrading 50% of the existing RNA molecules, *i.e.*, the half-life of RNA ($t_{1/2}$). Nevertheless, before half-life decay rate constant measurements, transcription of the sRNA of interest must be turned off with the use of rifampicin that inhibits bacterial DNA-dependent RNA polymerase (*see* Subheading 3.1).

1. The rate of disappearance of mRNA concentration (dC/dt) is $dC/dt = -k_{deg.}C$, with k_{deg} the rate constant of decay and C the sRNA concentration

   Thus, $dC/C = -k_{deg.}t$

   or $\left(C = C_0 \cdot e^{-k_{deg} \cdot t}\right)$, with C_0 the concentration of the sRNA before decay starts (at time 0 = addition of rifampicin)

   The half-life $t_{1/2}$ is found when C/C_0 = ½; thus, $\ln ½ = -k_{deg.}t_{1/2,}$ or $t_{1/2} = \ln 2/k_{deg}$ (*see* **Note 5**).

   The decay rate constant k_{deg} can be obtained by fitting data $C = f(t)$ to an exponential curve $\left(C = C_0 \cdot e^{-k_{deg} \cdot t}\right)$ (*see* **Note 6**).

   Half-lives are expressed as the mean +/− standard deviation of at least three independent experiments (*see* **Note 7**).

Note that the congruity of mRNA decay to simple first-order kinetics usually applies in prokaryotes, but that sometimes more complex models must be used. This is especially the case at the end of the kinetics probably due to the recruitment of different RNAses and cofactors (*see* **Note 8**).

4 Notes

1. The standard procedures exclude most RNAs less than 200 nucleotides in length. The modified procedure described here was adapted to get RNA preparations containing the small RNAs.
2. Choosing M-MLV reverse transcriptase in the RT reaction offers a good compromise between cost and efficiency as it is one of the cheapest enzyme available from several manufacturers, and it has been proven quite efficient in different reverse transcription benchmarks [38, 39].
3. In the time-course experiment, determination of sRNA half-life implies the analysis of several RNA samples collected at different intervals, and it is necessary to use an internal control to normalize the data. We suggest the use of the *rrsB* transcript for this goal.
4. The analyses described in this chapter also apply to measure half-life and concentration of mRNA targeted by sRNAs, which can be co-purified from the same culture.
5. The parameter *time constant* or *mean life time* τ may be used instead half-life $t_{1/2}$; time constant $\tau = 1/k_{deg}$.
6. Alternatively, the decay rate constant can be obtained from the slope of a semi-logarithmic plot of RNA concentration C as a function of time ($\ln C = \ln C_0 - k_{deg}.t$).
7. It is recommended to make cultures and analysis at least in triplicate ($n = 3$) to calculate the standard deviation of the mean.
8. We recommend analyzing in deeper detail the first part of the kinetics where first-order model applies.

Acknowledgments

This work was supported by the CNRS, CEA, and University Paris Diderot. We are particularly grateful to Bastien Cayrol (INRA/CIRAD) for his help in preparing this manuscript, and to Richard Lease (Ohio State Univ.) and Daniele Joseleau-Petit (Univ. Paris Diderot) for critical reading of the manuscript.

References

1. Hui MP, Foley PL, Belasco JG (2014) Messenger RNA degradation in bacterial cells. Annu Rev Genet 48:537–559
2. Gorna MW, Carpousis AJ, Luisi BF (2011) From conformational chaos to robust regulation: the structure and function of the multienzyme RNA degradosome. Q Rev Biophys 45:105–145
3. Taghbalout A, Yang Q, Arluison V (2014) The Escherichia coli RNA processing and degradation machinery is compartmentalized within an organized cellular network. Biochem J 458:11–22
4. Soper T, Mandin P, Majdalani N, Gottesman S, Woodson SA (2010) Positive regulation by small RNAs and the role of Hfq. Proc Natl Acad Sci U S A 107:9602–9607
5. Aiba H (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. Curr Opin Microbiol 10:134–139
6. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. Nat Rev Microbiol 9:578–589
7. Ikeda Y, Yagi M, Morita T, Aiba H (2011) Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli. Mol Microbiol 79:419–432
8. Mohanty BK, Maples VF, Kushner SR (2004) The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli. Mol Microbiol 54:905–920
9. Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci U S A 81:1991–1995
10. Andrade JM, Pobre V, Matos AM, Arraiano CM (2012) The crucial role of PNPase in the degradation of small RNAs that are not associated with Hfq. RNA 18:844–855
11. Saramago M, Barria C, Dos Santos RF, Silva IJ, Pobre V, Domingues S, Andrade JM, Viegas SC, Arraiano CM (2014) The role of RNases in the regulation of small RNAs. Curr Opin Microbiol 18:105–115
12. Andrade JM, Arraiano CM (2008) PNPase is a key player in the regulation of small RNAs that control the expression of outer membrane proteins. RNA 14:543–551
13. Viegas SC, Arraiano CM (2008) Regulating the regulators: how ribonucleases dictate the rules in the control of small non-coding RNAs. RNA Biol 5:230–243
14. Masse E, Escorcia FE, Gottesman S (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. Genes Dev 17:2374–2383
15. Viegas SC, Silva IJ, Saramago M, Domingues S, Arraiano CM (2011) Regulation of the small regulatory RNA MicA by ribonuclease III: a target-dependent pathway. Nucleic Acids Res 39:2918–2930
16. Marujo PE, Hajnsdorf E, Le Derout J, Andrade R, Arraiano CM, Regnier P (2000) RNase II removes the oligo(A) tails that destabilize the rpsO mRNA of Escherichia coli. RNA 6:1185–1193
17. Reichenbach B, Maes A, Kalamorz F, Hajnsdorf E, Gorke B (2008) The small RNA GlmY acts upstream of the sRNA GlmZ in the activation of glmS expression and is subject to regulation by polyadenylation in Escherichia coli. Nucleic Acids Res 36:2570–2580
18. Sledjeski DD, Whitman C, Zhang A (2001) HFq is necessary for regulation by the untranslated RNA DsrA. J Bacteriol 183:1997–2005
19. De Lay N, Schu DJ, Gottesman S (2013) Bacterial small RNA-based negative regulation: Hfq and its accomplices. J Biol Chem 288:7996–8003
20. Vanzo NF, Li YS, Py B, Blum E, Higgins CF, Raynal LC, Krisch HM, Carpousis AJ (1998) Ribonuclease E organizes the protein interactions in the Escherichia coli RNA degradosome. Genes Dev 12:2770–2781
21. Khemici V, Poljak L, Luisi BF, Carpousis AJ (2008) The RNase E of Escherichia coli is a membrane-binding protein. Mol Microbiol 70:799–813
22. Kuwano M, Ono M, Endo H, Hori K, Nakamura K, Hirota Y, Ohnishi Y (1977) Gene affecting longevity of messenger RNA: a mutant of Escherichia Coli with altered mRNA stability. Mol Gen Genet 154:279–285
23. Ono M, Kuwano M (1979) A conditional lethal mutation in an Escherichia Coli strain with a longer chemical lifetime of messenger RNA. J Mol Biol 129:343–357
24. McDowall KJ, Hernandez RG, Lin Chao S, Cohen SN (1993) The ams-1 and rne-3071 temperature-sensitive mutations in the ams gene are in close proximity to each other and cause substitutions within a domain that resembles a product of the Escherichia coli mre locus. J Bacteriol 175:4245–4249
25. Kido M, Yamanaka K, Mitani T, Niki H, Ogura T, Hiraga S (1996) RNase E polypeptides lacking a carboxyl-terminal half suppress a mukB mutation in Escherichia coli. J Bacteriol 178:3917–3925
26. Ow MC, Kushner SR (2002) Initiation of tRNA maturation by RNase E is essential for

cell viability in E. coli. Genes Dev 16: 1102–1115

27. Anupama K, Leela JK, Gowrishankar J (2011) Two pathways for RNase E action in Escherichia coli in vivo and bypass of its essentiality in mutants defective for Rho-dependent transcription termination. Mol Microbiol 82: 1330–1348
28. Jain C, Belasco JG (1995) RNase E autoregulates its synthesis by controlling the degradation rate of its own mRNA in Escherichia Coli: unusual sensitivity of the rne transcript to RNase E activity. Genes Dev 9:84–96
29. Morita T, Maki K, Aiba H (2005) RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. Genes Dev 19:2176–2186
30. Morita T, Kawamoto H, Mizota T, Inada T, Aiba H (2004) Enolase in the RNA degradosome plays a crucial role in the rapid decay of glucose transporter mRNA in the response to phosphosugar stress in Escherichia coli. Mol Microbiol 54:1063–1075
31. Mackie GA (2013) RNase E: at the interface of bacterial RNA processing and decay. Nat Rev Microbiol 11:45–57
32. Garrey SM, Mackie GA (2011) Roles of the 5′-phosphate sensor domain in RNase E. Mol Microbiol 80:1613–1624
33. Regnier P, Hajnsdorf E (1991) Decay of mRNA encoding ribosomal protein S15 of Escherichia coli is initiated by an RNase E-dependent endonucleolytic cleavage that removes the 3′ stabilizing stem and loop structure. J Mol Biol 217:283–292
34. Donovan WP, Kushner SR (1986) Polynucleotide phosphorylase and ribonuclease II are required for cell viability and mRNA turnover in Escherichia coli K-12. Proc Natl Acad Sci U S A 83:120–124
35. Yancey SD, Kushner SR (1990) Isolation and characterization of a new temperature-sensitive polynucleotide phosphorylase mutation in Escherichia coli K-12. Biochimie 72:835–843
36. Andrade JM, Hajnsdorf E, Regnier P, Arraiano CM (2009) The poly(A)-dependent degradation pathway of rpsO mRNA is primarily mediated by RNase R. RNA 15:316–326
37. Rasmussen AA, Eriksen M, Gilany K, Udesen C, Franch T, Petersen C, Valentin-Hansen P (2005) Regulation of ompA mRNA stability: the role of a small regulatory RNA in growth phase-dependent control. Mol Microbiol 58:1421–1429
38. Okello JB, Rodriguez L, Poinar D, Bos K, Okwi AL, Bimenya GS, Sewankambo NK, Henry KR, Kuch M, Poinar HN (2010) Quantitative assessment of the sensitivity of various commercial reverse transcriptases based on armored HIV RNA. PLoS One 5:e13931
39. Stahlberg A, Kubista M, Pfaffl M (2004) Comparison of reverse transcriptases in gene expression analysis. Clin Chem 50:1678–1680
40. Plumbridge JA, Dondon J, Nakamura Y, Grunberg-Manago M (1985) Effect of NusA protein on expression of the nusA, infB operon in E. coli. Nucleic Acids Res 13:3371–3388
41. Hajnsdorf E, Steier O, Coscoy L, Teysset L, Regnier P (1994) Roles of RNase E, RNase II and PNPase in the degradation of the rpsO transcripts of Escherichia oli: stabilizing function of RNase II and evidence for efficient degradation in an ams pnp rnb mutant. EMBO J 13:3368–3377
42. Ziolkowska K, Derreumaux P, Folichon M, Pellegrini O, Regnier P, Boni IV, Hajnsdorf E (2006) Hfq variant with altered RNA binding functions. Nucleic Acids Res 34:709–720
43. Arraiano CM, Yancey SD, Kushner SR (1988) Stabilization of discrete mRNA breakdown products in ams pnp rnb multiple mutants of Escherichia oli K-12. J Bacteriol 170: 4625–4633
44. Hajnsdorf E, Braun F, Haugel-Nielsen J, Regnier P (1995) Polyadenylylation destabilizes the rpsO mRNA of Escherichia coli. Proc Natl Acad Sci U S A 92:3973–3977
45. Nogueira T, de Smit M, Graffe M, Springer M (2001) The relationship between translational control and mRNA degradation for the Escherichia coli threonyl-tRNA synthetase gene. J Mol Biol 310:709–722
46. Masse E, Gottesman S (2002) A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. Proc Natl Acad Sci U S A 99:4620–4625
47. Lu F, Taghbalout A (2013) Membrane association via an amino-terminal amphipathic helix is required for the cellular organization and function of RNase II. J Biol Chem 288: 7241–7251

Part IV

Cofactors of sRNA-Based Regulation

Chapter 15

High-Resolution, High-Throughput Analysis of Hfq-Binding Sites Using UV Crosslinking and Analysis of cDNA (CRAC)

Brandon Sy, Julia Wong, Sander Granneman, David Tollervey, David Gally, and Jai J. Tree

Abstract

Small regulatory nonprotein-coding RNAs (sRNAs) have emerged as ubiquitous and abundant regulators of gene expression in a diverse cross section of bacteria. They play key roles in most aspects of bacterial physiology, including central metabolism, nutrient acquisition, virulence, biofilm formation, and outer membrane composition. RNA sequencing technologies have accelerated the identification of bacterial regulatory RNAs and are now being employed to understand their functions. Many regulatory RNAs require protein partners for activity, or modulate the activity of interacting proteins. Understanding how and where proteins interact with the transcriptome is essential to elucidate the functions of the many sRNAs. Here, we describe the implementation in bacteria of a UV-crosslinking technique termed CRAC that allows stringent, transcriptome-wide recovery of bacterial RNA–protein interaction sites in vivo and at base-pair resolution. We have used CRAC to map protein–RNA interaction sites for the RNA chaperone Hfq and ribonuclease RNase E in pathogenic *E. coli*, and toxins from toxin–antitoxin systems in *Mycobacterium smegmatis*, demonstrating the broad applicability of this technique.

Key words Protein–RNA interaction, Small RNA, Noncoding RNA, RBP, RNA-binding protein

1 Introduction

Bacterial pathogens colonize complex and changing host and environmental niches that require rapid signal perception and responses to survive. In the model prokaryote *E. coli str.* K-12, transcriptional responses are controlled by approximately 213 transcription factors [1, 2] that mediate some 3477 interactions [1] with the chromosome. These transcription factors integrate external and internal signals, and produce a transcriptional landscape tailored to the perceived environment. In recent years, it has become clear that there exists a post-transcriptional regulatory network of potentially comparable dimensions. This is controlled by small RNAs (sRNAs) that generally do not encode proteins but allow the integration of additional information into the regulatory

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_15,

response [3, 4]. The emerging picture is of a gene regulatory network that consists of both transcriptional and post-transcriptional inputs that ensures precise responses to complex signals.

Canonically, sRNAs control gene expression by forming short duplexes on mRNAs at the ribosomal-binding site, thereby blocking 30S ribosomal subunit recruitment and translation [5]. While mRNA targeting is directed by the sRNA "seed" sequence and structure, the interaction is facilitated by protein partners including the RNA chaperone Hfq [6], and the recently identified FinO/ProQ protein family [7, 8]. Further interactions with RNases (notably the major RNA degradosome component RNase E) are mediated by both Hfq–RNase and RNA–RNase interactions, and lead to degradation of the duplexed sRNA and/or mRNA [9]. In addition to this canonical pathway, a number of new regulatory mechanisms have emerged. Small RNA interactions have been described that promote translational activation [10], guide and inhibit processing of transcripts [9, 11], antagonize sRNA interactions by acting as "sponges" [12, 13], and control transcription termination [14–16]. Defining the in vivo binding sites of RNA chaperones and RNases within mRNAs and sRNAs has provided information that can be used to understand how these transcripts are regulated post-transcriptionally. High-resolution binding site data can define the sequence specificity of the protein, provide insights into the mechanism of protein function, and suggest functional outcomes for interactions based on identified protein-binding sites.

Here we present a protocol for defining Hfq-binding sites transcriptome-wide and with base-pair resolution in bacteria termed UV crosslinking and analysis of cDNA (CRAC) (Fig. 1). CRAC is differentiated from closely related CLIP-Seq protocols (HITS-CLIP, CLIP-Seq, iCLIP, and eCLIP) by the inclusion of tandem affinity purification and denaturing purification, providing higher stringency in the recovery of authentic RNA–protein interaction sites [17]. More recently, the Granneman lab has developed a very efficient UV-crosslinking device (Vari-X-linker) that makes it possible to monitor the dynamics of Hfq binding to its substrates

Fig. 1 (continued) panel are given on the right. Briefly, Hfq–RNA complexes are UV-crosslinked in vivo (panel 1), and the soluble fraction from cell lysates loaded onto M2 anti-FLAG resin. (Panel 2) Hfq–RNA complexes are purified and eluted using TEV protease cleavage, trimmed using a cocktail of RNase A and RNase T1, and loaded onto Ni-NTA resins for denaturing purification. (Panel 3) The 5′ and 3′ ends of the RNA are dephosphorylated and labeled with ^{32}P-ATP (indicated by star), and then ligated to RNA linkers on column. (Panel 4) The labeled Hfq–RNA complex is size selected on an SDS-PAGE gel and transferred to nitrocellulose where the appropriately sized Hfq–RNA complex is excised. The crosslinked RNA is then recovered by proteinase K digestion of Hfq and organic extraction of the released RNA. (Panel 5) The RNA is reverse transcribed and amplified by PCR to generate libraries compatible with Illumina sequencing platforms. High throughput sequencing data is analyzed using the pyCRAC software package

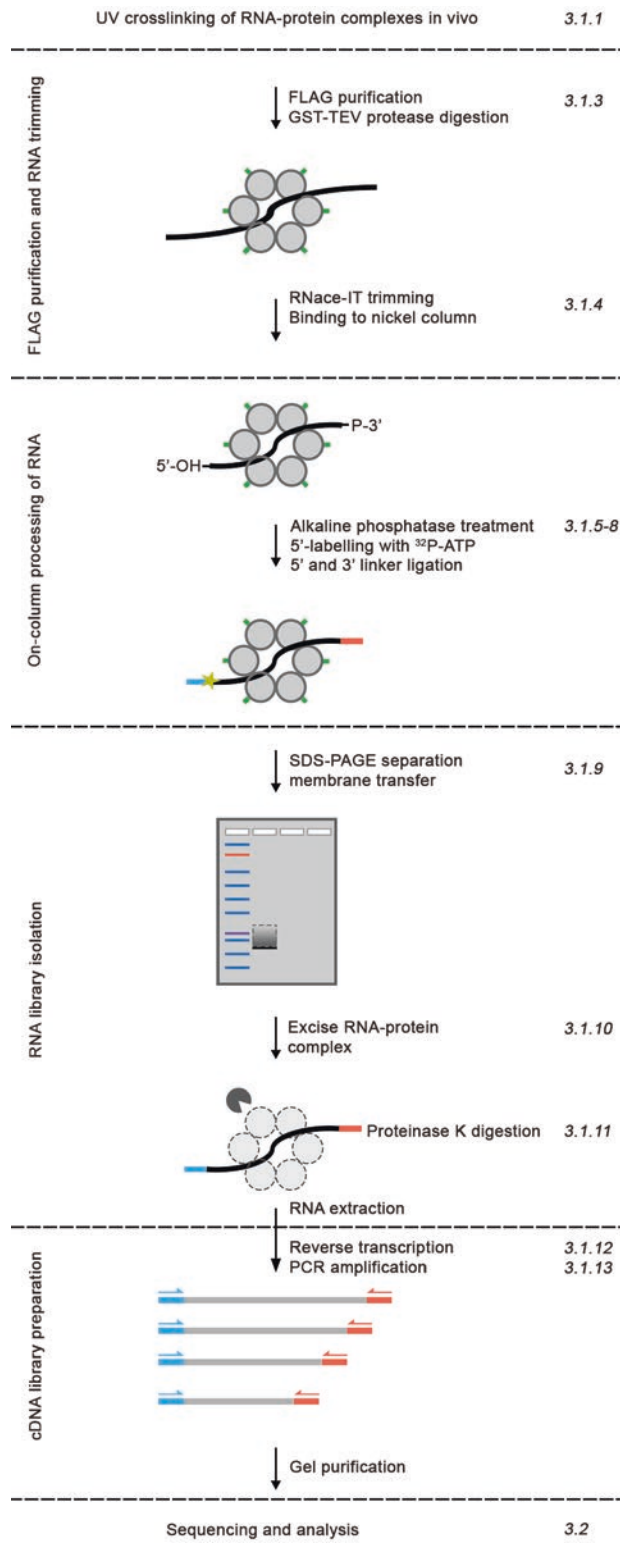


Fig. 1 Overview of the UV crosslinking and analysis of cDNA (CRAC) protocol used for identification of Hfq-binding sites. The steps of the protocol represented in each

in vivo at minute time-point resolution [18]. To facilitate Hfq purification, we modified the previously used His_6-TEV-Protein A tag [19] to incorporate a smaller FLAG affinity tag (His_6-TEV-FLAG [HTF]). This is better tolerated by the Hfq hexamer, and Hfq-HTF retains RNA chaperone activity. Importantly, the HTF tag was inserted into the chromosomal *hfq* gene to avoid off-target effects that might be associated with overexpression or variability on a plasmid-based system. To generate covalently linked RNA–Hfq complexes in vivo, cells were treated with UV-C (254 nm) generating RNA radicals that rapidly react with proteins in direct contact with the affected nucleotide (zero length crosslinking). The cells were then lysed and complexes purified using M2 anti-FLAG affinity resin. Hfq–RNA complexes were eluted and cross-linked RNAs trimmed using RNase A/T1, leaving a protected "footprint" of the protein-binding site on the RNA. Trimmed complexes were denatured, immobilized on Ni-NTA affinity resin, and washed under denaturing conditions to dissociate copurifying proteins and complexes. The enzymatic steps in which RNA 3′ and 5′ ends are prepared, labeled with ^{32}P, and linkers ligated were performed on-column. The linker-ligated Hfq-associated RNA was eluted from the Ni-NTA resin and size selected on an SDS-PAGE gel. Following elution, Hfq was degraded by treatment with Proteinase K digestion, releasing the bound RNA. The recovered RNA fragments were identified by reverse transcription, PCR amplification and sequencing using an Illumina platform.

CRAC offers the advantage of stringent purification and on-bead linker ligation that simplifies separation of reaction constituents during successive enzymatic steps. Because linkers are ligated to the protein–RNA complex, a disadvantage is that UV crosslinking of the RNA at, or near, the 5′ or 3′ end may sterically hinder on-column (de)phosphorylation or linker ligation.

In our own studies, we have used CRAC to identify the RNA targets of TA-systems in *Mycobacterium* and to map Hfq interactions in enterohemorrhagic *Escherichia coli* (EHEC) [12, 17]. Since this publication, there have been a number of transcriptome-wide interaction studies on sRNAs and their chaperones. CLIP-Seq (UV crosslinking and immunoprecipitation sequencing) has been applied to Hfq and CsrA in *Salmonella* Typhimurium revealing the binding specificity of these global posttranscriptional regulators [20]. A further high-throughput technique, termed Grad-Seq, has identified ProQ as an important RNA chaperone in *Salmonella* [7]. RNase E cleavage and binding sites have also been defined using TIER-Seq and CLASH [21, 22]. Significant advances have also been made in high throughput analysis of the sRNA-mRNA interactome (RIL-seq, CLASH, MAPS, GRIL-seq) [4, 22–24]. Overlaying this data with RNA–protein interaction networks will begin to define the full extent, and range of mechanisms, of posttranscriptional regulation in vivo.

2 Materials

2.1 Bacterial Strains and Culture Media

Purification of the RNA–protein complex requires that the protein of interest is tagged with the HTF dual affinity tag [22]. For our studies on the function of Hfq in enterohemorrhagic *E. coli* under type 3 secretion inducing conditions (supplemented MEM-HEPES), we have tagged the chromosomal copy of *hfq* to ensure native expression levels of the protein and to avoid off-target binding that may be associated with overexpression of RNA-binding proteins [12].

1. Bacterial strains: Enterohemorrhagic *Escherichia coli* O157:H7 [25] and the isogenic *hfq*-HTF mutant (JJT101) [12].
2. LB medium to culture inoculum.
3. MEM-HEPES (Sigma-Aldrich, M72781) supplemented with 0.1% glucose and 250 nM $Fe(NO_3)_3$.

2.2 Buffers and Solutions

All buffers (excepting those containing SDS) are chilled on ice before use.

1. Phosphate-buffered saline (PBS).
2. Lysis buffer: 50 mM Tris–HCl pH 7.8, 150 mM sodium chloride, 500 mM $MgCl_2$, 0.1% Nonidet P-40, 5 mM β-mercaptoethanol, 1 tablet of EDTA-free cOmplete protease inhibitor cocktail (Roche, 11697498001).
3. TNM1000 buffer: 50 mM Tris–HCl pH 7.8, 1 M NaCl, 500 mM $MgCl_2$, 0.1% nonidet P-40, 5 mM β-mercaptoethanol.
4. TNM150 buffer: 50 mM Tris–HCl pH 7.8, 150 mM NaCl, 500 mM $MgCl_2$, 0.1% nonidet P-40, 5 mM β-mercaptoethanol.
5. Wash buffer I: 6 M guanidine hydrochloride, 50 mM Tris–HCl pH 7.8, 300 mM NaCl, 0.1% nonidet P-40, 5 mM β-mercaptoethanol, 10 mM imidazole pH 8.0.
6. Wash buffer II: 50 mM Tris–HCl pH 7.8, 50 mM NaCl, 10 mM imidazole pH 8.0, 0.1% nonidet P-40, and 5 mM β-mercaptoethanol.
7. 1× PNK buffer: 50 mM Tris–HCl pH 7.8, 50 mM $MgCl_2$, 0.1% nonidet P-40, 5 mM β-mercaptoethanol.
8. 5× PNK buffer: 250 mM Tris–HCl pH 7.8, 50 mM $MgCl_2$, 50 mM β-mercaptoethanol.
9. Elution buffer: 50 mM Tris–HCl pH 7.8, 50 mM NaCl, 150 mM imidazole pH 8.0, 0.1% nonidet P-40, 5 mM β-mercaptoethanol.
10. Guanidine hydrochloride.
11. 5 M NaCl.
12. 2.5 mM imidazole pH 8.0.

13. Trichloroacetic acid (TCA).
14. Acetone.
15. Proteinase buffer: 50 mM Tris–HCl pH 7.8, 50 mM NaCl, 0.1% nonidet P-40, and 5 mM β-mercaptoethanol, 1% sodium dodecyl sulfate (v/v), 5 mM EDTA.
16. Proteinase K solution (20 mg/mL).
17. 3 M NaAc pH 5.2.
18. 25:24:1 phenol:chloroform:isoamylalcohol mixture.
19. 100% and 70% ethanol (stored at −20 °C).
20. 10× TBE buffer: 890 mM Tris base, 890 mM boric acid, 20 mM EDTA.
21. MilliQ water.

2.3 Enzymes and Enzymatic Reaction Components

1. RNaceIT (Agilent) diluted 1:60 in TNM150 buffer.
2. Homemade GST-TEV protease (do not use His-tagged TEV).
3. Thermosensitive alkaline phosphatase (TSAP) (Promega, M9910).
4. RNasIN RNase inhibitor (Promega, N2511).
5. T4 RNA ligase 1 (New England Biolabs, M0204S).
6. ^{32}P-γATP (Perkin-Elmer, BLU502A250UC).
7. 10 mM ATP.
8. 10 mM deoxyribonucleotides (Sigma-Aldrich, D7295).
9. Superscript III and accompanying 5× first strand buffer (Invitrogen, 18080044).
10. 100 mM DTT (Invitrogen, accompanies 18080044).
11. RNase H (New England Biolabs, M0297S).
12. LA Taq polymerase (TaKaRa, RR002M).
13. 10× LA Taq PCR Buffer (TaKaRa, accompanies RR002M).

2.4 Oligonucleotides

All oligonucleotides were supplied by Integrated DNA Technologies and are listed in Table 1. The forward and reverse PCR primers introduce sequences that allow binding of the PCR product to an Illumina flow cell. The sequence introduced by PE_miRCat_33_PCR is now obsolete and cannot be used for paired-end sequencing.

Adapters and PCR oligonucleotides for paired-end sequencing of CRAC libraries have been described in van Nues et al. (2017) [18].

2.5 Laboratory Equipment

1. Incubator with orbital shaker.
2. UV crosslinker (LT 40-205, Van Remmen UV Techniek).
3. Refrigerated centrifuge for 1 L bottles.

Table 1
List of linkers and primers

Oligonucleotide name	Oligonucleotide sequence
5′ and 3′ linkers for on-column library preparation[a]	
L5Aa	5′-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrU*rNrNrNrN*rUrArArGrCrN-3′
L5Ab	5′-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrU*rNrNrNrN*rArUrUrArGrCrN-3′
L5Ac	5′-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrU*rNrNrNrN*rGrCrGrCrArGrCrN-3′
L5Ad	5′-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrU*rNrNrNrN*rCrGrCrUrUrArGrCrN-3′
miRCat-33[b]	5′-AppNTGGAATTCTCGGGTGCCAAGG-ddC-3′.
App_PE	5′AppNAGATCGGAAGAGCACACGTCTG-ddC-3′
Primers for reverse transcription and PCR amplification	
miRCat-33-RT	5′-CCTTGGCACCCGAGAATT-3′
PE_miRCat-33_PCR	5′ CAAGCAGAAGACGGCATACGAGATCGGTCTCGGC ATTCCTGGCCTTGGCACCCGAGAATTCC-3′
P5	5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′
PE_reverse	5′-CAGACGTGTGCTCTTCCGATCT-3′
P3_BC1[c]	5′-CAAGCAGAAGACGGCATACGAGAT*CGTGAT*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′
P3_BC2[c]	5′-CAAGCAGAAGACGGCATACGAGAT*ACATCG*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′
P3_BC3[c]	5′-CAAGCAGAAGACGGCATACGAGAT*GCCTAA*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′

[a]5′ RNA linkers contain barcode sequences to allow multiplexing (underlined sequence) and random nucleotides to assess PCR duplication of reads (italicized sequence). The 5′ end contains an inverted dideoxy thymidine cap

[b]The miRCat-33 is 5′ adenylated (App) to facilitate ligation in the absence of ATP. The 3′ carries a dideoxy cytosine cap to prevent linker concatemerization. This linker has been discontinued, but the sequence can be custom synthesized with the appropriate 5′ and 3′ modifications

[c]The P3 reverse primer for amplifying the cDNA library contains a six-nucleotide barcode that corresponds to Illumina indices. This allows for demultiplexing during the sequencing reaction

4. Refrigerated centrifuge for 50 mL and 15 mL centrifuge tubes.
5. Rotary shaker stored at 4 °C.
6. Temperature controlled dry block (with range 16 °C–65 °C).
7. Refrigerated microcentrifuge.
8. Mini Trans Blot western transfer apparatus (BioRad).
9. Phophorimaging cassette.
10. Film processing machine for developing films.
11. Bunsen burner.
12. Thermocycler for cDNA synthesis.
13. Apparatus for agarose gel electrophoresis.
14. Qubit 3.0 Fluorometer (Thermo Scientific).
15. Vortex.

2.6 Other Consumables and Labware

1. 2 L conical flask (per sample).
2. Filtered and RNase-free pipette tips.
3. MEM-HEPES media (Sigma-Aldrich, M7278) supplemented with 0.1% glucose and 250 nM $Fe(NO_3)_3$ (800 mL per sample).
4. 0.1 mm Zirconia/silica beads (Daintree scientific).
5. M2 Anti-FLAG resin (Sigma, F3165).
6. Spin columns (Pierce, Snap Cap).
7. Ni-NTA resins (Qiagen, 30210).
8. 1.5 mL microcentrifuge tubes.
9. GlycoBlue (Ambion, AM9515).
10. NuPAGE bis-Tris 4–12% precast gradient gels (Invitrogen, NP0322BOX).
11. MOPS running buffer (Invitrogen, NP0001).
12. NuPAGE transfer buffer (Invitrogen, NP0006).
13. Nitrocellulose membranes (Thermo Scientific or GE Healthcare).
14. Kodak BioMax MS Autoradiography Film.
15. MinElute PCR purification kit (Qiagen, 28004).
16. MetaPhor high-resolution agarose (Lonza, 50181).
17. SYBRSafe (Life Technologies, S33102).
18. DNA ladder and loading dye (e.g., GeneRuler DNA Ladder Mix by Thermo Scientific, SM0331).
19. Pre-stained protein standard SeeBlue Plus2 (Life Technologies, LC5925).
20. Scalpels.
21. Qubit dsDNA HS Assay Kit (Life Technologies, Q32851).

3 Methods

3.1 Generating Hfq-CRAC Libraries for Sequencing

Appropriate negative controls and experimental replicates are required to determine the background signal and true positive binding sites. We routinely use the parental (untagged) strain as a negative control and perform a minimum of biological replicates for each sample. Alternative controls and experimental designs are discussed with our data analysis approach (Subheading 3.1.4).

3.1.1 Growth and Crosslinking of RNA-Hfq-His-TEV-FLAG (HTF) Complexes

1. Streak out glycerol stocks of the HTF tagged and negative control strains onto LB-agar plates and incubate at 37°C overnight.
2. Inoculate colonies into 25 mL of LB broth and incubate at 37°C shaking at 200 rpm overnight.
3. Inoculate 800 mL of supplemented MEM-HEPES with 8 mL of overnight culture and grow until the OD_{600} measurement reaches 0.8.
4. Decontaminate and warm-up the UV crosslinker before use.
5. Add bacterial culture into the UV crosslinker and expose for 100 s (1800 mJ). The chamber can be gently rocked to ensure circulation of the cell around the UV-lamp.
6. Pour the cultures into pre-chilled 1 L centrifuge bottles on ice. Centrifuge at 4000 × *g* for 20 min at 4°C.
7. Resuspend pellet in 40 mL of ice-cold PBS and transfer to a 50 mL Falcon tube. Centrifuge at 4000 × *g* for 20 min at 4°C.
8. Decant supernatant and record the pellet weight (subdivide the sample into 1 g pellets).
9. Snap freeze in liquid nitrogen, and store pellets at −80°C until required.

3.1.2 Cell Lysis

1. Add one volume of ice-cold LYSIS buffer and three volumes of 0.1 mm zirconia beads to the cell pellet. Vortex the cells five times at 1 min intervals, putting them on ice for 1 min between each vortex.
2. Add three volumes of LYSIS buffer and vortex once again. Centrifuge for 20 min at 4000 × *g* and at 4°C. Transfer the supernatant into chilled 1.5 mL microcentrifuge tubes and spin at 16,000 × *g* for 20 min at 4°C. Keep tubes on ice throughout the entire process.

3.1.3 Purification with M2 Anti-FLAG Resin

1. Wash 200 μL M2 anti-FLAG resin (50% slurry) (per sample) with 5 mL of LYSIS buffer. Gently invert the bottle of resin before use. Collect the resin by pulsing the centrifuge at 1000 rpm. Repeat the wash once more.

2. Remove the wash buffer from the resin before adding the supernatant from Subheading 3.2.2. Incubate the samples on an orbital shaker for 2 h at 4°C.
3. Pulse the tubes at 1000 × *g* to collect the resin. Remove the supernatant.
4. Wash the resin twice with 5 mL of TNM1000 and twice with TNM150.
5. Resuspend the resin in 600 µL of TNM150 and transfer to a microcentrifuge tube. Use wide orifice pipette tips or use a scalpel to cut the end off of a regular pipette tip for more efficient pipetting of the resin.
6. Add 1.5 µL (20–30 units) of TEV protease to the resin.
7. Incubate at 18°C for 2 h with shaking.
8. Spin the resin suspension through a microcentrifuge tube spin column to capture the resin. Store a sample of the eluate for troubleshooting.

3.1.4 RNase Digestion and Binding to Ni-NTA Resin

1. Prepare 1.5 mL microcentrifuge tubes containing 0.4 g of guanidine hydrochloride.
2. Prepare Ni-NTA resin (50% slurry) by washing 100 µL of resin (per sample) with 750 µL of Wash Buffer I twice. Leave the resin in about 100 µL of wash buffer until use.
3. Add 1 µL of a 1/60 dilution of RNace IT (0.15 units, Agilent; *see* **Note 1**) to 500 µL of the TEV eluate. Incubate for precisely 5 min at 37°C.
4. Immediately transfer the RNase-treated eluate into the pre-prepared microcentrifuge tubes containing guanidine hydrochloride. The final volume should be around 800 µL.
5. Add 27 µL of 5 M NaCl (300 mM) and 3 µL of 2.5 M pH 8.0 imidazole (10 mM) to the eluate.
6. Remove the remaining supernatant from the washed Ni-NTA resin. Add the eluate to the resin.
7. Incubate with gentle agitation at 4°C overnight.
8. Pulse the microcentrifuge tubes to pellet the resin. Remove the supernatant and wash twice with 750 µL of WASH BUFFER I and three times with 750 µL of 1× PNK buffer.
9. Transfer the resin to a microcentrifuge spin column.

3.1.5 Dephosphorylation of RNA 3′ P Ends Using Alkaline Phosphatase

1. Spin out the residual 1× PNK buffer and cap the bottom of the column. To each sample, add 80 µL of TSAP master mix that contains 16 µL of 5× PNK buffer, 8 µL of TSAP, and 2 µL of RNasIN.
2. Incubate at 37°C for 45 min.

3. Wash the resin once with 400 μL of WASH BUFFER I and three times with 400 μL of 1× PNK buffer.

3.1.6 On Bead Ligation of 3′ miRCat-33 Linker

1. Spin out the residual volume of 1× PNK buffer. Cap the bottom of the column. To the resin, add 76 μL of miRCat master mix containing 16 μL of 5× PNK buffer, 8 μL of 10 μM miRCat-33, 2 μL of RNasIN, and 50 μL of Milli-Q water (*see* **Note 2**).
2. To each sample add 4 μL of T4 RNA ligase I.
3. Incubate at 25°C for 6 h.
4. Wash the resin once with 400 μL of WASH BUFFER I and three times with 400 μL of 1× PNK buffer.

3.1.7 Phosphorylating the 5′ Ends of Crosslinked RNA

1. Spin out the residual volume of 1× PNK buffer. Cap the bottom of the column. To the resin, add 80 μL of PNK master mix containing 16 μL of 5× PNK buffer, 4 μL of T4 polynucleotide kinase, 56 μL of Milli-Q, and 4 μL of ^{32}P-γATP (10 μCi/μL). Incubate at 37°C for 40 min.
2. Add 1 μL of 100 mM ATP and continue the incubation for 20 min.
3. Wash the resin three times with 400 μL of WASH BUFFER I and three times with 1× PNK buffer.

3.1.8 On-Column Ligation of the 5′ Linker

1. Spin out the residual volume of 1× PNK buffer. Add 75 μL of 5′ linker master mix containing 16 μL of 5× PNK, 8 μL of 10 mM ATP, 2 μL of RNasIN, and 49 μL of Milli-Q water.
2. To each sample add 1 μL of bar-coded 5′ linker and 4 μL of T4 RNA ligase I.
3. Incubate overnight at 16°C.
4. Wash the resin three times with WASH BUFFER II.

3.1.9 Elution and Precipitation Hfq–RNA Complexes

1. Spin out the void volume. Cap the column and add 200 μL of ELUTION buffer.
2. Incubate the resin on ice for 5 min.
3. Collect the flow-through in a 1.5 mL microcentrifuge tube. Re-cap the column and repeat the elution with another 200 μL of ELUTION buffer (*see* **Note 3**). Collect the residual ELUTION buffer on the column by zip spinning the column.
4. Pool the eluates into a single microcentrifuge tube. Add 40 μg of GlycoBlue co-precipitant and 100 μL of trichloroacetic acid. Vortex and incubate on ice for an hour.
5. Centrifuge at top speed for 30 min at 4°C. Remove the supernatant (use the Geiger counter to ensure that the pellet hasn't been dislodged if the blue pellet isn't visible).

6. Add 800 μL of ice-cold acetone to the pellet and centrifuge for 20 min at 4°C (*see* **Note 4**).
7. Remove all the supernatant and air-dry the pellet for a few minutes in a fume hood. Make sure that you do not over-dry your samples as you may not be able to resuspend the pellet in loading dye (see below).

3.1.10 Size Selection of Hfq–RNA Complexes

1. Resuspend the pellet in 30 μL of 1× NuPAGE LDS sample buffer. Load the sample onto a NuPAGE 4–12% gradient gel. Run for 50 min at 200 V or until the dye reaches the bottom of the gel.
2. Transfer the protein–RNA complex to nitrocellulose using a wet transfer western blotting system. Transfer the proteins at 100 V for 1.5 h (*see* **Notes 5** and **6**). Expose the membrane to a high-sensitivity X-ray film. If samples are highly radioactive, a 10–30 min exposure time should suffice. Ensure that chemiluminescent marker is included to realign and film after developing.
3. Develop the X-ray film and align it to the membrane. Cut out the smear corresponding to the size of the protein–RNA complex (*see* **Note 7**).

3.1.11 Recovery of Trimmed and Adapter Ligated RNA

1. Incubate the membrane slices with 400 μL of Proteinase buffer (50 mM Tris–HCl pH 7.8, 50 mM NaCl, 0.1% NP-40, 5 mM β-mercaptoethanol, 1% SDS, 5 mM EDTA) containing 100 μg of Proteinase K (Roche). Make sure the membrane slices are completely submerged. To be able to more accurately quantify the background signal, membrane slices containing from the control and HTF samples can be pooled so that the differentially bar-coded RNA samples can be reverse transcribed and PCR amplified in the same reaction tube [18].
2. Incubate at 55°C for 2 h in a shaking heating block.
3. Add 50 μL of 3 M sodium acetate (pH 5.2) and 500 μL of phenol:chloroform:isoamyl alcohol (25:24:1). Vortex and centrifuge for 5 min in a microcentrifuge at full speed at room temperature.
4. Transfer the aqueous phase to clean microcentrifuge tube containing 500 μL of phenol-chloroform-isoamyl alcohol (25:24:1). Vortex vigorously and centrifuge for 3 min in a microcentrifuge at full speed at room temperature.
5. Transfer the supernatant to a tube containing 400 μL of chloroform. Vortex vigorously and centrifuge for 3 min in a microcentrifuge at full speed at room temperature. Transfer the supernatant to a new 1.5 mL Eppendorf tube.
6. Add 1 mL of ice-cold absolute ethanol and 20 μg of GlycoBlue. Incubate at −80°C for 30 min and centrifuge at 16,000 × *g*

and 4°C for 30 min. Wash the pellet with 500 μL of ice-cold 70% ethanol and centrifuge for 20 min. Aspirate the pellet and air dry.

3.1.12 Reverse Transcription of Purified RNA

1. Resuspend the RNA pellet in 9 μL of MilliQ water and transfer into a 0.2 mL PCR tube. Add 2 μL of RT oligo and 2 μL of 5 mM dNTPs. Perform subsequent incubation in a thermocycler.
2. Heat the samples to 80°C for 3 min, then chill on ice for 5 min.
3. To each sample, add 4 μL of 5× First Strand buffer (Invitrogen), 1 μL of 100 mM DTT, and 1 μL of RNasIN.
4. Incubate at 50°C for 3 mins and add 1 μL of SuperScript III (Invitrogen).
5. Incubate at 50°C for 1 h.
6. Inactivate the SuperScript III by incubating the samples at 65°C for 15 min.
7. Add 2 μL of RNase H and incubate for 30 min at 37°C.

3.1.13 PCR Amplification of cDNA Libraries and Gel Purification of Amplicons

1. To 2 μL of cDNA template, add 48 μL of PCR master mix containing: 5 μL of 10× LA Taq buffer, 1 μL of 10 μM P5 Solexa primer, 1 μL of 10 μM paired_end_miRCat.R reverse primer, 2.5 μL of 5 mM dNTPs, 0.5 μL of LA TaKaRa Taq polymerase, and 38 μL of nuclease-free water. We prepare 3–10 PCR reactions per sample to increase the complexity of our libraries.
2. The reaction is run with cycling conditions (*see* **Note 8**):

Temp. (°C)	Time	Cycle
95	2 min	
98 52 68	20 s 20 s 20 s	21 cycles
72	5 min	

3. Pool PCR reactions into clean microcentrifuge tube and precipitate with 0.1 V sodium acetate (pH 5.2) and 2.5 V of ice-cold absolute ethanol. Incubate at −80 °C for 30 min. Centrifuge at 16,000 × *g* and 4 °C for 30 min. Remove the supernatant and air dry the pellet. Resuspend in 20 μL of MillQ water.
4. Prepare a 2% Metaphor agarose gel using TBE buffer (*see* **Note 9**).
5. Add 4 μL of 6× DNA gel loading dye to precipitated sample and load the entire volume onto the prepared 2% Metaphor agarose gel.

6. Run the gel at 100 V for approximately 1.5 h or until the bromophenol blue dye front reaches 2–3 cm from the bottom of the gel.
7. Image the gel. We use a Typhoon FLA9500 laser scanner (GE Life sciences) for increased sensitivity and print the gel images at 1:1 scaling. The cDNA libraries from Hfq-HTF appear as a smear running above primer dimers that should be apparent in the negative control samples.
8. Excise the libraries using a sterile scalpel by cutting from the bottom of the smear to the top. Transfer the gel slices to 1.5 mL microcentrifuge tubes.
9. Add 750 μL of Buffer QG from the MinElute PCR purification kit (QIAgen) and incubate the gel slices at 42 °C for 30 min to dissolve the agarose. The gel slice should be less than or equal to 0.4 g. If you have more than that, separate the gel slices over two tubes.
10. Transfer the volume to a MinElute column fitted to collection tubes and spin for 1 min at 16,000 × *g* (*see* **Note 10**). Discard the flowthrough.
11. Add 500 μL of QG buffer and spin through the column.
12. Add 750 μL of Buffer PE (QIAgen) to the columns and incubate for 10 min at room temperature.
13. Spin the wash out at 16,000 × *g* for 1 min at room temperature and discard the flowthrough.
14. Dry the columns by spinning at 16,000 × *g* for 2 min at room temperature. Transfer the columns to clean 1.5 mL microcentrifuge tubes.
15. Add 20 μL of nuclease-free MilliQ water to the column and incubate at room temperature for 5 min.
16. Elute the purified cDNA by spinning at 16,000 × *g* for 2 min at room temperature.
17. Quantify the cDNA using a Qubit high sensitivity DNA assay kit and fluorometer.
18. The samples can be submitted for single-end sequencing on Illumina MiSeq, HiSeq, NextSeq platforms.

3.2 Analysis of CRAC Datasets

A generally accepted analysis pipeline for the many RNA–protein UV-crosslinking protocols (CRAC, CLIP-Seq, HITS-CLIP, iCLIP, eCLIP, PAR-CLIP) has yet to emerge and most labs use custom scripts and software packages to process the sequencing data. Here, we focus on the pipeline we currently use for bacterial CRAC datasets and highlight some alternative approaches where appropriate (*see* reference [25] for a detailed review of alternative methodologies).

Controls and replicates. The inclusion of negative control experiments is required to identify background signal, usually from abundant RNA species. We include the parental (untagged) strain *E. coli* O157:H7 str. Sakai as a negative control. These controls generate very low complexity libraries and pose some problems for normalization between background and signal samples. Control samples generated without UV crosslinking are often used in CLIP-Seq experiments and may be used to establish dependence on covalent crosslinking, although we note that under native purification conditions Hfq–RNA complexes are readily purified without any crosslinking. Size matched input (SMinput) controls from pre-immunoprecipitated total RNA have been used to generate high complexity control libraries that can be used for background estimation and have been shown to increase enrichment of true positive binding sites in eCLIP datasets [26]. We minimally process biological duplicate experiments with duplicate negative control samples, which allows us to identify peaks that are present in replicate datasets and absent from controls. For differential analysis of peak height between signal and background samples, a minimum of biological triplicates for experimental and control samples are required.

Read depth. The read depth required for sufficient coverage of binding sites will depend on the number of RBP-binding sites and complexity of the library generated (i.e., number of PCR duplicates). Hfq binds ~20% of the transcriptome and we have generated usable CRAC data from as little as 4.2 M reads. We generally aim to generate 17–35 nt trimmed RNA fragments that contain enough sequence for a unique alignment, and short enough to limit the sequence space for the protein interaction. We routinely use Illumina 50 bp single-end sequencing that is long enough to sequence into the 3′ adapter sequence. However, paired-end sequencing can be useful to identify high-confidence mutations that are introduced during the cDNA library preparation step. These crosslinking mutations can be used to map the exact protein crosslinking sites [19].

Data analysis. Our data analysis can be broken into six stages: (1) adapter trimming and quality filtering, (2) demultiplexing samples, (3) aligning reads to a genomic template, (4) counting reads overlapping genomic features, (5) identification of binding sites, and (6) statistical analysis of peaks. We use the pyCRAC software package (https://bitbucket.org/sgrann/pycrac) for many of these processing steps [27].

3.2.1 Adapter Trimming and Quality Filtering

Flexbar (https://github.com/seqan/flexbar) is used to trim 3′ adapter sequences and allows quality filtering of the reads [28].

```
flexbar -r input.fastq -qf i1.8 -t trimmed.fastq -n 10 -
    ao 7 -as TGGAATTCTCGGGTGCCAAGG -qt 30
```

where input.fastq and trimmed.fastq are the input and output fastq file names, respectively.

3.2.2 Demultiplexing Samples Using 5′ Linker Barcodes

The 5′ barcode sequence introduced by the L5Ax series of 5′ linkers (Table 1) allows sample multiplexing and also incorporates three random nucleotides for accurate quantification of PCR duplicates (i.e., reads with the same start and end positions that arise from PCR duplication of a single cDNA rather than independent linker ligation events). We use pyBarcodeFilter.py (version 2.3.4) from the pyCRAC package to demultiplex the samples [27]. This script additionally recognizes the random nucleotides within the barcode sequence and appends the random sequence to the read header for downstream processing.

```
pyBarcodeFilter.py -f trimmed.fastq -b barcodes.txt -m 1
```

where barcodes.txt is a file containing a tab delimited list of barcodes and sample names. The "–m" flag indicates the number of mismatches allowed in the barcode sequence part of the read during demultiplexing.

We remove PCR duplicates from our datasets using pyFastqDuplicateRemover.py (version 0.0.2), also from the pyCRAC package.

```
pyFastqDuplicateRemover.py -f demultiplexed.fastq -o collapsed.fasta
```

where demultiplexed.fastq and collapsed.fasta are the input and output file names, respectively.

3.2.3 Aligning Reads to a Genomic Template

A number of tools are available to align short read data to genomic templates. We use novoalign (version 3.04.06) as it tolerates mutations well, and these are introduced during the CRAC protocol. SAM format files generated by bowtie2 are also compatible with the downstream processing steps presented here (with some additional flags).

```
novoalign -d novo.index -f collapsed.fasta -r Random > collapsed.novo
```

where novo.index is the genome-specific index file generated by novoindex, and collapsed.novo is the output file name. The "-r Random" flag ensures that reads that map to multiple genomic locations are randomly distributed to all possible mapping locations.

3.2.4 Counting Reads Overlapping Genomic Features

To convert aligned reads into read density across the genome, we use the pyCRAC script pyReadCounters.py (version 0.5.3). The output is a GTF format file that can be used as input for a number of analysis scripts within the package. The script also generates a "hit table" of reads counts mapping to each genomic feature that is useful as a first glance at dataset for enriched RNAs.

```
pyReadCounters.py -f collapsed.novo --gtf annotation.gtf -v --rpkm -o readcounters_output
```

where annotation.gtf is a GTF format file of the genome annotation and readcounters_output is the file name for pyReadCounters output files (*see* **Note 11**). The "--rpkm" flag adds an extra table with the number of reads per kilobase transcript per million reads for each genomic feature.

At this stage, we also generate plots of read density across the transcriptome using pyGTF2bedGraph.py (version 0.0.3). The BEDGRAPH format files can be normalized to reads per million and visualized using software like the integrated genome browser (http://bioviz.org/igb/).

```
pyGTF2bedGraph.py --gtf readcounters_output.gtf --count
    -v --permillion -o bedgraph_output
```

where readcounters_output.gtf is the GTF format output from pyReadCounters.py and bedgraph_output is the output file name for pyGTF2bedGraph.py.

3.2.5 Identification of Binding Sites

A number of approaches are available for identifying binding sites in CRAC and CLIP-Seq datasets [29]. We currently use two approaches; both look for read clusters (groups of overlapping reads) and require the user to empirically determine thresholds for read overlap, read density, or minimum cluster height.

```
pyClusterReads.py -f readcounters_output.gtf -r 100
    --cic X --co Y -o cluster_output
```

where –r is the amount of flanking sequence surrounding each genomic feature, X is the threshold for number of reads within a clusters, and Y is minimum overlap for a read to be included in the cluster.

More recently we have adopted the peak calling approached described by Holmqvist et al. [20, 29] with good results. Read clusters are identified using blockbuster [30] and read peaks identified by merging overlapping read clusters into a single feature. The blockbuster pipeline requires a BED file format input of aligned reads and this can be generated from the pyReadCounters.py GTF output.

```
pyGTF2bed.py --gtf read_counters.gtf -o read_counters.
    bed
```

where read_counters.bed is the output file name. The BED format file can be used as an input for the pipeline described previously by Holmqvist et al. [20].

3.2.6 Statistical Analysis of Peak Enrichment

After identification of read clusters in both experimental and control samples, we look to define those that are statistically significant when compared to background. We use the pyCRAC script pyCalculateFDR.py that heuristically determines the probability of recovering a given peak height by randomly redistributing reads within a genomic feature and reporting the maximum random peak height.

The process is iterated to determine the distribution of random peak heights, which is used to calculate the probability of recovering the observed peak height. We then look for peaks (read clusters) that are present in replicate experiments and absent from background samples. The advantage of this approach is that it can generate false discovery rates from a single experimental sample. This approach also avoids the need to normalize between experimental and backgrounds samples (e.g., by median centering) that may amplify the background signal and obscure true positives. High complexity control libraries can be generated from a fraction of the input material and may allow more robust identification of true positive binding sites when using approaches that required normalization between experimental and background samples [26].

The FindSignificantPeaks_pipeline.py pipeline that performs the below steps is available at https://bitbucket.org/sgrann/

```
pyCalculateFDR.py -f readcounters_output.gtf --gtf an-
    notation.gtf -c genome.length -r 100 -o FDR_output -m
    0.05 --min=10 --iterations=500
```

where:

genome.length is a tab delimited file containing the chromosome name and length

-r is the length of sequence on either side of a feature to include

-is the output file name

-m is the FDR threshold for reporting peaks

--min is the minimum read coverage

--iterations is the number of times the reads are randomly distributed over each genomic feature.

The GTF format output from pyCalculateFDR reports regions of statistically significant peaks and can be used to identify statistically significant read clusters from pyClusterReads.py and the blockbuster pipeline using intersectBed from the bedtools package.

In some cases the interval lengths reported by pyCalculateFDR.py can be less than 5 nucleotides, which is not useful for downstream analyses, such as motif analyses. In these cases we recommend to using pyNormalizeIntervalLengths.py to extend the length of the short intervals. We set a minimum interval length of 20 nucleotides:

```
pyNormalizeIntervalLengths.py -f FDR_output.gtf -o FDR_
    output_min20.gtf
```

Next we use intersectBed from the bedtools suite (REF) on each sample allowing to collate data between samples:

```
intersectBed -s -u -a cluster_output -b FDR_output_
    min20.gtf > clusters_FDR.gtf
```

Next we merge overlapping peaks within each sample using mergeBed from the bedtools package, and append the sample name to each peak to facilitate peak counting between samples.

```
sort -k1,1 -k4n,4 clusters_FDR.gtf > clusters_FDR_sorted.gtf

mergeBed -s -d -1 –i clusters_FDR_sorted.gtf > clusters_FDR_merged.temp

awk '{print $1"\t"$2"\t"$3"\t"FILENAME"\t1\t"$4}' clusters_FDR_merged.temp > clusters_FDR_merged.bed
```

The number of overlapping statistically significant peaks found in the experimental or background samples can then be counted and the source (signal or background) added to each feature and output in BED format.

```
cat *signal_cluster_FDR_merged.bed | sort –k1,1 –k2n,2 > all_sig_cluster_FDR_merged.bed

cat *bkg_cluster_FDR_merged.bed | sort –k1,1 –k2n,2 > all_bkg_cluster_FDR_merged.bed

mergeBed -s -d -1 -c 4 -o count_distinct –i all_sig_cluster_FDR_merged.bed | awk '{print $1"\t"$2"\t"$3"\tsignal_count\t"$5"\t"$4}' > signal_merged.bed

mergeBed -s -d -1 -c 4 -o count_distinct –i all_bkg_cluster_FDR_merged.bed | awk '{print $1"\t"$2"\t"$3"\tbackground_count\t"$5"\t"$4}' > bkg_merged.bed
```

where *cluster_FDR_merged.bed lists all of the clusters_FDR_merged.bed files generated for experimental or background samples. Finally, we tabulate overlapping, statistically significant peaks found in the experimental and background samples using mergeBed.

```
cat signal_merged.bed bkg_merged.bed | sort -k1,1 -k2n,2 > all_samples_merged.bed

mergeBed -s -d -1 -c 4,5 -o collapse –i all_samples_merged.bed > peaks_FDR_count.bed

FDRbed2table.py peaks_FDR_count.bed X Y > peaks_FDR_count.table
```

where X and Y are the number of experimental and background samples, respectively. We consider statistically significant peaks that are found in at least replicate experimental samples, and are absent from all background samples, to be valid protein-binding sites.

Once statistically significant peaks have been defined, these regions can be analyzed for enriched sequence motifs and RNA structures that will allow extrapolation of the protein–RNA interaction data.

4 Notes

1. The concentration of RNace IT used to footprint (trim) RNAs onto protein of interest is determined empirically. Ideally, the reads will be long enough to map uniquely (~17 nt) but short enough to give good resolution of the protein-binding site. We aim to generate and average RNA length of ~35 nt.
2. As of February 2016, the sequences introduced by the miRCat-33 3′ linker, RT_oligo, and PE_miRCat_33_PCR primers no longer allow paired-end sequencing on Illumina platforms. Single end read sequencing can still be performed on libraries generated using the 5′ and 3′ sequences listed in Table 1.
3. Ensure that the elution flowthrough is radioactive. If after two elutions the flowthrough remains only slightly radioactive, more elutions can be added as desired.
4. If the TCA pellet is very large, this means that some salt has precipitated during the TCA step. Should this be the case, resuspend the pellet in acetone before spinning.
5. For larger complexes, we find that transferring overnight at 30 V gives better transfer.
6. Using the wet transfer system, we usually get about an 80% transfer efficiency. It is important to check whether the transfer has been completed as you could otherwise lose a lot of your material.
7. First incision can be made just above the band corresponding to the protein of interest—this should represent the protein bound to a few nucleotides of RNA. Once membrane fragments have been cut out, they can be stored overnight at −20°C.
8. The number of cycles used to amplify cDNA libraries should be optimized for the template to avoid artifacts due to overamplification. Generally, 21 cycles has been sufficient to produce complex libraries from cDNA generated from Hfq-HTF; however, we often adjust the cycles between 19 and 24 cycles and will increase number of independent PCR reactions (up to 10) for samples with low abundance cDNA.
9. Cooling the Metaphor agarose-TBE buffer mixture for 15 min on ice prior to melting the agarose in the microwave prevents foaming of the mixture during heating. Similarly, cooling the Metaphor gel in a caster at 4 °C for at least 30 min prior to use helps with handling of the fragile gel after casting.
10. For samples with >700 μL of buffer/agarose multiple of sample loading should be performed to bind the entire sample to the column matrix.

11. A GTF format file for genome annotation is required by the pyCRAC software and is critical to the interpretation of the output of the pyCRAC pipeline (i.e., the genomic features to which the significant clusters map). pyCRAC is sensitive to the formatting within the GTF file and we find it useful to check the annotated GTF file using the pyCheckGTFfile.py command to ensure that the GTF file is suitable for use with the pyCRAC software.

References

1. Gama-Castro S, Salgado H, Santos-Zavaleta A et al (2016) RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. Nucleic Acids Res 44:D133–D143
2. Ishihama A (2010) Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. FEMS Microbiol Rev 34:628–645
3. Waters LS, Storz G (2009) Regulatory RNAs in bacteria. Cell 136:615–628
4. Melamed S, Peer A, Faigenbaum-Romm R et al (2016) Global mapping of small RNA-target interactions in bacteria. Mol Cell 63:884–897
5. Bouvier M, Sharma CM, Mika F et al (2008) Small RNA binding to 5′ mRNA coding region inhibits translational initiation. Mol Cell 32:827–837
6. Storz G, Vogel J, Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 43:880–891
7. Smirnov A, Förstner KU, Holmqvist E et al (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. Proc Natl Acad Sci U S A 113:11591–11596
8. Smirnov A, Wang C, Drewry LL et al (2017) Molecular mechanism of mRNA repression in *trans* by a ProQ-dependent small RNA. EMBO J 36:1029–1045
9. Bandyra KJ, Said N, Pfeiffer V et al (2012) The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. Mol Cell 47:943–953
10. Soper T, Mandin P, Majdalani N et al (2010) Positive regulation by small RNAs and the role of Hfq. Proc Natl Acad Sci 107:9602–9607
11. Papenfort K, Sun Y, Miyakoshi M et al (2013) Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. Cell 153:426–437
12. Tree JJ, Granneman S, McAteer SP et al (2014) Identification of bacteriophage-encoded anti-sRNAs in pathogenic Escherichia coli. Mol Cell 55:199–213
13. Miyakoshi M, Chao Y, Vogel J (2015) Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. EMBO J 34:1478–1492
14. Bossi L, Schwartz A, Guillemardet B et al (2012) A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. Genes Dev 26:1864–1873
15. Rabhi M, Espéli O, Schwartz A et al (2011) The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. EMBO J 30:2805–2816
16. Sedlyarova N, Shamovsky I, Bharati BK et al (2016) sRNA-mediated control of transcription termination in E. coli. Cell 167:111–121.e13
17. Winther K, Tree JJ, Tollervey D et al (2016) VapCs of Mycobacterium tuberculosis cleave RNAs essential for translation. Nucleic Acids Res 44:9860–9871
18. van Nues R, Schweikert G, de Leau E et al (2017) Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress. Nat Commun 8:12
19. Granneman S, Kudla G, Petfalski E et al (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc Natl Acad Sci U S A 106:9613–9618
20. Holmqvist E, Wright PR, Li L et al (2016) Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. EMBO J 35:e201593360
21. Chao Y, Li L, Girodat D et al (2017) In vivo cleavage map illuminates the central role of RNase E in coding and noncoding RNA pathways. Mol Cell 65:39–51
22. Waters SA, McAteer SP, Kudla G et al (2017) Small RNA interactome of pathogenic E. coli

revealed through crosslinking of RNase E. EMBO J 36:374–387

23. Lalaouna D, Carrier MC, Semsey S et al (2015) A 3′ external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. Mol Cell 58: 393–405
24. Han K, Tjaden B, Lory S (2016) GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation. Nat Microbiol 16239:1–10
25. Dahan S, Knutton S, Shaw RK et al (2004) Transcriptome of enterohemorrhagic *Escherichia coli* O157 adhering to eukaryotic plasma membranes. Infect Immun 72:5452–5459
26. Van Nostrand EL, Pratt GA, Shishkin AA et al (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods 13:508–514
27. Webb S, Hector RD, Kudla G et al (2014) PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. Genome Biol 15:R8
28. Dodt M, Roehr J, Ahmed R et al (2012) FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. Biology 1:895–905
29. Uhl M, Houwaart T, Corrado G et al (2016) Computational analysis of CLIP-seq data. Methods 118–119:60–72
30. Langenberger D, Bermudez-Santana C, Hertel J et al (2009) Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics 25:2298–2301

Chapter 16

Producing Hfq/Sm Proteins and sRNAs for Structural and Biophysical Studies of Ribonucleoprotein Assembly

Kimberly A. Stanek and Cameron Mura

Abstract

Hfq is a bacterial RNA-binding protein that plays key roles in the post-transcriptional regulation of gene expression. Like other Sm proteins, Hfq assembles into toroidal discs that bind RNAs with varying affinities and degrees of sequence specificity. By simultaneously binding to a regulatory small RNA (sRNA) and an mRNA target, Hfq hexamers facilitate productive RNA•••RNA interactions; the generic nature of this chaperone-like functionality makes Hfq a hub in many sRNA-based regulatory networks. That Hfq is crucial in diverse cellular pathways—including stress response, quorum sensing, and biofilm formation—has motivated genetic and "RNAomic" studies of its function and physiology (in vivo), as well as biochemical and structural analyses of Hfq•••RNA interactions (in vitro). Indeed, crystallographic and biophysical studies first established Hfq as a member of the phylogenetically conserved Sm superfamily. Crystallography and other biophysical methodologies enable the RNA-binding properties of Hfq to be elucidated in atomic detail, but such approaches have stringent sample requirements, viz.: reconstituting and characterizing an Hfq·RNA complex requires ample quantities of well-behaved (sufficient purity, homogeneity) specimens of Hfq and RNA (sRNA, mRNA fragments, short oligoribonucleotides, or even single nucleotides). The production of such materials is covered in this chapter, with a particular focus on recombinant Hfq proteins for crystallization experiments.

Key words Hfq, Sm, sRNA, RNA chaperone, RNA-binding protein, Crystallization, In vitro transcription

Abbreviations

3D	Three-dimensional
AU	Asymmetric unit
CV	Column volume
DEPC	Diethyl pyrocarbonate
HDV	Hepatitis δ virus
HDVD	Hanging-drop vapor diffusion
IMAC	Immobilized metal affinity chromatography
MW	Molecular weight
MWCO	Molecular weight cut-off

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_16,

nt	Nucleotide
PDB	Protein Data Bank
RNP	Ribonucleoprotein
RT	Room temperature
SDVD	Sitting-drop vapor diffusion

1 Introduction

The bacterial protein Hfq, initially identified as a host factor required for the replication of bacteriophage Qβ RNA [1], plays a central role in RNA biology: both in RNA-based regulation of gene expression and in modulating RNA stability and lifetime in vivo [2]. Hfq functions broadly as a chaperone, facilitating contacts between small noncoding RNAs (sRNAs) and their cognate mRNAs [3]. The RNA interactions may either stimulate or inhibit expression, depending on the identity of the mRNA–sRNA pair and the molecular nature of the interaction (high or low affinity, stable or transient, etc.) [4, 5]. In many cases, Hfq is required for these pairings to be effective [6], and knockdown of the *hfq* gene results in pleiotropic phenotypes such as increased UV sensitivity, greater susceptibility to oxidative or osmotic stress, and decreased growth rates [7]. A flood of "RNAomics"-type studies, over the past decade, has shaped what we know about Hfq-associated RNAs [2, 8, 9]. Hfq has been linked to many cellular pathways that rely on rapid responses at the level of post-transcriptional/mRNA regulation, including stress responses [10–12], quorum sensing [13], biofilm formation [14], and virulence factor expression [12, 15].

Hfq homologs are typically ≈80–100 amino acids in length, with the residues folding as an α-helix followed by five β-strands arranged into a highly bent, antiparallel β-sheet [16, 17]. Hfq monomers self-assemble into a toroidal hexamer, the surface of which features at least three distinct regions that can bind RNA. The *proximal* face of the hexamer (proximal with respect to the *N*′-terminal α-helix) is known to bind U-rich sequences [16, 18], while the *distal* face of the $(Hfq)_6$ ring binds preferentially to A-rich RNA elements [19, 20]. Recently a third, lower-affinity, *lateral* surface on the outer rim of the Hfq ring has been shown to bind RNA [21] and aid in sRNA•••mRNA annealing [22]. This lateral site likely has a preference for U-rich segments [23], but also may interact fairly nonspecifically with RNA because of an arginine-rich region that is found in some homologs. While the exact mechanism by which Hfq facilitates productive RNA•••RNA interactions remains unclear, it is thought that the distal face binds the 3′-poly(A) tails of mRNAs while the proximal face binds to 3′ U-rich regions of sRNAs [3]. The lateral surface may act either to cycle different RNAs onto Hfq [22] or as an additional surface for binding to internal, U-rich regions of sRNAs. A recent study suggested that this mechanistic model holds for only a subset of

sRNAs, termed "Class I" sRNAs [24]. A second subset of sRNAs ("Class II") appears to bind both the proximal and distal sites of Hfq; the mRNA targets of these Class II sRNAs are predicted to bind preferentially to the lateral region of the Hfq ring.

A detailed understanding of how different sRNAs interact with Hfq, and with target RNAs in a ternary RNA·Hfq·RNA complex, requires atomic-resolution structural data. While multiple structures have been determined for short ($\lesssim$10-nucleotide) RNAs bound to either the proximal [16, 18], distal [19, 20], or lateral [23] sites of Hfq (Table 1), as of this writing only one structure of an Hfq bound to a full-length sRNA has been reported [25]. In that Hfq·RNA complex, comprised of *E. coli* Hfq bound to the *Salmonella* RydC sRNA (Fig. 1a), the 3′-end of the sRNA encircles the pore, toward the proximal face of the hexamer, while an internal U–U dinucleotide binds in one of the six lateral pockets on the periphery of the Hfq ring (Fig. 1b). Though other regions of the sRNA were found to further contact a neighboring Hfq ring in the lattice (Fig. 1c), the stoichiometry of the Hfq·RydC complex in vivo, at limiting RNA concentrations, is thought to be 1:1. (Interestingly, two distinct interaction/binding modes were seen between RydC and the lateral rim of an adjacent hexamer [Fig. 1c].) The Hfq·RydC complex offers a valuable window into our understanding of Hfq•••sRNA interactions, limited mainly by the relatively low resolution (3.48 Å) of the refined structure. For this and other Hfq·RNA complexes, many questions can be addressed by leveraging different types of structural and biophysical approaches. Ideally, the methods used would provide a variety of complementary types of information (i.e., the underlying strategy in taking a "hybrid methods" approach [26, 27]).

Atomic-resolution information may be obtained in several ways. Historically, the premier methodologies have been X-ray crystallography and solution NMR spectroscopy; these well-established approaches are described in many texts, such as [28, 29]. Though beyond the scope of this chapter, note that much progress in recent years has positioned electron cryo-microscopy (cryo-EM) as a powerful methodology for high-resolution (nearly atomic) structural studies of macromolecular assemblies [30, 31], including ribonucleoprotein (RNP) complexes such as the ribosome [32–34], telomerase [35], and, most recently, the spliceosome [36, 37]. Thus far, all Hfq and Hfq·RNA structures deposited in the Protein Data Bank (PDB), listed in Table 1, have been determined via X-ray crystallography. The molecular weight (MW) of a typical Hfq hexamer is ≈60 kDa while sRNAs, which range in length from ≈ 50 to 500 nucleotides (nt), have MWs of ≈16–1600 kDa. An RNP complex of this size is ideally suited to macromolecular crystallography.

In this chapter, we describe how to prepare and crystallize Hfq·sRNA complexes for structure determination and analysis via

Table 1

A comprehensive list of Hfq structures in the PDB, including co-crystal structures with nucleotides and RNAs

PDB ID (year)	Space-group	d_{min} (Å)	Solvent content (by vol) (%)	Macromolecular complex [species]; other notes	Crystallization information (format, precipitants, other solution conditions/ notes)	Citation
1KQ1 (2002)	$P2_1$	1.55	35.9	$(Hfq)_6$ hexamer [*Staphylococcus aureus*]	HDVD at 298 K; pH 4.6; $(NH_4)_2SO_4$, NaOAc	[16]
1KQ2 (2002)	$C222_1$	2.71	43.3	$(Hfq)_6$·r(AU_5G) RNA, bound at proximal site; [*S. aureus*]	HDVD at 298 K; pH 7.5; HEPES, PEG-550, $MgCl_2$, KCl	[16]
1HK9 (2003)	$P6_1$	2.15	33.0	$(Hfq)_6$ hexamer [*Escherichia coli*]	SDVD at 293 K; pH 4.6; 25% PEG-4000, 0.2 M NH_4-OAc, 0.2 M NaOAc	[68]
1U1S (2005)	$P2_12_12_1$	1.6	50.0	$(Hfq)_6$ hexamer [*Pseudomonas aeruginosa*]	HDVD at 295 K; pH 8.5; 200 mM NH_4Cl, 12% PEG-4000, 50 mM Tris–HCl, 5 mM $CdCl_2$	[53]
1U1T (2005)	$P2_12_12_1$	1.9	51.7	$(Hfq)_6$ hexamer [*P. aeruginosa*]	HDVD at 295 K; pH 6.5; 100 mM MES, 0.6 M $(NH_4)_2SO_4$, 1 M Li_2SO_4	[53]
2QTX (2007)	$P2_1$	2.5	45.3	$(Hfq)_6$ hexamer [*Methanococcus jannaschii*]	SDVD at 277 K; pH 8.5; 0.1 M Tris, 0.2 M NH_4OAc, 25% PEG-3350	[69]
3GIB (2009)	$P2_12_12$	2.4	45.0	$(Hfq)_6$·r($A)_9$ RNA, bound to distal site; [*E. coli*]	HDVD at 298 K; pH 9.5; 0.1 M CHES, 40% v/v MPD	[19]
3HFN (2009)	$P3$	2.3	42.7	$(Hfq)_6$ hexamer [*Anabaena sp. PCC 7120*]	SDVD at 277 K; pH 3.5; 0.1 citric acid, 2 M $(NH_4)_2SO_4$	[70]
3HFO (2009)	$F222$	1.3	36.1	$(Hfq)_6$ hexamer [*Synechocystis sp. PCC 6803*]	SDVD at 292 K; pH 7; 60% Tacsimate (*see* **Note 10**)	[70]
3INZ (2010)	$P2_12_12_1$	1.7	41.2	$(Hfq)_6$ hexamer, H57T mutant; [*P. aeruginosa*]	VD at 295 K; pH 8.5; 50 mM NaCl, 100 mM NH_4Cl, 7.5% PEG-MME 550, 50 mM TrisCl, 10 mM $CdCl_2$	[71]

(continued)

Table 1
(continued)

PDB ID (year)	Space-group	d_{min} (Å)	Solvent content (by vol) (%)	Macromolecular complex [species]; other notes	Crystallization information (format, precipitants, other solution conditions/ notes)	Citation
3M4G (2010)	*P*1	2.05	41.0	$(Hfq)_6$ hexamer, H57A mutant; [*P. aeruginosa*]	HDVD at 293 K; pH 8.5; 50 mM NaCl, 100 mM NH_4Cl, 7.5% PEG-MME 550, 50 mM TrisCl, 10 mM $ZnCl_2$	[71]
2Y90 (2011)	*P*6	2.25	30.0	$(Hfq)_6$ hexamer [*E. coli*]	HDVD at 293 K; pH 8; 0.1 M Tris, 1.6 M $(NH_4)_2SO_4$	[72]
2YHT (2011)	*P*1	2.9	30.0	$(Hfq)_6$ hexamer [*E. coli*]	SDVD at 293 K; pH 5.4; 0.1 M Na^+citrate, 30% PEG-3350	[72]
2YLB (2011)	$P6_1$	1.15	39.0	$(Hfq)_6$ hexamer [*Salmonella typhimurium*]	pH 7; 0.1 M HEPES, 0.5% Jeffamine, 1.1 M malonate	[18]
2YLC (2011)	*P*6	1.3	40.0	$(Hfq)_6$·r$(U)_6$ RNA, bound at proximal pore; [*S. typhimurium*]	pH 8; 0.2 M NaSCN, 20% PEG-3350	[18]
3AHU (2011)	*F*222	2.2	42.0	$(Hfq)_6$·r$(AG)_3$A RNA (SELEX-derived aptamer), bound to distal site; [*Bacillus subtilis*]	HDVD at 293 K; pH 6.5; 0.2 M MES, 1.8 M $(NH_4)_2SO_4$, 0.01 M $CoCl_2$	[73]
3HSB (2011)	*I*422	2.2	39.6	$(Hfq)_6$·r$(AG)_3$A RNA (SELEX-derived aptamer), bound to distal site; [*B. subtilis*]	HDVD at 293 K; pH 6.5; 0.1 M MES, 1.8 M $(NH_4)_2SO_4$, 0.015 M $CoCl_2$	[73]
3QHS (2011)	*P*1	2.85	37.7	$(Hfq)_6$ hexamer; full-length protein; [*E. coli*]	HDVD at 277 K; pH 6.5; 0.1 M Bis-Tris, 30% v/v PEG-MME 550, 0.05 M $CaCl_2$	[74]
3RER (2011)	*P*1	1.7	43.7	$(Hfq)_6$·r(AU_6A) RNA·ADP [*E. coli*]	HDVD at 283 K; pH 6.2; 0.1 M cacodylate, 100 mM NaCl, 12% w/v PEG-8000	[75]
3RES (2011)	*I*2	2.0	42.6	$(Hfq)_6$·ADP [*E. coli*]	HDVD at 283 K; pH 4.2; 200 mM NH_4OAc, 100 mM NaOAc, 22% w/v PEG-4000	[75]

(continued)

Table 1
(continued)

PDB ID (year)	Space-group	d_{min} (Å)	Solvent content (by vol) (%)	Macromolecular complex [species]; other notes	Crystallization information (format, precipitants, other solution conditions/ notes)	Citation
3SB2 (2011)	$P2_1$	2.63	43.0	$(Hfq)_6$ hexamer [*Herbaspirillum seropedicae*]	SDVD at 291 K; pH 7.0; 0.1 M PCB (Na-propionate, Na-cacodylate, Bis-tris propane), 25% w/v PEG-1500	[54]
3QO3 (2012)	*C*2	2.15	48.3	$(Hfq)_6$·ATP [*E. coli*]	SDVD at 295 K; pH 7.5; 0.1 M HEPES, 10% w/v PEG-8000, 8% v/v ethylene glycol	[76]
3QSU (2012)	*P*3	2.2	35.9	$(Hfq)_6$·r(A)$_7$ RNA, bound to distal site; [*S. aureus*]	HDVD at 298 K; pH 6.5; 0.1 M Na-cacodylate, 12% v/v MPD, 0.2 M $Zn(OAc)_2$, 0.1 M KCl	[20]
3QUI (2013)	$P2_12_12_1$	1.93	37.9	$(Hfq)_6$·{ADP, AMP-PNP} (*see* **Note 11**) [*P. aeruginosa*]	HDVD at 295 K; pH 8; 0.2 M $(NH_4)_2SO_4$, 0.2 M NaCl, 50 mM TrisCl	[77]
3VU3 (2013)	*I*222	2.85	58.9	$(Hfq)_6$·catalase HPII (*see* **Note 12**) [*E. coli*]	HDVD at 293 K; pH 9; 0.1 M TrisCl, 0.18 M NaCl, 10% w/v PEG-4000	[78]
4HT8 (2013)	*C*2	1.9	43.0	$(Hfq)_6$·r(A)$_7$ RNA [*E. coli*]	HDVD at 283 K; pH 7.9; 200 mM NH_4OAc, 100 mM Tris, 26% v/v isopropanol	[79]
4HT9 (2013)	*I*2	1.8	33.0	$(Hfq)_6$·r(A)$_7$·r(AU$_6$A) RNAs [*E. coli*]	HDVD at 283 K; pH 6.2; 0.1 M cacodylate, 0.1 M NaCl, 12% w/v PEG-8000	[79]
4J5Y (2013)	$P2_12_12_1$	2.1	44.8	$(Hfq)_6$·ATP [*P. aeruginosa*]	HDVD at 295 K; pH 8; 0.2 M $(NH_4)_2SO_4$, 0.2 M NaCl, 50 mM TrisCl	[77]
4J6W (2013)	$P2_12_12_1$	1.8	46.3	$(Hfq)_6$·CTP [*P. aeruginosa*]	HDVD at 295 K; pH 8; 0.2 M $(NH_4)_2SO_4$, 0.2 M NaCl, 50 mM TrisCl	[77]

(continued)

Table 1
(continued)

PDB ID (year)	Space-group	d_{min} (Å)	Solvent content (by vol) (%)	Macromolecular complex [species]; other notes	Crystallization information (format, precipitants, other solution conditions/ notes)	Citation
4J6X (2013)	$P2_12_12_1$	2.22	44.9	$(Hfq)_6$·UTP [*P. aeruginosa*]	HDVD at 295 K; pH 8; 0.2 M $(NH_4)_2SO_4$, 0.2 M NaCl, 50 mM TrisCl	[77]
4J6Y (2013)	$P2_12_12_1$	2.14	44.1	$(Hfq)_6$; GTP not found in density; [*P. aeruginosa*]	HDVD at 295 K; pH 8; 0.2 M $(NH_4)_2SO_4$, 0.2 M NaCl, 50 mM TrisCl, 4.75 mM GTP	[77]
4JLI (2014)	*H3*	1.79	35.7	$(Hfq)_6$ hexamer, F42 W mutant; [*E. coli*]	HDVD; pH 8.0–9.0; 0.1 M Tris, 22–28% PEG-3350, 26–32% isopropanol	[80]
4JRI (2014)	$P2_1$	1.83	38.0	$(Hfq)_6$ hexamer, F39 W mutant; [*E. coli*]	HDVD; pH 8.0–9.0; 0.1 M Tris, 22–28% PEG-3350, 26–32% isopropanol	[80]
4JRK (2014)	$P22_12_1$	1.89	39.9	$(Hfq)_6$ hexamer, F11 W mutant; [*E. coli*]	HDVD; pH 8.0–9.0; 0.1 M Tris, 22–28% PEG-3350, 26–32% isopropanol	[80]
4JUV (2014)	$P2_1$	2.19	40.2	$(Hfq)_6$ hexamer, Y25W mutant; [*E. coli*]	HDVD at 295 K; pH 8.0–9.0; 0.1 M Tris, 22–28% PEG-3350, 26–32% isopropanol	[80]
4MMK (2014)	$P2_1$	2.16	40.3	$(Hfq)_6$ hexamer, Q8A mutant; [*P. aeruginosa*]	HDVD at 303 K; pH 6.5; 50 mM TrisCl, 7% w/v PEG-2000 MME, 2% v/v MPD, 20 μM $ZnCl_2$	[81]
4MML (2014)	*P6*	1.8	36.5	$(Hfq)_6$ hexamer, D40A mutant; [*P. aeruginosa*]	HDVD at 303 K; pH 6.5; 50 mM TrisCl, 7% w/v PEG-2000 MME, 2% v/v MPD, 20 μM $ZnCl_2$	[81]
4NL2 (2014)	$P2_122_1$	2.6	43.1	$(Hfq)_6$ hexamer [*Listeria monocytogenes*]	HDVD at 298 K; pH 7.5; 0.1 M HEPES, 40% 1,2-propanediol	[82]

(continued)

Table 1
(continued)

PDB ID (year)	Space-group	d_{min} (Å)	Solvent content (by vol) (%)	Macromolecular complex [species]; other notes	Crystallization information (format, precipitants, other solution conditions/ notes)	Citation
4NL3 (2014)	*C*2	3.1	48.7	$(Hfq)_6$·r$(U)_6$ RNA, bound at proximal pore; [*L. monocytogenes*]	HDVD at 298 K; pH 7.5; 0.1 M HEPES, 40% 1,2-propanediol	[82]
4NOY (2014)	$P2_122_1$	2.8	43.3	$(Hfq)_6$ hexamer, F43 W mutant; [*L. monocytogenes*]	HDVD at 298 K; pH 7.5; 0.1 M HEPES, 40% 1,2-propanediol	[82]
4PNO (2014)	*P*6	0.97	33.4	$(Hfq)_6$·r$(U)_6$ RNA, bound at proximal pore; [*E. coli*]	SDVD at 293 K; pH 8.0; 0.1 M HEPES/NaOH, 12% w/v PEG-3350, 0.25 M KSCN	[83]
4V2S (2014)	$P2_12_12_1$	3.48	55.0	$(Hfq)_6$·RydC sRNA (65 nt) [*E. coli* Hfq, *S. enterica* RydC]	SDVD; pH 6.5; 0.2 M tri-sodium citrate, 0.1 M Na-cacodylate, 15% isopropanol	[25]
4QVC (2015)	$P2_12_12_1$	1.99	49.3	$(Hfq)_6$·r(AUAACUA) RNA [*E. coli*]	HDVD at 281 K; pH 5.5; 0.1 M citrate, 12% w/v PEG-4000	[84]
4QVD (2015)	$P2_12_12_1$	1.97	49.5	$(Hfq)_6$·r(AACUAAA) RNA [*E. coli*]	HDVD at 281 K; pH 7.2; 0.1 M HEPES, 16% w/v MPEG-5000 (*see* **Note 13**)	[84]
4RCB (2015)	*P*6	1.63	38.3	$(Hfq)_6$ hexamer [*E. coli*]	SDVD; 1.6 M $(NH_4)_2SO_4$, 0.5 M LiCl	[85]
4RCC (2015)	*P*6	1.98	38.3	$(Hfq)_6$ hexamer [*E. coli*]	SDVD; pH 8.5; 0.1 M TrisCl, 1.5 M $(NH_4)_2SO_4$, 15% v/v glycerol	[85]
4Y91 (2015)	$P2_12_12_1$	2.66	39.0	$(Hfq)_6$·r$(U)_6$ RNA, bound at proximal pore; [*Thermotoga maritima*]	HDVD at 291 K; pH 8.5; tri-potassium citrate, 30% w/v PEG-3350	n/a
4X9C (2016)	$P2_12_12_1$	1.4	41.7	$(Hfq)_6$ hexamer [*M. jannaschii*]	HDVD at 296 K; pH 8; 0.1 M TrisCl, 50% v/v PEG-200	[86]
4X9D (2016)	$P2_12_12_1$	1.5	41.9	$(Hfq)_6$·UMP [*M. jannaschii*]	HDVD at 296 K; pH 8; 0.1 M TrisCl, 50% v/v PEG-200	[86]

(continued)

Table 1 (continued)

PDB ID (year)	Space-group	d_{min} (Å)	Solvent content (by vol) (%)	Macromolecular complex [species]; other notes	Crystallization information (format, precipitants, other solution conditions/ notes)	Citation
5DY9 (2016)	$P2_1$	1.6	34.9	$(Hfq)_6$·AMP (Y68T mutant of Hfq); [*M. jannaschii*]	HDVD at 296 K; pH 8; 0.1 M TrisCl, 50% v/v PEG-200	[86]
5I21 (2016)	*P*6	1.55	37.0	$(Hfq)_6$ hexamer, Y55W mutant; [*P. aeruginosa*]	HDVD at 303 K; pH 6.5; 50 mM TrisCl, 7% w/v PEG-2000 MME, 2% v/v MPD	n/a
5SZD (2017)	*P*1	1.49	40.2	$(Hfq)_6$ hexamer [*Aquifex aeolicus*]	SDVD at 291 K; pH 5.5; 0.1 M Na-cacodylate, 5% w/v PEG-8000, 35% v/v MPD, 0.1 M $[Co(NH_3)_6]Cl_3$	[23]
5SZE (2017)	*P*6	1.5	46.1	$(Hfq)_6$·r(U)$_6$ RNA, bound at lateral rim; [*A. aeolicus*]	SDVD at 291 K; pH 5.5; 0.1 M Na-cacodylate, 5% w/v PEG-8000, 35% v/v MPD, 1.0 M GndCl	[23]

the classic, single-crystal X-ray diffraction approach. However, if crystallographic efforts with a particular Hfq or Hfq·sRNA complex prove difficult because of a lack of well-diffracting crystals—or, even when such crystals can be reproducibly obtained—then one can also consider investigating the Hfq-based complex via complementary approaches. Two main families of alternative methodological approaches are available: (1) NMR and other spectroscopic methods (e.g., EPR [38]), and (2) cryo-EM and other scattering-based approaches (e.g., SAXS [39]). NMR and cryo-EM are routinely used for smaller or larger-sized biomolecular complexes, respectively, though methodological developments are continuously redefining these limitations. The current upper size limit for de novo NMR structure determination is ≈40 kDa, this limit being reached via the application of techniques such as TROSY, as well as relatively recently developed approaches for deriving distance restraints (e.g., paramagnetic relaxation enhancement). NMR applications to RNP complexes have been recently reviewed [40]. In the reverse direction, from large to small, cryo-EM was recently used to determine the structure of a protein as small as ≈170 kDa [41]; the highest-resolution cryo-EM structure reported thus far has reached a near-atomic 2.2 Å [42].

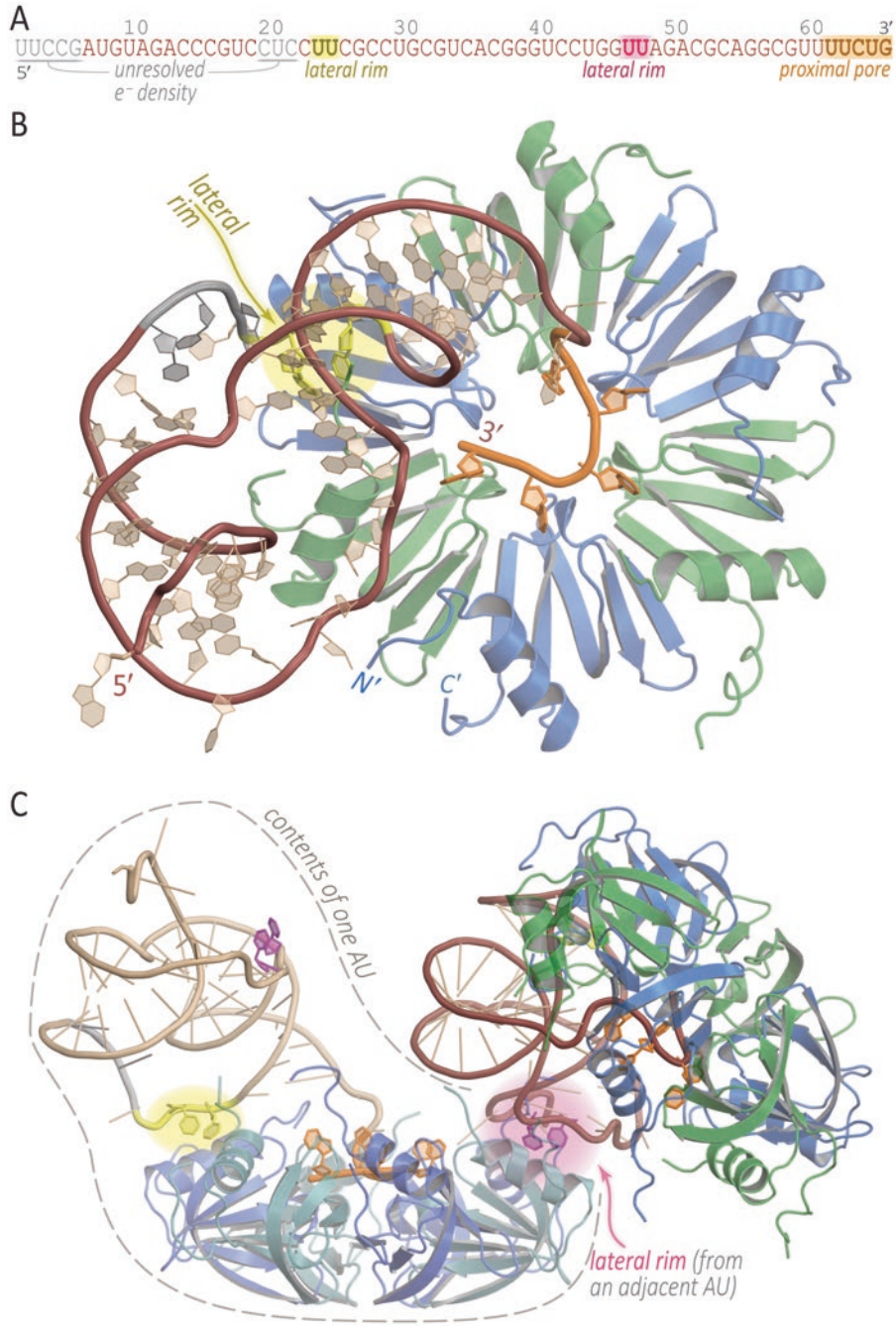


Fig. 1 Crystal structure of Hfq in complex with RydC sRNA (PDB 4V2S). (**a**) The sequence of *S. enterica* RydC sRNA is shown. The gray residues were not discernible in the crystal structure and were manually modeled in (**b**) and (**c**). Residues that bind Hfq at the lateral and proximal sites are highlighted. (**b**) In this cartoon ribbon representation of the *E. coli* Hfq hexamer, alternating monomeric subunits are colored blue and cyan. N′- and C′-termini are labeled for the monomer at the 6-o'clock position. The RydC RNA backbone is shown as a tan-colored tube, with the termini labeled. The 3′ end of the RydC RNA wraps around the *proximal* pore of the Hfq ring, and an internal region of the RNA binds to the *lateral* rim (yellow arrow). Uracil bases involved in binding Hfq at the *proximal* and *lateral* sites are thickened and colored orange and yellow (respectively). (**c**) The RydC sRNA mediates crystal contacts via binding to the *lateral* pocket of an adjacent Hfq hexamer, as indicated by the red arrow. The same coloring scheme is used as in (**b**), with the uridines that facilitate crystal contacts thickened and colored red. This figure was created with PyMOL

As alluded to above, RNA and RNP complexes pose particular challenges in crystallographic structure determination [43, 44]. Most proteins typically adopt a discrete, well-defined three-dimensional (3D) structure, but populations of RNAs tend to sample broad ranges of conformational states, yielding greater structural heterogeneity; notably, this holds even if the sample is technically monodisperse (i.e., homogeneous in terms of MW). Therefore, RNA and RNP crystals often exhibit significant disorder and correspondingly poorer diffraction, as gauged by resolution, mosaicity, and other quality indicators [45]. Synthesis and purification of an RNA construct through in vitro transcription (*see* Subheading 3.2) generates large quantities of chemically homogeneous RNA, and also conveniently lends itself to the engineering of constructs that may be more crystallizable, or that exhibit improved diffraction. Alongside crystallization efforts, chemical probing [46] and structure prediction/modeling methods [47] can be used to examine the secondary structure of the RNA of interest, as well as identify potential protein-binding sites. Then, in designing a more crystallizable construct, extraneous regions of RNA can be either removed or replaced with more stable secondary structures (e.g., stem-loops incorporating tetraloop/tetraloop-receptor pairs); these rigid structural elements can aid crystal contacts and enhance lattice order [44, 48]. As an example of judiciously choosing (and/or designing) an RNA system for crystallographic work, the aforementioned RydC sRNA (Fig. 1a) is a favorable candidate for crystallization efforts because (1) it is relatively small and compact (forming a pseudoknot), and (2) it features multiple U-rich regions that can potentially bind to both its cognate Hfq (within a single RNP complex) and other Hfq proteins across the lattice. In the crystal structure (Fig. 1b, c), the RNA was found to span two Hfq hexamers, forming intermolecular contacts that helped stitch together a stable crystal lattice.

Once crystals that diffract to even low resolution (e.g., $\lesssim 4$ Å) are obtained, a native (underivatized) X-ray diffraction dataset can be collected. From this dataset alone, much can be learned [49], including the likely stoichiometry in the specimen that crystallized (e.g., 1:1 or 2:1 Hfq:RNA?) and whether or not the complex found in the crystalline asymmetric unit (AU) features any additional (non-crystallographic) symmetry. Calculation of an initial electron density map from the diffraction data requires approximate phases for each X-ray reflection. Such phases can be estimated, de novo, via a family of computational approaches based on the two fundamental ideas of multiple isomorphous replacement (MIR) or multi-wavelength anomalous dispersion (MAD); these general approaches require derivatization of native crystals, either via soaking with heavy atoms (MIR/SIR/etc.) or covalent introduction of an anomalously scattering atom (MAD/SAD/etc.) such as selenium. For more information on approaches to de novo estimation of initial sets of phases, *see* [28].

If a known 3D structure is similar to the (unknown) structure that one seeks to determine, then the phase problem can be greatly simplified. In such cases, a phasing approach known as molecular replacement (MR) can be used to estimate initial phases for the unknown structure (the "target") using a known structure (the "probe"), or a suitable modification thereof (e.g., a homology model). Essentially, the MR approach can be thought of as a "fit," via rigid-body transformations that sample the three rotational and three translational degrees of freedom, of the probe structure to the unknown phases of the diffraction data (which, in turn, directly result from the detailed 3D coordinates of the target structure). Because 3D structures are available for Hfq homologs from many species (Table 1), the phase problem is much simplified by using MR. Similarly, the phases computed in refining a given *apo* Hfq structure can then be used as an initial phase estimate for X-ray data collected for a corresponding Hfq·sRNA complex. The protocols below assume that one can successfully estimate initial phases via MR; if such is not the case, e.g., if there is an unexpected/complicated stoichiometry in the AU (say four Hfq rings and three RNAs, in an odd geometric arrangement), then one must resort to de novo phasing methods.

2 Materials

2.1 Hfq Purification

1. DNA sample that contains the *hfq* gene of interest (e.g., genomic DNA).
2. Inducible expression vector capable of encoding a His6× affinity tag (e.g., pET-22b(+) or pET-28b(+)).
3. Chemically competent BL21(DE3) *E. coli* cells.
4. Lysogeny broth (LB) agar plate, supplemented with appropriate antibiotic (e.g., 100 μg/mL ampicillin or 50 μg/mL kanamycin).
5. LB liquid media, supplemented with a suitable antibiotic (e.g., 100 μg/mL ampicillin or 50 μg/mL kanamycin).
6. 1 mM isopropyl β-D-1-thiogalactopyranoside (a 1000× stock [1 M] can be prepared, partitioned into 1-mL aliquots, and stored at −20 °C).
7. Lysis buffer: 50 mM Tris pH 7.5, 750 mM NaCl.
8. Chicken egg-white lysozyme (100× stock at 1 mg/mL).
9. 0.2-μm syringe filters.
10. His-Trap HP pre-packed sepharose column (GE Healthcare).
11. 200 mM Ni_2SO_4.
12. Wash buffer: 50 mM Tris pH 7.5, 200 mM NaCl, 10 mM imidazole.

13. Elution buffer: 50 mM Tris pH 7.5, 200 mM NaCl, 600 mM imidazole.
14. Bovine thrombin (200 U/mL).
15. p-Aminobenzamidine-agarose resin (Sigma).
16. 3-kDa molecular weight cut-off (MWCO) filter-concentrators (Amicon).

2.2 sRNA Purification

1. Template DNA (≈1 μg/μL for a 3-kb linearized plasmid).
2. 10× transcription buffer: 500 mM Tris–HCl pH 8.0, 100 mM NaCl, 60 mM $MgCl_2$, 20 mM spermidine.
3. 10× rNTP mix: rATP, rCTP, rUTP, rGTP, each at 20 mM.
4. 100 mM dithiothreitol (DTT).
5. Diethyl pyrocarbonate (DEPC)-treated H_2O (RNase-free).
6. T7 RNA polymerase (20 U/μL).
7. RNase-free DNase I (50 U/μL).
8. Denaturing (8 M urea) 5% polyacrylamide gel.
9. Elution buffer: 20 mM Tris–HCl pH 7.5, 1 mM EDTA, 1% (w/v) SDS.
10. Phenol:chloroform (1:1).
11. 96% ethanol.

2.3 Co-crystallization Trials

1. Purified Hfq protein (*see* Subheading 3.1).
2. Purified sRNA construct (*see* Subheading 3.2).
3. The "Natrix" and "Crystal Screen" sparse-matrix crystallization screens (Hamton Research).
4. Intelli-Plate 96-3 Microplates (Hampton Research).
5. Sealing tape (Hampton Research).
6. 24-well VDX plates with sealant (Hampton Research).
7. Siliconized glass cover slips (Hampton Research).
8. Vacuum grease.
9. Light stereomicroscope, with cross-polarizing lenses (e.g., Zeiss Discovery V20).

3 Methods

3.1 Hfq Purification

Crystallization efforts typically require large quantities of highly purified and concentrated material, e.g., on the scale of >100 μL at >10 mg/mL of the biomolecule. To achieve this, Hfq is often expressed in a standard *E. coli* K12 laboratory strain, using a plasmid-based construct created via standard recombinant DNA techniques. The expression vector, e.g., an inducible T7*lac*-based

system (pET series), can be used to add various affinity tags to the N′ or C′ termini of the wild-type sequence. Then, overexpressed Hfq can be readily purified via affinity chromatographic means. Further steps, detailed below and in Fig. 2, may be required to remove co-purifying proteins and nucleic acids. Because at least some Hfq homologs bind nucleic acids fairly indiscriminately, the ratio of absorbances at 260 nm and 280 nm (A_{260}/A_{280}) should be monitored over the various stages of purification in order to detect the presence of contaminating nucleic acids. Pure nucleic acid is characterized by an A_{260}/A_{280} ratio of ≈1.5–2.0, versus ≈0.7 for pure protein; rapid colorimetric assays can be used alongside absorbance readings to discern whether a contaminant is mostly RNA or DNA [50].

In terms of solution behavior, experience has shown that Hfq homologs generally remain soluble in aqueous buffers at temperatures >70 °C, and that they resist chemical denaturation (e.g., treatment with 6 M GndCl); in many cases, depending on the species of origin, Hfq samples are insoluble at low temperatures. We have found that the common practice of purifying/storing proteins at 4 °C can be unwise with Hfq homologs: if visible precipitation occurs at ≈4 °C, we recommend that purification be conducted at ambient room temperature (≈18–22 °C), and that elevated temperatures be considered for long-term storage of purified protein (e.g., ≈37–42 °C works well for an Hfq homolog from the hyperthermophile *Aquifex aeolicus*). In terms of protein expression behavior, purification strategies, solubility properties (temperature- and ionic strength-dependence), etc., we have found that the in vitro behavior of many Hfq constructs resembles the overall properties of Hfq orthologs (Sm proteins) from the archaeal domain of life [17].

Previously, His-tagged [18, 51] and self-cleaving intein-tags [16, 21, 52] have been used for affinity purification of Hfq, although it has also been purified without the use of a tag. Untagged Hfq has been purified using immobilized metal affinity chromatography (IMAC), as the native protein has been shown to associate with the resin [4]. Poly(A)-sepharose [10] and butyl-sepharose [53, 54] columns also have been utilized to purify untagged Hfq, leveraging the RNA-binding properties and partially hydrophobic nature of the surface of the protein (respectively). Below, we outline the purification of recombinant Hfq using a His6×-tagged construct. This tag can be removed at a later step through protease treatment; this is a crucial feature, as it is possible that even a modestly sized His6× tag can interfere with the oligomerization behavior and binding properties of Hfq [55]. The intein-mediated purification with an affinity chitin-binding tag (IMPACT) scheme, used to both clone and purify intein-tagged constructs, has been successfully applied to Hfq by multiple labs (e.g., [21, 52]). This system is available as a kit from NEB, so that method will not be

described herein. Note that many of the considerations and notes described below, for His6×-tagged Hfq, also apply when purifying and working with any Hfq construct, intein-based or otherwise.

1. Clone the *hfq* gene, using a genomic sample as PCR template, into an appropriate expression vector; ideally, such a vector will add a His-tag. A compatible vector from the pET series of plasmids often works well (e.g., pET-28b(+), which fuses an *N*-terminal His6× tag).
2. Transform competent BL21(DE3) *E. coli* with the recombinant Hfq plasmid and plate onto LB agar supplemented with antibiotic (e.g., 50 μg/mL kanamycin if using pET-28b(+)).
3. Grow-out the transformed cells in LB media at 37 °C with shaking (225 rpm) to an optical density at 600 nm (OD_{600}) of ≈0.6–0.8. Then, induce overexpression of Hfq by adding IPTG to a final concentration of 1 mM. Optionally, immediately before adding IPTG take a 1-mL aliquot of the cell culture as a $t = 0$ (pre-induction) sample; this sample can be stored at −20 °C and later analyzed alongside a post-induction sample (by SDS-PAGE) in order to assess overexpression levels.
4. Incubate the cell cultures for an additional 3–4 h at 37 °C, with continued shaking, and then centrifuge at 15000 × *g* for 5 min to pellet. Optionally, take a 1-mL aliquot of the cell culture at $t \approx 2$–3 h post-induction; this sample can be stored at −20 °C and analyzed by SDS-PAGE.
5. Resuspend the cell pellet from the previous step in Lysis Buffer (Subheading 2.1). Optionally, DNase I and RNase A can be added at this stage in order to hydrolyze any nucleic acids, Hfq-associated or otherwise (*see* **Note 1**).
6. Incubate the lysate with 0.01 mg/mL lysozyme for 30 min (if a more thorough chemical lysis is required), at either RT or 37 °C; gently shake/invert the sample a few times during this incubation.
7. Mechanically lyse the cells using a sonicator or other similar means (e.g., a microfluidizer or French press). Remove cellular debris from the lysate by centrifugation at 16,000 × *g* for 5–10 min at RT.
8. Initial purification of Hfq can be achieved by using a heat-cut to precipitate endogenous, mesophilic *E. coli* proteins. Proceed by incubating the supernatant from the last step (i.e., clarified lysate) at ≈70–80 °C for ≈10–15 min (*see* **Note 2**); a substantial amount of white precipitate should develop within minutes. Next, use a high-speed centrifugation step (e.g., 33,000 × *g* for 30 min) to clarify the soluble, Hfq-containing supernatant; for pilot studies, the supernatant and pellet

fractions from this step can be saved in case SDS-PAGE analysis becomes necessary.

9. If nucleic acid is still present in the heat-treated sample, as assessed by A_{260}/A_{280} ratios, colorimetric assays [50], or dye-binding assays (e.g., cyanine-based stains such as PicoGreen or SYBR-Gold), then a chaotropic agent such as urea or GndCl can be added to the sample, to a concentration of up to 8 M or 6 M, respectively (*see* **Note 3** and Fig. 2).
10. Pass the latest Hfq-containing sample through a 0.2-μm filter (syringe or vacuum line) to remove any particulate matter, prior to applying the material to a high-performance liquid chromatography (HPLC) or fast protein liquid chromatography (FPLC) system in the next step.
11. Isolate the His6×-tagged Hfq via IMAC, using an iminodiacetic acid sepharose resin in a pre-packed column connected to an HPLC or FPLC instrument. All buffers should be vacuum-filtered (0.45-μm filters) and sonicated before use. In brief, this IMAC step entails the following sub-steps:

    (a) Prepare the resin by washing with 3–4 column volumes (CVs) of dH_2O, and then 3–4 CVs of Wash Buffer.

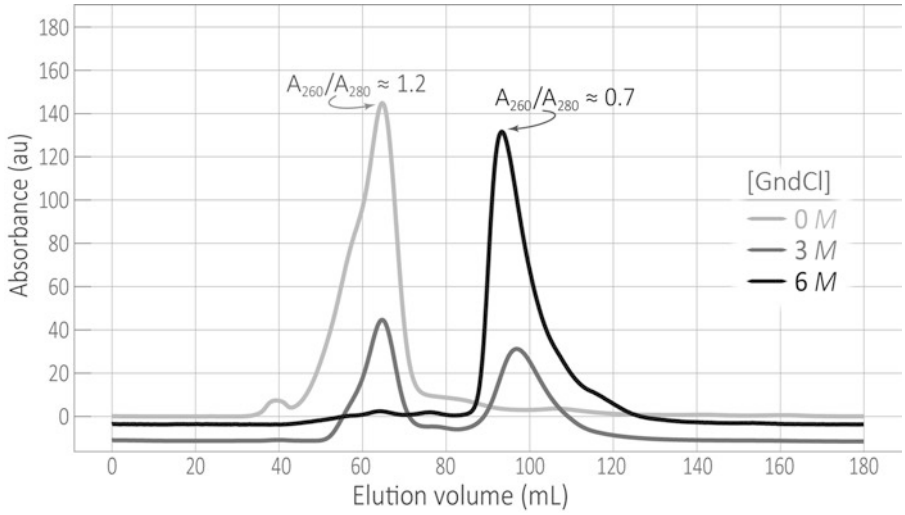


Fig. 2 Size-exclusion chromatography of Hfq samples reveals the impact of a chaotrope such as guanidinium chloride (GndCl) on elution profiles and co-purifying nucleic acid content. In particular, high concentrations of GndCl can disrupt Hfq•••RNA interactions, as shown here via preparative-scale SEC chromatograms for recombinant His-tagged *A. aeolicus* Hfq constructs that were previously purified by IMAC either in the absence (0 M) or in the presence (at 3 M, 6 M) of GndCl. The peak that elutes at ≈60 mL corresponds to Hfq associated with nucleic acids, as indicated by the higher molecular weight (versus Hfq alone) and the high A_{260}/A_{280} absorbance ratio for this eluate; the peak at ≈100 mL corresponds to pure Hfq protein. Note the smooth shift from nucleic acid-bound Hfq to free protein as [GndCl] increases

(b) Charge the resin with Ni^{2+} by loading at least 1–2 CVs of 200 mM $NiSO_4$ (*see* **Note 4**).

(c) Load the crude (unpurified) Hfq-containing protein sample (collect the flow-through), and then wash the column with several CVs of Wash Buffer (until the A_{280} trace drops near baseline).

(d) Elute the Hfq protein by applying a linear gradient of Elution Buffer, from 0 → 100% over 10 CVs.

(e) Combine the fractions thought to contain Hfq (as assessed by A_{280} and the elution profile), and dialyze into 50 mM Tris pH 7.5, 200 mM NaCl, 12.5 mM EDTA in order to remove any residual Ni^{2+}.

(f) To regenerate a column for subsequent use, strip the resin with 4–5 CVs of 100 mM EDTA; remove the EDTA by washing with 5–6 CVs of dH_2O (and, for long-term storage, wash with 20% EtOH).

12. To proteolytically remove the His6×-tag (*see* **Note 5**), incubate the sample overnight with thrombin at a 1:600 mass ratio of thrombin:Hfq.

13. To remove thrombin from the latest sample, either apply the material to a benzamidine column or mix it with free resin (in batch mode).

14. Additional chromatographic steps, such as size-exclusion chromatography (Fig. 2), may be necessary in order to isolate the various populations of hexameric, "free" Hfq versus any subpopulations with RNA bound.

3.2 Large-Scale Synthesis and Purification of the sRNA

In addition to purified protein, crystallizing an RNP complex also requires milligram quantities of RNA of sufficient quality. Here, "quality" means that the ideal RNA sample will be (1) *chemically uniform*, in terms of sequence, length, and phosphate end-chemistry (i.e., uniform covalent structure), and also (2) *structurally homogeneous* (i.e., narrow distribution of conformational states in solution). The first issue—monodispersity—is a fairly straightforward matter of chemistry, and is within one's control (e.g., use an RNA synthesis scheme that minimizes heterogeneity of the 3′-termini of the product RNA molecules). The second issue, concerning structural heterogeneity, is a matter of physics: one can anneal RNAs by heating/cooling, adjusting pH, ionic strength, etc., to try and modulate the solution-state behavior of an RNA, but the intrinsic structural/dynamical properties of RNA are generally not easily regulated; ultimately, one must empirically monitor the RNA and its properties of interest (e.g., "crystallizability").

For sRNAs, which range from ≈50 to 200 nt, a sufficient quantity of material can be readily synthesized via run-off

transcription, in vitro, using phage T7 RNA polymerase and a linearized plasmid as the DNA template (*see* **Note 6**). Traditional in vitro transcription is limited by the fact that T7 polymerase strongly prefers guanosine at the 5′-end of the transcript [56], thus adversely affecting yields for target RNAs lacking a 5′ G. In addition, the polymerase typically incorporates a few nucleotides at the 3′-end of the transcript in a random, template-independent manner, giving an RNA population that is heterogeneous in length and 3′ sequence. Both of these limitations can be avoided by including, 5′ and 3′ to the RNA sequence of interest, a pair of *cis*-acting, self-cleaving ribozymes [57, 58]. The flanking ribozymes ensure that the population of RNA products is accurate and chemically uniform, given the single-nt precision with which ribozymes self-cleave at the scissile bond. In principle, any self-cleaving ribozyme can be used (hammerhead, hairpin, hepatitis δ virus (HDV), etc.). In practice, an engineered 5′ hammerhead and 3′ HDV ribozyme have been found to work well [48], and impose virtually no sequence constraints on the target RNA product; some obligatory base-pairing interactions between the 5′ hammerhead ribozyme and the target RNA sequence does mean that this region will need to be redesigned for each new target RNA construct that one seeks to produce.

A DNA template suitable for the in vitro transcription reaction can be generated by cloning the construct into a high copy number plasmid containing the T7 promoter upstream of a multiple cloning site (MCS). The plasmid will need to be linearized using a restriction enzyme with selectivity to a site that is 3′ of the sequence of interest. The individual components for the in vitro transcription reaction may be prepared by the user or purchased from a manufacturer; whole kits are also commercially available (e.g., MEGAscript T7 transcription kit, Invitrogen). Large quantities of T7 RNA polymerase can be produced in-house in a cost-effective manner by using a His-tagged construct and affinity purification (similar to that described above for Hfq). In general, the concentrations of rNTPs, $MgCl_2$, and T7 polymerase in the transcription reaction will require optimization for each new RNA construct/system. General guidelines and examples can be found in [57–59]. Using the method outlined below, it is ideally possible to generate milligram quantities of RNA.

1. Clone the DNA construct into a high copy number plasmid containing a T7 promoter upstream of a MCS (e.g., the pBluescript or pGEM series; *also see* **Note 7**).
2. Linearize the plasmid using a restriction enzyme for a site 3′ to the sequence of interest.
3. Mix the following components (final concentrations are noted) in the listed order, and incubate at 37 °C for 1–2 h:
   (a) 1× transcription buffer (50 mM Tris–HCl pH 8.0, 10 mM NaCl, 6 mM $MgCl_2$, 2 mM spermidine).

(b) 2 mM rNTP mix.

(c) 10 mM DTT.

(d) Template DNA (≈0.05 μg/μL for a 3-kb linearized plasmid).

(e) DEPC-treated H_2O to bring to volume.

(f) 0.5 U/μL T7 RNA polymerase.

4. To digest the original template, add RNase-free DNase I (2 U DNase I per 1 μg DNA template) and incubate for 30 min.
5. Purify the RNA by first separating on a denaturing (≈8 M urea) 10% w/v polyacrylamide gel and then excising the band corresponding to the transcript (*see* **Note 8**).
6. Add the gel slice to a tube containing 400 μL Elution Buffer and incubate for several hours at 4 °C. Centrifuge at 10,000 × *g* for 10 min at 4 °C and transfer the supernatant to a new tube.
7. Extract the RNA by phase separation with 1–2 V phenol:chloroform. Centrifuge at 10,000 × *g* for 20 min at 4 °C and transfer the aqueous phase to a new tube.
8. Precipitate with 2–3 V ice-cold ethanol.
9. Resuspend in dH_2O or an appropriate buffer (e.g., Tris–HCl pH 7.5, 0.1 mM EDTA).

3.3 Crystallization of the Hfq·sRNA Complex

There is no reliable way to predict the conditions that will yield well-diffracting crystals of a macromolecule or macromolecular complex, such as an Hfq·sRNA assembly: the process is almost entirely empirical. The word "almost" appears in the last sentence because the process of crystallizing biomolecules is one of guided luck. Many of the biochemical properties and behavior of a system that are most salient to crystallization—idiosyncratic variations in solubility with pH, metal ions, presence of ligands, etc.—become manifest as the knowledge that one develops after many hours of working with a biomolecular sample at the bench. This implicit knowledge is highly system-specific (sometimes varying for even a single-residue mutant in a given system), it accumulates in a tortuously incremental manner, and it directly factors into the decision-making steps that ultimately dictate the success of a crystallization effort. Thus, the best advice for crystallizing an Hfq·sRNA complex is to work as extensively as possible to characterize the Hfq and sRNA components, as well as the assembled RNP, prior to extensive crystallization trials.

Ideally, one's samples will be structurally homogeneous, thus increasing the likelihood of successful crystallization. Even given that, still it is often necessary to empirically screen through myriad potential crystallization conditions. High-throughput kits are available for the rapid screening of the many conditions that have successfully yielded crystals for various proteins in the past

(a technique commonly referred to as sparse-matrix screening [60]). We recommend the Natrix Screen (Hampton Research), as it is specifically tailored to nucleic acid and protein–nucleic acid complexes; other commonly used screens, such as the Crystal Screen and PEG-Ion Screen, are also advised. These kits are available in 15-mL or 1-mL high-throughput (HT) formats. Once a potential crystallization condition is identified, further optimization is usually required in order to improve the quality of the crystalline specimen. Often, this is pursued via "grid screens." In grid screens, one or two free parameters are varied in a systematic manner; these parameters often include the buffer and pH, protein concentration, salts (types, concentrations), types and concentrations of other precipitants (e.g., PEGs), inclusion of small-molecule additives, temperature, etc. Further information on crystallization can be found in many excellent texts (e.g., [61]) and other resources, such as the Crystal Growth 101 literature available online (https://hamptonresearch.com/growth_101_lit.aspx).

In general, the purified Hfq and sRNA must be prepared and then assessed for homogeneity and stability before crystallization trials begin (often this is done via biophysical approaches or, ideally, via functional assays). Also, we advise adhering as closely as possible to RNase-free procedures (e.g., use DEPC-treated water, RNase Zap) both in biochemical characterization steps and in handling Hfq, sRNA, and Hfq·sRNA specimens for crystallization trials. The following is a general protocol to get started:

1. Dialyze the purified Hfq into a suitable crystallization buffer. This should be the simplest, most minimalistic buffer in which the biomolecule is stable and soluble, to a concentration of at least 1–2 mg/mL; for instance, a buffer such as 20 mM TrisCl pH 7.5, 200 mM NaCl has often worked well in our experience. Because RNA is involved, inclusion of salts of divalent cations, such as $MgCl_2$, may be found to aid in crystallization and overall diffraction quality.
2. Bring the [Hfq] to ≈15 mg/mL via concentration or dilution, as necessary (*see* **Note 9**); concentration is often achieved via centrifugal filtration devices with a suitable MWCO.
3. To potentially enhance the conformational homogeneity of the sRNA via annealing, incubate at 80 °C and slowly cool to RT; another approach worth trying is to heat the RNA sample and then snap-cool to ≈4 °C on ice.
4. As a cautionary (and troubleshooting) step, one can test for background RNase activity in the crystallization sample by incubating the Hfq and RNA together for ≈2 weeks and assay degradation via PAGE or other methods (a molar ratio of between 1:1 and 2:1 protein:RNA is recommended as a starting point [25, 48, 62]).

Finally, to begin crystallization trials one should follow the manufacturer's protocol for the particular sparse-matrix screen. Crystal trays should be stored in a temperature and humidity-controlled environment. Crystals can take between hours and months to develop. We recommend checking trays for crystals relatively frequently at the start of the process—e.g., after 1, 2, 4, 8, 16, 32 days. Experience suggests that, ideally, precipitation should occur in roughly one-third to one-half of the conditions within minutes of setting up the crystallization drop; if this is not the case, the Hfq, sRNA, or Hfq·sRNA concentrations may need to be adjusted accordingly. Once a potential crystallization condition has been identified, it should be re-made in-house (using one's own reagents) in order to ensure reproducibility. Then, large-scale grid screening and further optimization can be pursued.

Intriguingly, a survey of the PDB identifies several crystallization agents that seem to recur in the crystallization of Hfq and Hfq·RNA complexes, as detailed in Table 1. The most commonly occurring reagents are (1) sodium cacodylate and citrate buffers, (2) PEG 3350 and 2-methyl-2,4-pentanediol (MPD) precipitants, and (3) $MgCl_2$, $CoCl_2$, and KCl salts as additives. Other divalent cations and polyamines, such as metal hexammines (e.g., hexammine cobalt(III) chloride, $[Co(NH_3)_6]Cl_3$), spermine, and spermidine, have been found to aid in the crystallization of many protein·nucleic acid complexes [48, 62].

Several methods can be applied to verify that new crystals are indeed of an Hfq·sRNA complex. A three-well Intelli-Plate (Hampton Research, HR3-118) may be used during sparse-matrix screening in order to test, in parallel, multiple components for each crystallization condition (e.g., the Hfq·sRNA complex, the Hfq alone, and the buffer alone). If crystals appear only in the drop containing Hfq·sRNA complex, there is a high likelihood that the crystals are composed of the complex. Also, macromolecular crystals can be washed, dissolved, and run on an SDS-PAGE or native polyacrylamide gel, or they can be subjected to the flame of a Bunsen burner (biomolecular crystals melt, whereas salt crystals survive this trial by fire [63]). Small-molecule dyes, such as crystal violet or methylene blue, are taken up by macromolecular crystals but not by salt crystals, and thus can be used to distinguish between the two [64]. Finally, obtaining a diffraction dataset is the ultimate way to determine if a given crystal is macromolecular and, if so, the likelihood of a successful structure determination from that specimen.

4 Notes

1. Hfq is known to protect RNAs [65], and nucleic acid may still remain even after nuclease treatment. This potential pitfall should be monitored by A_{260}/A_{280}. If protein degradation is detected by SDS-PAGE or other means, then the Lysis Buffer

used to resuspend the frozen cell pellet should be supplemented with a protease inhibitor cocktail, either commercial or homemade (including such compounds as PMSF, AEBSF, EDTA, aprotinin, and leupeptin).

2. Experience with many recombinant Sm-like protein constructs suggests that the efficacy of the heat-cut step (i.e., degree of purification achieved) can vary greatly with temperature: we have found that many (>5) more *E. coli* proteins retain solubility in the clarified lysate after a 70 °C heat-cut, versus at 75 °C, at least for the BL21(DE3) strain. In purifying a new Hfq homolog, one can test 500 μL-aliquots of the clarified lysate at a series of temperatures near this range, say 65, 70, 75, and 80 °C.
3. In our experience, Hfq withstands treatment with conventional chaotropic agents, such as high concentrations (≈6–8 M) of urea or GndCl. While such treatment may not fully denature the protein, we have found that it can disrupt potential Hfq•••nucleic acid interactions. Adding such denaturants to the wash and elution buffers used in the IMAC stage can help mitigate nucleic acid contamination [23].
4. The divalent cation Co^{2+} has a lower affinity (than Ni^{2+}) for the imidazole side chain of histidine, but it also features less nonspecific binding to arbitrary proteins; if necessary because of persistent contaminants in the Hfq eluate, one can try Co^{2+} in place of Ni^{2+} in the critical IMAC purification step.
5. Often, proteins are crystallized with an intact His6×-tag. Protein tags can potentially interfere with structure or function, although this is less likely with the small His6×-tag. His-tags can also deleteriously affect crystallizability, by increasing the length of a disordered tail or by forming spurious (and weak) crystal contacts that lead to lattice disorder. We recommend cleaving the tag if possible, as this better replicates the wild-type sequence. If crystals cannot be obtained with the untagged protein, the tagged construct should be considered for crystallization too. As two practical anecdotes from our work with the Sm-like archaeal protein (SmAP) homologs of Hfq, we note the following: (1) with *Pyrobaculum aerophilum* SmAP1, a C-terminal His6× tag was found to interfere with oligomerization in vitro, and ultimately the tag was cleaved off in order for crystallization to succeed [66], and (2) for that same recombinant construct, attempts to remove the His-tag via treatment with thrombin failed (even though the linker between the tag and the native protein sequence was designed to include a thrombin recognition site), but the tag could be successfully removed by proteolytic treatment with trypsin. In such work, we generally use mass spectrometry (typically MALDI-TOF, sometimes electrospray) to assess the accuracy of the cut-site and completeness of proteolysis.

6. Previous work has examined the RNA-binding properties of Hfq using either (1) free ribonucleotides of various forms (e.g., rNTPs, rNMPs, etc.), (2) short oligoribonucleotides of ≲30-nt, e.g., the rU_6 oligo co-crystallized by Stanek et al. [23], or (3) longer, full-length sRNAs, such as the 65-nt *Salmonella* RydC sRNA co-crystallized with *E. coli* Hfq [25]. The nucleotides in category (1) are readily purchased, and the oligonucleotides in category (2) are readily obtained via stepwise, solid-phase chemical synthesis (such RNAs are available from various suppliers, e.g., Dharmacon). In contrast, such approaches are inefficient for the longer (≳30-nt) oligonucleotides of (3), and these can be efficiently generated by enzymatic synthesis in vitro, using RNA polymerase as described above.
7. The plasmids pUC18 and pUC19 are also commonly used for in vitro transcription. The T7 promoter will need to be cloned into these plasmids as well.
8. If the RNA construct contains ribozymes, be careful to not overload gels in order to enable the correctly processed, self-cleaved RNA transcript to be isolated from other products that differ by only a few nucleotides.
9. Typically, proteins are crystallized at concentrations between ≈5 and 20 mg/mL. Nevertheless, concentrations well outside this range have been required for Hfq and other Sm proteins; for instance, *A. aeolicus* Hfq crystallized at 4 mg/mL [23], while *P. aerophilum* SmAP3 was at 85 mg/mL [67]. Hfq concentrations may well be limited by protein solubility (not just supply), and likely will need to be varied in any successful set of crystallization trials.
10. Tacsimate is "a mixture of titrated organic acid salts" that contains 1.8 M malonic acid, 0.25 M ammonium citrate tribasic, 0.12 M succinic acid, 0.3 M DL-malic acid, 0.4 M sodium acetate trihydrate, 0.5 M sodium formate, and 0.16 M ammonium tartrate dibasic (for more information, *see* http://hamptonresearch.com/documents/product/hr000175_what_is_tacsimate_new.pdf).
11. This structure contains molecules of both ADP and the non-hydrolyzable ATP analog AMP-PNP.
12. In this serendipitous co-crystal structure of *E. coli* Hfq and catalase HPII, an Hfq hexamer was found to bind each subunit of a HPII tetramer.
13. MPEG, an acronym for methoxypolyethylene glycol (also known as PEG monomethyl ether), has a covalent formula of $CH_3(OCH_2CH_2)_nOH$, versus $H(OCH_2CH_2)_nOH$ for simple PEGs.

Acknowledgments

We thank L. Columbus (UVa) for helpful discussions. This work was funded by NSF Career award MCB–1350957.

References

1. Franze de Fernandez MT, Eoyang L, August JT (1968) Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. Nature 219(5154):588–590
2. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. Nat Rev Microbiol 9(8):578–589. https://doi.org/10.1038/nrmicro2615
3. Sauer E (2013) Structure and RNA-binding properties of the bacterial LSm protein Hfq. RNA Biol 10(4):610–618. https://doi.org/10.4161/rna.24201
4. Soper T, Mandin P, Majdalani N, Gottesman S, Woodson SA (2010) Positive regulation by small RNAs and the role of Hfq. Proc Natl Acad Sci 107(21):9602–9607. https://doi.org/10.1073/pnas.1004435107
5. De Lay N, Schu DJ, Gottesman S (2013) Bacterial small RNA-based negative regulation: Hfq and its accomplices. J Biol Chem 288(12):7996–8003. https://doi.org/10.1074/jbc.R112.441386
6. Jousselin A, Metzinger L, Felden B (2009) On the facultative requirement of the bacterial RNA chaperone, Hfq. Trends Microbiol 17(9):399–405. https://doi.org/10.1016/j.tim.2009.06.003
7. Tsui HC, Leung HC, Winkler ME (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. Mol Microbiol 13(1):35–49
8. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S (2001) Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev 15(13):1637–1651. https://doi.org/10.1101/gad.901001
9. Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet 4(8):e1000163. https://doi.org/10.1371/journal.pgen.1000163
10. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Mol Cell 9(1):11–22
11. Sledjeski DD, Whitman C, Zhang A (2001) Hfq is necessary for regulation by the untranslated RNA DsrA. J Bacteriol 183(6):1997–2005. https://doi.org/10.1128/jb.183.6.1997-2005.2001
12. Fantappie L, Metruccio MM, Seib KL, Oriente F, Cartocci E, Ferlicca F, Giuliani MM, Scarlato V, Delany I (2009) The RNA chaperone Hfq is involved in stress response and virulence in Neisseria meningitidis and is a pleiotropic regulator of protein expression. Infect Immun 77(5):1842–1853. https://doi.org/10.1128/iai.01216-08
13. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae. Cell 118(1):69–82. https://doi.org/10.1016/j.cell.2004.06.009
14. Mika F, Hengge R (2013) Small regulatory RNAs in the control of motility and biofilm formation in E. coli and salmonella. Int J Mol Sci 14(3):4560–4579. https://doi.org/10.3390/ijms14034560
15. Chao Y, Vogel J (2010) The role of Hfq in bacterial pathogens. Curr Opin Microbiol 13(1):24–33. https://doi.org/10.1016/j.mib.2010.01.001
16. Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. EMBO J 21(13):3546–3556. https://doi.org/10.1093/emboj/cdf322
17. Mura C, Randolph PS, Patterson J, Cozen AE (2013) Archaeal and eukaryotic homologs of Hfq: a structural and evolutionary perspective on Sm function. RNA Biol 10(4):636–651. https://doi.org/10.4161/rna.24538
18. Sauer E, Weichenrieder O (2011) Structural basis for RNA 3′-end recognition by Hfq. Proc Natl Acad Sci U S A 108(32):13065–13070. https://doi.org/10.1073/pnas.1103420108
19. Link TM, Valentin-Hansen P, Brennan RG (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. Proc Natl

Acad Sci U S A 106(46):19292–19297. https://doi.org/10.1073/pnas.0908744106

20. Horstmann N, Orans J, Valentin-Hansen P, Shelburne SA 3rd, Brennan RG (2012) Structural mechanism of Staphylococcus aureus Hfq binding to an RNA A-tract. Nucleic Acids Res 40(21):11023–11035. https://doi.org/10.1093/nar/gks809
21. Sun X, Wartell RM (2006) *Escherichia coli* Hfq binds A_{18} and DsrA domain II with similar 2:1 Hfq_6/RNA stoichiometry using different surface sites. Biochemistry 45(15):4875–4887. https://doi.org/10.1021/bi0523613
22. Panja S, Schu DJ, Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. Nucleic Acids Res 41(15):7536–7546. https://doi.org/10.1093/nar/gkt521
23. Stanek KA, Patterson-West J, Randolph PS, Mura C (2017) Crystal structure and RNA-binding properties of an Hfq homolog from the deep-branching Aquificae: conservation of the lateral RNA-binding mode. Acta Crystallogr D Struct Biol 73(Pt 4):294–315. https://doi.org/10.1107/s2059798317000031
24. Schu DJ, Zhang A, Gottesman S, Storz G (2015) Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. EMBO J 34(20):2557–2573. 10.15252/embj.201591569
25. Dimastrogiovanni D, Frohlich KS, Bandyra KJ, Bruce HA, Hohensee S, Vogel J, Luisi BF (2014) Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. elife 3. https://doi.org/10.7554/eLife.05375
26. Schröder GF (2015) Hybrid methods for macromolecular structure determination: experiment with expectations. Curr Opin Struct Biol 31:20–27. https://doi.org/10.1016/j.sbi.2015.02.016
27. Schlundt A, Tants JN, Sattler M (2017) Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition. Methods 118-119:119–136. https://doi.org/10.1016/j.ymeth.2017.03.015
28. Sherwood D, Cooper JB, Press OU (2011) Crystals, X-rays and proteins: comprehensive protein crystallography. Oxford University Press, Oxford
29. Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ, Rance M (2010) Protein NMR spectroscopy: principles and practice. Elsevier Science, Amsterdam
30. Bai XC, McMullan G, Scheres SHW (2015) How cryo-EM is revolutionizing structural biology. Trends Biochem Sci 40(1):49–57. https://doi.org/10.1016/j.tibs.2014.10.005
31. Cheng Y, Grigorieff N, Penczek PA, Walz T (2015) A primer to single-particle cryo-electron microscopy. Cell 161(3):438–449. https://doi.org/10.1016/j.cell.2015.03.050
32. Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell 102(5):615–623
33. Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature 407(6802):340–348
34. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science (New York, NY) 289(5481):905–920
35. Cohen SB, Graham ME, Lovrecz GO, Bache N, Robinson PJ, Reddel RR (2007) Protein composition of catalytically active human telomerase from immortal cells. Science (New York, NY) 315(5820):1850–1853. https://doi.org/10.1126/science.1138596
36. Nguyen THD, Galej WP, X-c B, Savva CG, Newman AJ, Scheres SHW, Nagai K (2015) The architecture of the spliceosomal U4/U6•U5 tri-snRNP. Nature 523(7558):47–52
37. Agafonov DE, Kastner B, Dybkov O, Hofele RV, Liu WT, Urlaub H, Luhrmann R, Stark H (2016) Molecular architecture of the human U4/U6.U5 tri-snRNP. Science (New York, NY) 351(6280):1416–1420. https://doi.org/10.1126/science.aad2085
38. Tangprasertchai NS, Zhang X, Ding Y, Tham K, Rohs R, Haworth IS, Qin PZ (2015) An integrated spin-labeling/computational-modeling approach for mapping global structures of nucleic acids. Methods Enzymol 564:427–453. https://doi.org/10.1016/bs.mie.2015.07.007
39. Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40(3):191–285. https://doi.org/10.1017/s0033583507004635
40. Yadav DK, Lukavsky PJ (2016) NMR solution structure determination of large RNA-protein complexes. Prog Nucl Magn Reson Spectrosc 97:57–81. https://doi.org/10.1016/j.pnmrs.2016.10.001

41. Lu P, X-c B, Ma D, Xie T, Yan C, Sun L, Yang G, Zhao Y, Zhou R, Scheres SHW, Shi Y (2014) Three-dimensional structure of human γ-secretase. Nature 512(7513):166–170. https://doi.org/10.1038/nature13567
42. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S (2015) 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. Science (New York, NY) 348(6239):1147–1151. https://doi.org/10.1126/science.aab1576
43. Ferre-D'Amare AR, Zhou K, Doudna JA (1998) A general module for RNA crystallization. J Mol Biol 279(3):621–631. https://doi.org/10.1006/jmbi.1998.1789
44. Ferre-D'Amare AR, Doudna JA (2001) Methods to crystallize RNA. Curr Protoc Nucleic Acid Chem Chapter 7:Unit 7 6. https://doi.org/10.1002/0471142700.nc0706s00
45. Evans P (2006) Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr 62(Pt 1):72–82. https://doi.org/10.1107/S0907444905036693
46. Deigan KE, Li TW, Mathews DH, Weeks KM (2009) Accurate SHAPE-directed RNA structure determination. Proc Natl Acad Sci 106(1):97–102. https://doi.org/10.1073/pnas.0806929106
47. Eddy SR (2014) Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. Annu Rev Biophys 43:433–456. https://doi.org/10.1146/annurev-biophys-051013-022950
48. Ke A, Doudna JA (2004) Crystallization of RNA and RNA-protein complexes. Methods 34(3):408–414. https://doi.org/10.1016/j.ymeth.2004.03.027
49. Sawaya MR (2007) Characterizing a crystal from an initial native dataset. Methods Mol Biol 364:95–120. https://doi.org/10.1385/1-59745-266-1:95
50. Patterson J, Mura C (2013) Rapid colorimetric assays to qualitatively distinguish RNA and DNA in biomolecular samples. J Vis Exp 72:e50225. https://doi.org/10.3791/50225
51. Sukhodolets MV, Garges S (2003) Interaction of Escherichia coli RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. Biochemistry 42(26):8022–8034. https://doi.org/10.1021/bi020638i
52. Møller T, Franch T, Højrup P, Keene DR, Bächinger HP, Brennan RG, Valentin-Hansen P (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol Cell 9(1):23–30
53. Nikulin A, Stolboushkina E, Perederina A, Vassilieva I, Blaesi U, Moll I, Kachalova G, Yokoyama S, Vassylyev D, Garber M, Nikonov S (2005) Structure of Pseudomonas aeruginosa Hfq protein. Acta Crystallogr D Biol Crystallogr 61(Pt 2):141–146. https://doi.org/10.1107/s0907444904030008
54. Kadowaki MA, Iulek J, Barbosa JA, Pedrosa Fde O, de Souza EM, Chubatsu LS, Monteiro RA, de Oliveira MA, Steffens MB (2012) Structural characterization of the RNA chaperone Hfq from the nitrogen-fixing bacterium Herbaspirillum seropedicae SmR1. Biochim Biophys Acta 1824(2):359–365. https://doi.org/10.1016/j.bbapap.2011.11.002
55. Obregon KA, Hoch CT, Sukhodolets MV (2015) Sm-like protein Hfq: composition of the native complex, modifications, and interactions. Biochim Biophys Acta 1854(8):950–966. https://doi.org/10.1016/j.bbapap.2015.03.016
56. Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res 15(21):8783–8798
57. Price SR, Ito N, Oubridge C, Avis JM, Nagai K (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. J Mol Biol 249(2):398–408
58. Ferre-D'Amare AR, Doudna JA (1996) Use of cis- and trans-ribozymes to remove 5′ and 3′ heterogeneities from milligrams of in vitro transcribed RNA. Nucleic Acids Res 24(5):977–978
59. Beckert B, Masquida B (2011) Synthesis of RNA by in vitro transcription. In: Nielsen H (ed) RNA: methods and protocols. Humana Press, Totowa, NJ, pp 29–41. https://doi.org/10.1007/978-1-59745-248-9_3
60. McPherson A, Gavira JA (2014) Introduction to protein crystallization. Acta Crystallogr Sect F Struct Biol Cryst Commun 70(Pt 1):2–20. https://doi.org/10.1107/S2053230X13033141
61. McPherson A (1999) Crystallization of biological macromolecules. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
62. Obayashi E, Oubridge C, Krummel DP, Nagai K (2007) Crystallization of RNA-protein complexes. In: Walker JM, Doublié S (eds) Macromolecular crystallography protocols: Volume 1, Preparation and crystallization of macromolecules. Humana Press, Totowa, NJ, pp 259–276. https://doi.org/10.1007/978-1-59745-209-0_13
63. Raghunathan K, Harris PT, Arvidson DN (2010) Trial by fire: are the crystals macromolecules? Acta Crystallogr Sect F Struct Biol Cryst Commun 66(Pt 5):615–620. https://doi.org/10.1107/S1744309110012078

64. Salt or Protein Crystal? Hampton Research. https://hamptonresearch.com/documents/growth_101/20.pdf
65. Folichon M, Arluison V, Pellegrini O, Huntzinger E, Regnier P, Hajnsdorf E (2003) The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. Nucleic Acids Res 31(24):7302–7310
66. Mura C, Cascio D, Sawaya MR, Eisenberg DS (2001) The crystal structure of a heptameric archaeal Sm protein: implications for the eukaryotic snRNP core. Proc Natl Acad Sci U S A 98(10):5532–5537. https://doi.org/10.1073/pnas.091102298
67. Mura C, Phillips M, Kozhukhovsky A, Eisenberg D (2003) Structure and assembly of an augmented Sm-like archaeal protein 14-mer. Proc Natl Acad Sci U S A 100(8):4539–4544. https://doi.org/10.1073/pnas.0538042100
68. Sauter C, Basquin J, Suck D (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. Nucleic Acids Res 31(14):4091–4098
69. Nielsen JS, Boggild A, Andersen CB, Nielsen G, Boysen A, Brodersen DE, Valentin-Hansen P (2007) An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from *Methanococcus jannaschii*. RNA 13(12):2213–2223
70. Bøggild A, Overgaard M, Valentin-Hansen P, Brodersen DE (2009) Cyanobacteria contain a structural homologue of the Hfq protein with altered RNA-binding properties. FEBS J 276(14):3904–3915
71. Moskaleva O, Melnik B, Gabdulkhakov A, Garber M, Nikonov S, Stolboushkina E, Nikulin A (2010) The structures of mutant forms of Hfq from *Pseudomonas aeruginosa* reveal the importance of the conserved His57 for the protein hexamer organization. Acta Crystallogr Sect F Struct Biol Cryst Commun 66(Pt 7):760–764
72. Bonnefond L, Schellenberger P, Basquin J, Demangeat G, Ritzenthaler C, Chênevert R, Balg C, Frugier M, Rudinger-Thirion J, Giegé R, Lorber B, Sauter C (2011) Exploiting protein engineering and crystal polymorphism for successful X-ray structure determination. Cryst Growth Des 11(10):4334–4343
73. Someya T, Baba S, Fujimoto M, Kawai G, Kumasaka T, Nakamura K (2012) Crystal structure of Hfq from *Bacillus subtilis* in complex with SELEX-derived RNA aptamer: insight into RNA-binding properties of bacterial Hfq. Nucleic Acids Res 40(4):1856–1867
74. Beich-Frandsen M, Vecerek B, Sjoblom B, Blasi U, Djinovic-Carugo K (2011) Structural analysis of full-length Hfq from *Escherichia coli*. Acta Crystallogr Sect F Struct Biol Cryst Commun 67(Pt 5):536–540
75. Wang W, Wang L, Zou Y, Zhang J, Gong Q, Wu J, Shi Y (2011) Cooperation of *Escherichia coli* Hfq hexamers in DsrA binding. Genes Dev 25(19):2106–2117
76. Hammerle H, Beich-Frandsen M, Vecerek B, Rajkowitsch L, Carugo O, Djinovic-Carugo K, Blasi U (2012) Structural and biochemical studies on ATP binding and hydrolysis by the *Escherichia coli* RNA chaperone Hfq. PLoS One 7(11):e50892
77. Murina V, Lekontseva N, Nikulin A (2013) Hfq binds ribonucleotides in three different RNA-binding sites. Acta Crystallogr D Biol Crystallogr 69(Pt 8):1504–1513
78. Yonekura K, Watanabe M, Kageyama Y, Hirata K, Yamamoto M, Maki-Yonekura S (2013) Post-transcriptional regulator Hfq binds catalase HPII: crystal structure of the complex. PLoS One 8(11):e78216
79. Wang W, Wang L, Wu J, Gong Q, Shi Y (2013) Hfq-bridged ternary complex is important for translation activation of *rpoS* by DsrA. Nucleic Acids Res 41(11):5938–5948
80. Robinson KE, Orans J, Kovach AR, Link TM, Brennan RG (2014) Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. Nucleic Acids Res 42(4):2736–2749
81. Murina VN, Melnik BS, Filimonov VV, Uhlein M, Weiss MS, Muller U, Nikulin AD (2014) Effect of conserved intersubunit amino acid substitutions on Hfq protein structure and stability. Biochemistry (Mosc) 79(5):469–477
82. Kovach AR, Hoff KE, Canty JT, Orans J, Brennan RG (2014) Recognition of U-rich RNA by Hfq from the Gram-positive pathogen *Listeria monocytogenes*. RNA 20(10):1548–1559
83. Schulz EC, Barabas O (2014) Structure of an *Escherichia coli* Hfq:RNA complex at 0.97 Å resolution. Acta Crystallogr F Struct Biol Commun 70(Pt 11):1492–1497
84. Wang L, Wang W, Li F, Zhang J, Wu J, Gong Q, Shi Y (2015) Structural insights into the recognition of the internal A-rich linker from OxyS sRNA by *Escherichia coli* Hfq. Nucleic Acids Res 43(4):2400–2411
85. Feng SQ, Si YL, Song CY, Wang PQ, Su JY (2015) Limited proteolysis improves *E. coli* Hfq crystal structure resolution. Chinese J Biochem Mol Biol 31(10):1102–1108
86. Nikulin A, Mikhailina A, Lekontseva N, Balobanov V, Nikonova E, Tishchenko S (2017) Characterization of RNA-binding properties of the archaeal Hfq-like protein from *Methanococcus jannaschii*. J Biomol Struct Dyn 35(8):1615–1628

Chapter 17

Single-Molecule FRET Assay to Observe the Activity of Proteins Involved in RNA/RNA Annealing

Thierry Bizebard, Véronique Arluison, and Ulrich Bockelmann

Abstract

In recent years, single-molecule fluorescence resonance energy transfer (smFRET) has emerged as a powerful technique to study macromolecular interactions. The chief advantages of smFRET analysis compared to bulk measurements include the possibility to detect sample heterogeneities within a large population of molecules and the facility to measure kinetics without needing the synchronization of intermediate states. As such, the methodology is particularly well adapted to observe and analyze RNA/RNA and RNA/protein interactions involved in small noncoding RNA-mediated gene regulation networks. In this chapter, we describe and discuss protocols that can be used to measure the dynamics of these interactions, with a particular emphasis on the advantages—and experimental pitfalls—of using the smFRET methodology to study sRNA-based biological systems.

Key words Small noncoding RNA, Single-molecule FRET, RNA annealing, Hfq, sRNA

Abbreviations

smFRET	Single-molecule Förster resonance energy transfer
sRNA	Small noncoding RNA

1 Introduction

Förster resonance energy transfer (FRET) is a mechanism of energy transfer between two fluorophores, namely the donor and acceptor. One condition for FRET to occur is that donor emission and acceptor excitation spectra partially overlap. Upon donor excitation, nonradiative energy transfer to the acceptor occurs. The efficiency of energy transfer is inversely proportional to the sixth power of the distance between the donor and the acceptor, making FRET happen only when fluorophores are in close proximity (typically less than 100 Å, Fig. 1a).

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_17,

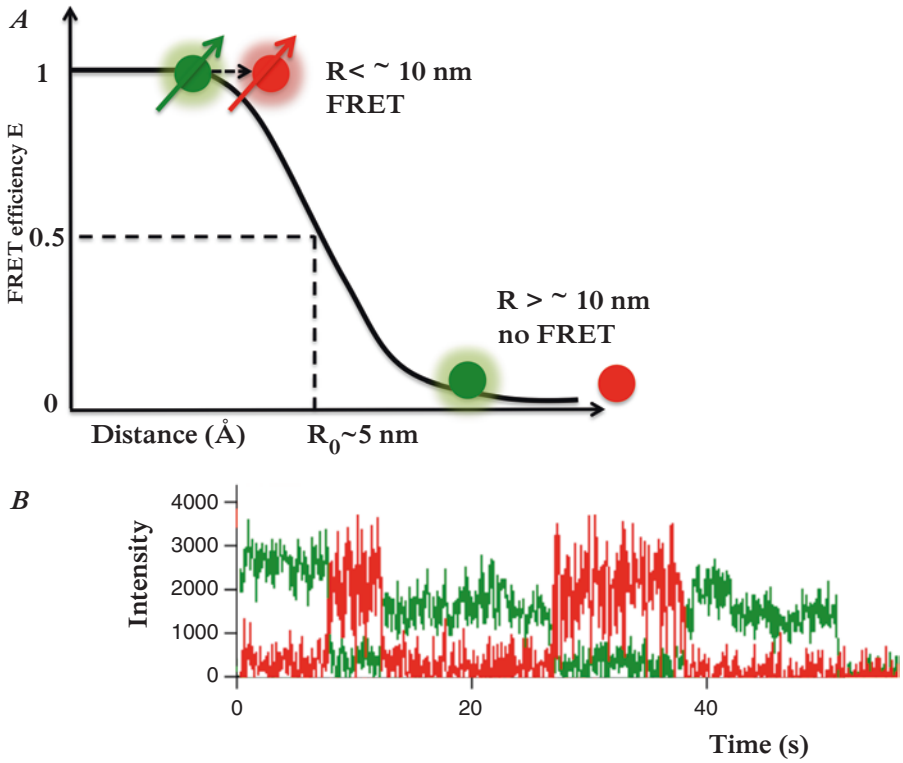


Fig. 1 Principle of FRET. (**a**) FRET efficiency E (the ratio of energy transfer events per donor excitation events) depends on the distance between donor (red) and acceptor (green), the spectral overlap of donor emission and acceptor absorption spectra and relative orientation of donor emission dipole moment (red arrow) and acceptor absorption dipole moment (green arrow). FRET typically occurs in the range of 1–10 nm. (**b**) a typical single-molecule fluorescence trajectory: donor emission (green) and acceptor emission (red). The anti-correlation of the donor and acceptor fluorescence emissions is indicative of FRET

Precisely, the efficiency of energy transfer (*E*) is given as:

$$E = \frac{R^{-6}}{R^{-6} + R_0^{-6}}$$

where R is the distance between the two dyes and R_0 is the Förster radius at which energy transfer is half-maximal ($E = 0.5$; Fig. 1a). R_0 is usually in the range of ~1.5–6 nm, depending on the donor/acceptor couple used. Indeed, R_0 depends mainly on the refractive index of the medium (n), the dipoles orientation factor (κ^2), the fluorescence quantum yield of the donor in the absence of the acceptor (Φ_D), and the spectral overlap integral of donor and acceptor (J).

Recent progress in optics (i.e., laser light sources and ultrasensitive cameras) allowed the development of single-molecule FRET techniques (smFRET). In the case of smFRET, a pair of donor and

acceptor dyes is excited and detected at the single-molecule level, not in ensemble as in the case of conventional bulk FRET. This can be achieved by analyzing either diffusing molecules in a small volume or with molecules immobilized on a surface. For instance, confocal microscopy can be used to follow conformational motion of individual proteins [1]. In this case, a single fluorescent protein is excited when it diffuses through the small excitation volume of the confocal microscope. Nevertheless, to observe interaction between different partners and make a statistics on a population, imaging of surface immobilized molecules by total internal reflection fluorescence (TIRF) microscopy is preferred. In TIRF microscopy, an evanescent field of excitation light only penetrates by ~100 nm into the immediate adjacency to the surface, allowing the selective excitation of individual surface-bound fluorophores, while non-bound molecules are not excited. Exciting an area of approximately 1000–5000 μm^2 allows imaging hundreds of distinct molecules in parallel and to observe kinetics of energy transfer between dyes (Fig. 1b). Thus, the advantage of using TIRF for smFRET analysis mainly relies on: (1) the detection of sample heterogeneity within a large population of molecules, (2) the detection of transient intermediate states that can be hidden in bulk measurements, thus allowing building more complete mechanisms, and (3) the measurement of kinetics without the need for synchronization of molecules (Fig. 2).

Recently, three-color smFRET, i.e., measurement of FRET between one donor and two alternative acceptors, has also emerged as a powerful tool in order to analyze complex reactions (Fig. 3). This is particularly useful for investigations involving more than 2 partners, as in the case of regulatory RNA annealing to target mRNA, which is in turn usually dependent on a protein cofactor. Within the gene regulation networks involving RNA/RNA interactions, bacterial small noncoding RNA (sRNA) are of particular interest as they control response to cellular stresses by base pairing with their mRNA targets. They usually inhibit translation by covering the *Ribosome Binding Site* (RBS), with important consequences on RNA stability. As most of these sRNA need a protein cofactor (Hfq or ProQ) to facilitate base pairing, sRNA annealing process thus usually involves three cellular partners [2, 3]. Furthermore, as RNA melting and/or RNA strand exchange may occur simultaneously, three-color smFRET also permit following these different processes in parallel [4, 5].

In this chapter, we will describe and discuss protocols that can be used to measure kinetics of RNA/RNA annealing in vitro, with a particular focus on sRNA/mRNA/Hfq complexes. Advantage and drawbacks of the use of immobilized molecules and of protein labeling, together with information that can be obtained with smFRET experiments, will be discussed in detail.

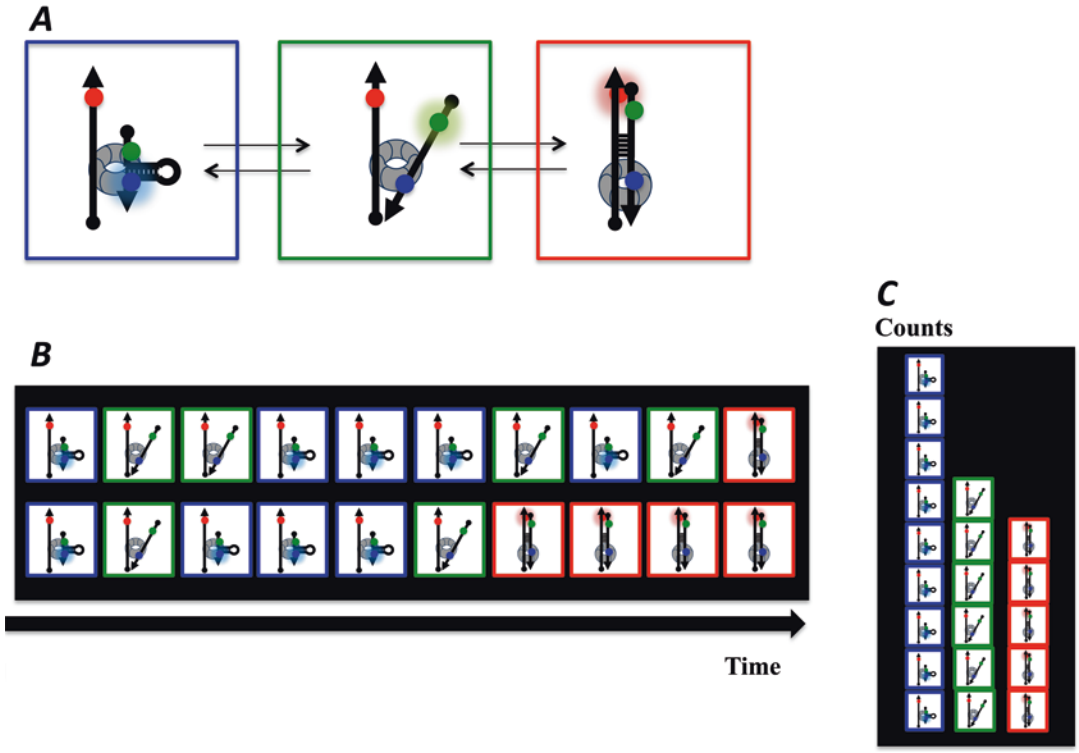


Fig. 2 Information obtained by smFRET analysis. (**a**) Observation of different FRET intermediates. (**b**) Single-molecule kinetics allows characterizing intermediate states: in schematic drawing there is always a green conformation between the blue conformation and the red conformation. (**c**) Different populations can also be quantified and shown as single-molecule histograms

2 Materials

2.1 Microscopy

1. Coverslips: 24 × 32 mm #1 (Menzel-Gläser).
2. Microscope Slides 76 × 26 mm, cut edges, frosted end (Menzel-Gläser).
3. Inverted microscope (Zeiss Axiovert 200).
4. Back-illuminated EM-CCD camera 512 × 512 pixel (Andor iXon+).
5. Beam expander 2–8×, λ = 532 nm (Linos).
6. Beam expander 2–8×, λ = 633 nm (Linos).
7. Beam combiner dichroic z532bcm (Chroma Technology Corp.).
8. Fluorescence dichroic z532/633rpc (Chroma Technology Corp.).
9. Emission filter HQCy3/Cy5m (Chroma Technology Corp.).

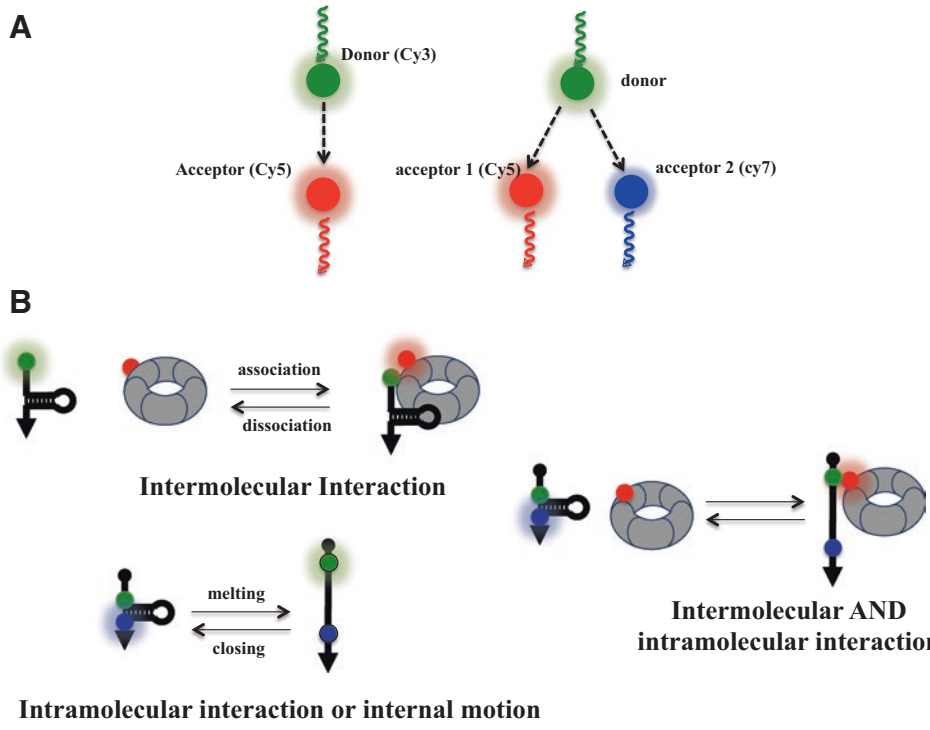


Fig. 3 Three-color FRET. Single-molecule three-color FRET measures FRET between one donor (green) and two alternative acceptors (red and blue). This allows to observe correlated melting of RNA and binding to a protein (for instance, Hfq protein, grey torus). 5′ end of RNA is depicted as a sphere and 3′ end is as an arrowhead

10. Image-splitter beam divider 630dcrx (Chroma Technology Corp.).
11. Image-splitter (Cairn research Optosplit II).
12. CW 532 nm 50 mW DPSS laser (Cobolt Samba).
13. CW 638 nm 25 mW laser (Coherent Cube).
14. TIRF objective (Zeiss alpha "Plan-Apochromat" 100×/1.46 Oil DIC).
15. Kwik-cast sealant (WPI).
16. Softwares: Matlab (MathWorks); vbFRET [6]; LabView (National Instruments); Igor (WaveMetrics Inc.)

2.2 Reagents and Enzymes

1. HPLC grade methanol.
2. HPLC grade acetone.
3. Glucose oxidase.
4. Catalase.
5. Streptavidin.
6. Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) (from Sigma-Aldrich).

7. Vectabond (aminosilane reagent) (from Vector).
8. N-Hydroxysuccinimide (NHS) polyethylene glycol (PEG; MW 5000) (from Laysan Bio).
9. Biotin-PEG-NHS ester (from Laysan Bio).
10. Tris[2-carboxyethyl]phosphine (TCEP).
11. Cy5-maleimide (from GE Healthcare).
12. Ni-NTA resin

2.3 Buffers and Solutions

1. Washing buffer 1: 20 mM Tris–HCl pH 8, 20 mM imidazole, 0.3 M NaCl, 0.1 mM TCEP.
2. Elution buffer: 20 mM Tris–HCl pH 8, 250 mM imidazole, 0.3 M NaCl, 0.1 mM TCEP.
3. Working buffer: 50–300 mM NaCl, 0–5 mM $MgCl_2$, 50 mM Tris–HCl pH 8.
4. Washing buffer 2: 50 mM NaCl, 10 mM Tris–HCl pH 8.
5. Imaging buffer: 50–300 mM NaCl, 0–5 mM $MgCl_2$, 50 mM Tris–HCl pH 8, 0.8%(w:v) glucose, 165 U/mL glucose oxidase, 2000 U/mL catalase, 1.5–2 mM Trolox.

3 Methods

3.1 Choice of Dyes and Buffers for Single Molecule Fluorescent Experiment

3.1.1 Dyes

For smFRET analysis, dyes must be bright, which means usually an extinction coefficient $\varepsilon > 50{,}000\ M^{-1}\ cm^{-1}$ and a quantum yield $\Phi > 0.1$ [7]. Dyes must be photostable and it is recommended to choose FRET pairs with a large spectral separation between donor and acceptor emission. Consequently, we suggest using organic fluorophores such as Cyanine, Alexa, and ATTO dyes. Cyanine dyes (Cy3, Cy5, Cy5.5, and Cy7) are our favorites, although other dyes (ATTO550 or Alexa555 instead of Cy3, for instance) may work properly (Table 1). It is finally better to choose two fluorophores with similar quantum yields in order to have similar donor and acceptor signals. In brief, common cyanine pairs used for FRET and three-color FRET are Cy3 as donor, Cy5 as acceptor 1, and Cy5.5 or Cy7 as acceptor 2.

To ensure the reproducibility of the experiments described below, we recommend using cyanine-labeled oligoribonucleotides from Eurogentec. The labeling efficiencies are usually >90% and not less than 70%. As necessary for the experiment, fluorophores can be placed at either extremity (5′ or 3′) of the RNA oligoribonucleotide, or even at an internal location.

Some of the fluorescent oligoribonucleotides will be used for immobilization, in order to be able to perform smFRET studies on immobilized molecules. To achieve this goal, we suggest using a biotin at an extremity of the oligoribonucleotide (5′ or 3′),

Table 1
Properties of cyanine dyes used for smFRET measurement

Cyanine name	$\lambda_{Excitation}$ (nm)	$\lambda_{Emission}$ (nm)	Quantum yield Φ
Cy3	550	565	0.12
Cy5	650	670	0.27
Cy5.5	675	694	0.28
Cy7	753	775	0.28

depending on where the fluorescent modification has been designed on the same oligoribonucleotide. We recommend adding a chemical spacer between oligonucleotidic RNA sequence and the biotin used for immobilization in order to avoid interferences with surface proximity during RNA annealing. Oligoribonucleotides providers propose different spacers of various lengths.

3.1.2 Buffer Optimization

Even though the buffer condition may need to be optimized depending on specific protein, we suggest starting with the "working buffer." $MgCl_2$ may have to be included depending on the system of interest. In the specific case of Hfq, a high-salt buffer (containing 300 mM NaCl) is used to dissociate the protein from RNA [5].

Additionally, when imaging, as oxygen is a notorious source of photobleaching, it is almost always mandatory to use as an "imaging buffer" the "working buffer" supplemented with an oxygen scavenging system—we routinely use 0.8%(w:v) glucose; 1 mg/mL glucose oxidase; 0.05 mg/mL catalase (all final concentrations)—and a triplet state quencher such as the vitamin E analog Trolox. Trolox can be added at final concentration of 1.5–2 mM (Trolox stock is prepared at 200 mM in dimethyl sulfoxide DMSO).

3.2 Use of Fluorescently Labeled Protein

Labeled proteins can also be efficient tools to observe association to RNA (Fig. 3b). In this case, different approaches are available for protein labeling.

We will present here how to label a protein with a small organic fluorophore, such as a cyanine dye.

3.2.1 Choice of Labeling Position

Mainly, two approaches can be used (1) site-specific conjugation on a cysteine; this can be achieved with a maleimide-ester dye that forms a covalent bond with the cysteine sulfhydryl group. Maleimide-linked cyanine dyes are easily available from different companies, such as GE Healthcare. (2) Conjugation on a lysine; amine-reactive conjugates such as NHS-esters.

The advantage of cysteine residue labeling is specificity as Cys is not an abundant amino acid residue, while lysine residues are frequent in proteins, especially in proteins that interact with nucleic acids; therefore Lys labeling is likely to result in binding many fluorophores. Note that cysteine can be absent in the protein of interest and that it then may be necessary to mutate a surface-accessible serine or alanine into a cysteine to label the protein. This is, for instance, the case for the sRNA cofactor Hfq, where a cysteine has to be introduced. Position of labeling may vary, but two successful positions for cysteine labeling have been reported: a cysteine residue introduced in position 38 (S38C) [8] or in position 65 (S65C) [9]. Note that depending on the activity measured, one mutant can be preferred: S38 is located on the distal face, involved in binding of DNA and RNA A-rich sequences, while S65 is located on the proximal surface involved in uridine-containing RNA binding.

3.2.2 Example of Hfq S38C Labeling with Cy5

The cysteine mutant plasmid was prepared by substituting Ser38 in wild-type Hfq with a cysteine using the QuikChange mutagenesis kit (Agilent). *E. coli* Hfq (without His_6-tag) can be expressed and purified using a Ni-NTA resin as previously described in [10]. Note that it is usually important to use a Δ*hfq* BL21(DE3) strain to avoid the formation of mixed mutated and non-mutated hexamers. Hfq S38C protein is then labeled with Cy5-maleimide (GE Healthcare) as follows:

1. Prepare "Washing Buffer 1" by supplementing a solution containing 20 mM Tris–HCl pH 8, 20 mM imidazole, and 0.3 M NaCl with a 10 mM TCEP stock solution (final TCEP concentration: 0.1 mM). TCEP is used as it interferes less with the thiol-reactive dye than DTT and β-mercaptoethanol.
2. Dissolve 1 mg of Cy5-maleimide in 30 μL DMSO. Add a two-fold molar excess of Cy5-maleimide to the protein (protein concentration is usually around 1 mg/mL).
3. Incubate overnight on a rotatory shaker at 4 °C.
4. Load Ni-NTA resin with the protein (we prefer using a batch resin rather than a pre-packed column to adapt the ratio resin/protein).
5. Wash with 30 mL of "Washing Buffer 1" (until the solution which is flowing through the column becomes clear).
6. Elute with "Elution Buffer."
7. Process to an electrophoresis gel to check protein labeling. Gel must be imaged with a fluorescence imager, such as a G-Box (Syngene).
8. To evaluate the efficiency of labeling, OD of the sample is measured (the blank is done on "Elution buffer"). The OD is

measured at 650 nm and 277 nm. OD at 277 nm corresponding to the protein is corrected from Cy5 absorption as follows: 5% of A_{650nm} is subtracted from A_{277nm} ($A_{277}^{corr} = A_{277} - 5\%A_{650}$)

[Cy5] concentration is determined using the extinction coefficient of Cy5 ε_{650nm} = 250,000 M^{-1} cm^{-1}

Hfq concentration [Hfq] is determined using the extinction coefficient of hexameric Hfq ε_{277nm} = 23,333 M^{-1} cm^{-1}

The ration of [Cy5]/[Hfq] allows to determine the efficiency of labeling, which is usually around 50%.

3.3 Optical Setup for smFRET Experiments

We use a custom-built objective-type TIRF setup for single-molecule imaging (Fig. 4). The excitation beam is focused on the back focal plane of the TIRF objective. A movable mirror is used to bring the incident beam close to the edge of the objective aperture and cause total internal reflection. The fluorescence light is collected by the same objective, passes a dichroic and an emission filter (placed in the microscope filter cube) before being split with a commercial image splitter. Two separate images corresponding to emissions from the Cy3 and Cy5 fluorescent dyes are formed on a cooled back-illuminated electron-multiplying CCD.

3.4 Analysis with Label-Free Protein

3.4.1 Preparation of Coverslips and Slides for Immobilization of Fluorescent RNA

Polymer-coated glass coverslips are used as imaging surfaces [7]. Coverslips and slides are prepared for capture of biotinylated molecules as follows:

1. Coverslips are cleaned by sonication (15 min) in 1 M KOH and rinsed in milliQ water.
2. Then coverslips are dried and aminosilane reaction is carried out with aminosilane reagent (Vectabond, Vector) in acetone. Submerge coverslips into Vectabond (in acetone) for 5 min. Rinse in acetone and in milli-Q water.
3. Coverslips are then incubated for 3 h with a solution of PEG-NHS (120 mg/mL) including a fraction (~0.25%) of biotin-PEG-NHS ester in 100 mM sodium bicarbonate buffer at pH 8.25 (Fig. 5). PEG forms a passivated layer to prevent nonspecific binding of proteins to the surface.
4. Glass slides must be customized as follows: at least 2 drilled holes (diameter ~1 mm, *see* Fig. 6 for explanations of where holes must be designed) are made in the slide (with your local rotary drill machine).
5. The slides are then cleaned and PEG-passivated following exactly the same protocol (**steps 1**, **2**, and **3**) than for coverslips except that biotin-PEG is not used.
6. Finally, PEG-slides and coverslips are stored (up to a few months) at −20 °C in well-closed Falcon tubes filled with an inert gas, such as argon.

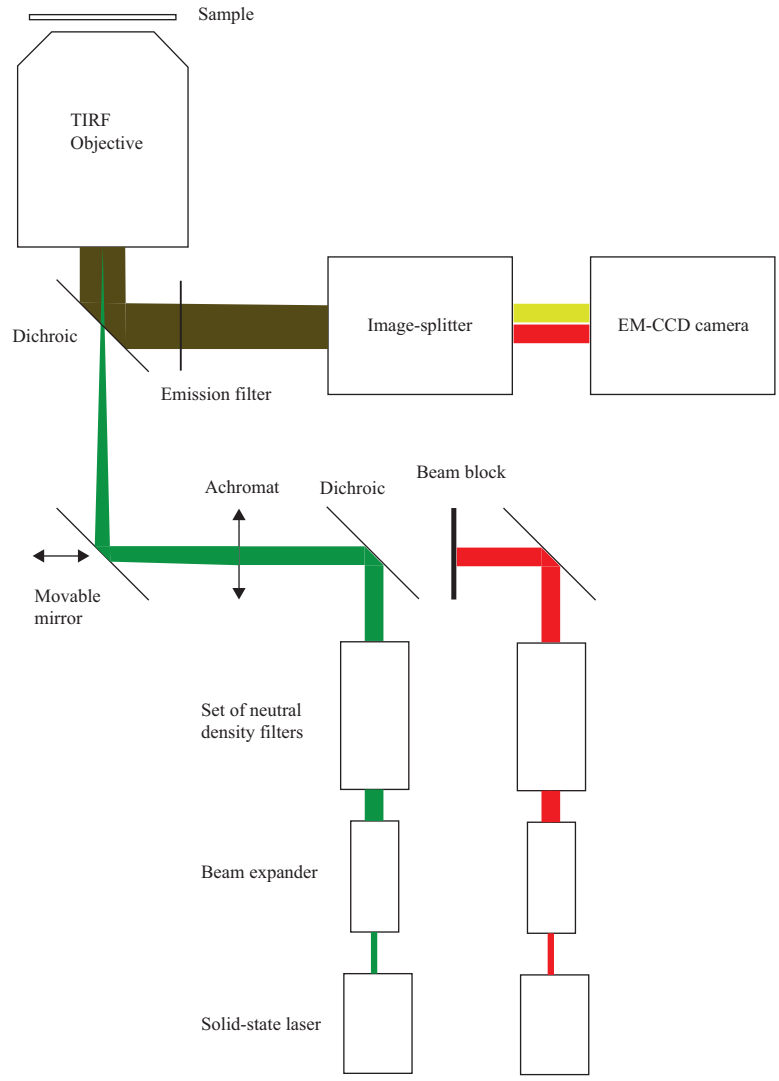


Fig. 4 Schematic view of the setup for smFRET TIRF microscopy. In the main configuration, a green laser is used to excite the donor fluorophores. The corresponding beam path is represented in green. For some control measurements acceptor molecules are directly excited using a red laser. In this case, the beam block is removed and the laser beam shown in red is coupled into the excitation path using a dichroic. In either case, an achromatic doublet focuses the laser beam onto the back focal plane of the water-immersion TIRF objective. The position of the focal point on the objective is adjusted with a movable mirror in order to reach total internal reflection mode with the TIRF objective, assuring that only a sheet of about 100 nm depth above the bottom coverslip is excited inside the sample chamber. The same objective collects the fluorescence light emitted by the fluorophores located within this sheet. The light is reflected by a dichroic, passes an emission filter exhibiting very efficient blocking of the excitation wavelengths before entering the image-splitter. The image splitter device generates two distinct images, where the first one arises from emission of the donor fluorophores and the second one from emission of the acceptor fluorophores. These two images are formed on two laterally separated regions of the back-illuminated chip of a cooled electron-multiplying charged-coupled device (EMCCD) camera exhibiting single-photon sensitivity

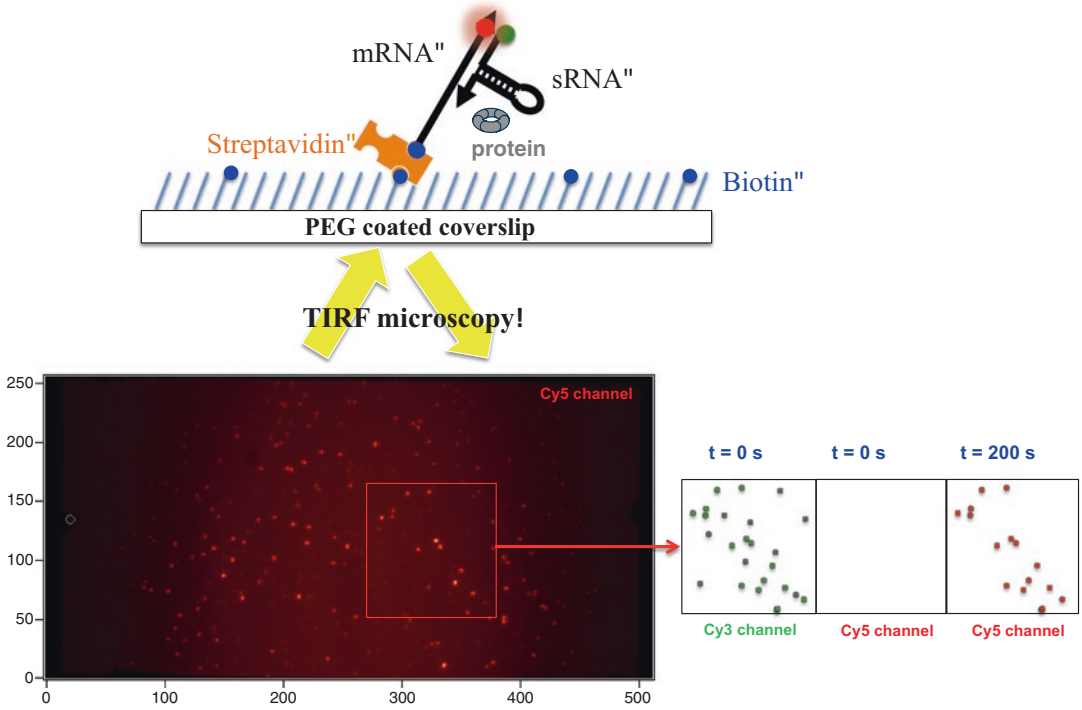


Fig. 5 Single-molecule FRET experimental setup. In each experiment, one RNA is immobilized on the surface thanks to the presence of biotin. Then, other molecules in solution are injected in the sample chamber. FRET can be observed in parallel for hundred of molecules. Single-molecule kinetics as that shown in Fig. 1b can be recorded

3.4.2 Flow Chamber and Sample Preparation

A PEG-passivated glass slide with drilled holes is customized for flow in- and output. Typically, tubing is glued and sealed above each hole using the Kwik-cast sealant (*see* Fig. 6 as an example). This allows flow in of samples and buffers and removal of waste solutions.

1. Just prior to the fluorescence measurements, a biotin-PEG coverslip is affixed to the slide using double-sided tape—typically one or a few injection channels of ca. 10–50 μL can thus be designed, with the corresponding number of holes in the slide (Fig. 6).
2. The slide and coverslip are sealed using wax or commercial sealant (Kwik-cast sealant, WPI).

   From this point, all experiments are performed at room-controlled temperature (typically 20 or 25 °C).

   Successively inject into the sample chamber(s):
3. 150–200 μL of "washing buffer 2."
4. 50–75 μL of streptavidin (0.2 mg/mL) in "washing buffer 2." Incubate for 5–10 min at room temperature.
5. 150–200 μL of "washing buffer 2" to rinse out excess streptavidin.

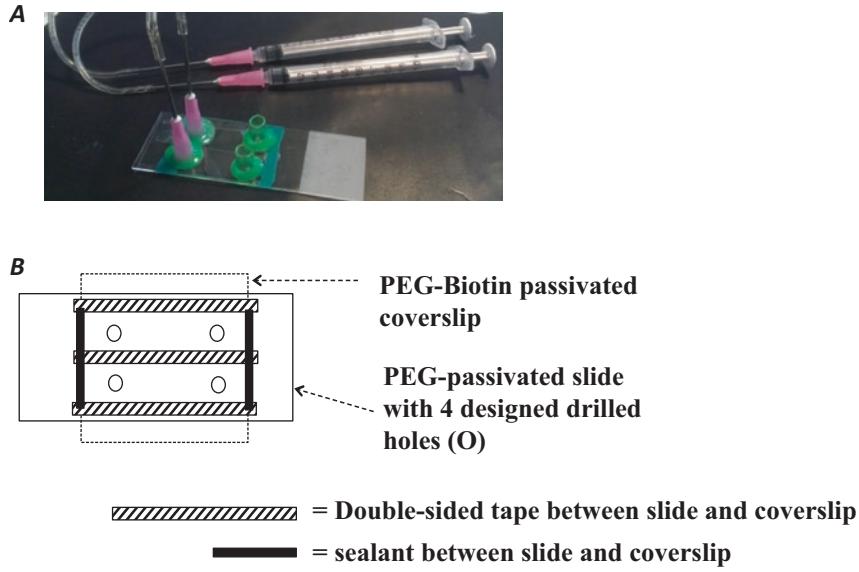


Fig. 6 Preparation of flow chamber. (**a**) Picture of one of our typical flow chamber: buffer/samples are injected with a pipet into the customized yellow cones sealed to the slide just above the 2 right drilled holes; the solution is aspirated into the flow chamber using the hand-manipulated syringes connected by teflon tubing to the cut needles sealed to the slide just above the 2 left drilled holes. An important practical point while aspirating buffers/solutions into the flow chamber is to avoid at all costs to introduce bubbles into the chamber! (**b**) Schematic layout for a 2-injection compartments flow chamber

6. 50–75 μL of biotinylated RNA solution (10–100 pM—*see* **Note 1**) in "working buffer." Incubate for 10 min at room temperature.
7. 150–200 μL of "working buffer" to rinse out excess fluorescent RNA molecules.
8. (When relevant) 50–75 μL of protein and/or complementary RNA in "working buffer" are injected and incubated (necessary time must be adjusted). Afterward, the flow chamber is rinsed with 150–200 μL of "working buffer" to remove biomolecules in excess.

3.4.3 Visualization of Molecules

In order to be able to observe individual fluorescent molecules during a sufficiently long time (typically one or a few minutes) to allow analysis, the visualization buffer should be "imaging buffer," i.e., "working buffer" supplemented with an oxygen scavenging system and a triplet state quencher—as described previously.

Proceed as follows:

1. Mix all the components of the "imaging buffer" just prior to the fluorescence measurements, and incubate it in a microtube for 2 min at room temperature (to allow the enzymes to catalyze the removal of most of the oxygen originally present in the solution).

2. Inject into the flow chamber and immediately start the fluorescent measurements (*see* **Note 2**).
3. Carefully install the flow chamber on the sample frame of the inverted microscope. A small area (approx. 1000 μm^2) of the sample solution is directly excited when the appropriate laser is switched on (*see* **Note 3**).

3.4.4 Acquisition and Analysis

Concerning controls necessary to validate the quality of the collected smFRET data (*see* **Note 4**).

1. A custom-written program in LabView (available on request) is used to set data acquisition parameters and to collect raw data. Typically, each image on the CCD camera corresponds to an acquisition time of 60 ms (lower limit 30 ms, limited by CCD camera deadtime). Gain value is set to optimize S/N ratio. We usually use a gain of 300.
2. Data analysis is performed using a custom-written program in Igor (available on request). First, donor and acceptor channels are aligned, including any translation, rotation, or distortion between the 2 channels. Second, isolated spots, mostly corresponding to single fluorescent molecules, are detected by the software in donor (or acceptor) channel—and spots corresponding to the same molecule in acceptor (or donor) channel are automatically mapped and isolated (Fig. 5). Third, spot intensities are summed on a user-defined area (we typically use a circle with diameter of 9 pixels—1 pixel corresponds to 16 μm on our camera) and background is subtracted from the area surrounding the spot (here again user-defined). Too closely localized spots are rejected (again, the rejection criterion is user-defined).
3. Fluorescent intensity vs. time trajectories for individual spots are extracted from all images in both donor and acceptor channels. The fluorescent intensity trajectories are then converted into FRET trajectories as a function of time using equation: $E = \mathrm{Ia}/(\mathrm{Id} + \mathrm{Ia})$.
4. Apparent FRET efficiency trajectories are analyzed by hidden Markov chain modeling. We typically use the software vbFRET for this analysis [6], which allows to extract discrete FRET states from the data containing noise from various sources.
5. From raw trajectories or from vbFRET idealized trajectories, custom-written program in Matlab are used:
   (a) To plot histograms of apparent FRET efficiency for any subset or for all molecules analyzed (typically in this case, ca. 10 short movies, 5–10 s each, taken in different areas of the sample are used for the analysis)
   (b) To compute dwell time statistics of the single molecules in the different idealized FRET states.

3.4.5 Observation of RNA/Protein Interaction by PIFE

Protein binding to cyanine-labeled RNA substrates can often be monitored by fluorescence intensity increase (Fig. 6). This effect is named protein-induced fluorescence enhancement (PIFE) and consists in the increase of fluorescence intensity when the protein binds a cyanine probe [11]. This typical enhancement of cyanine fluorescence occurs when the dye is sterically constrained, for instance, in a more viscous environment (solvent) or, as this is the case here, by being confined in a protein-binding site. PIFE can be observed with analysis in bulk (Fig. 7a) or at the single-molecule level (Fig. 7b).

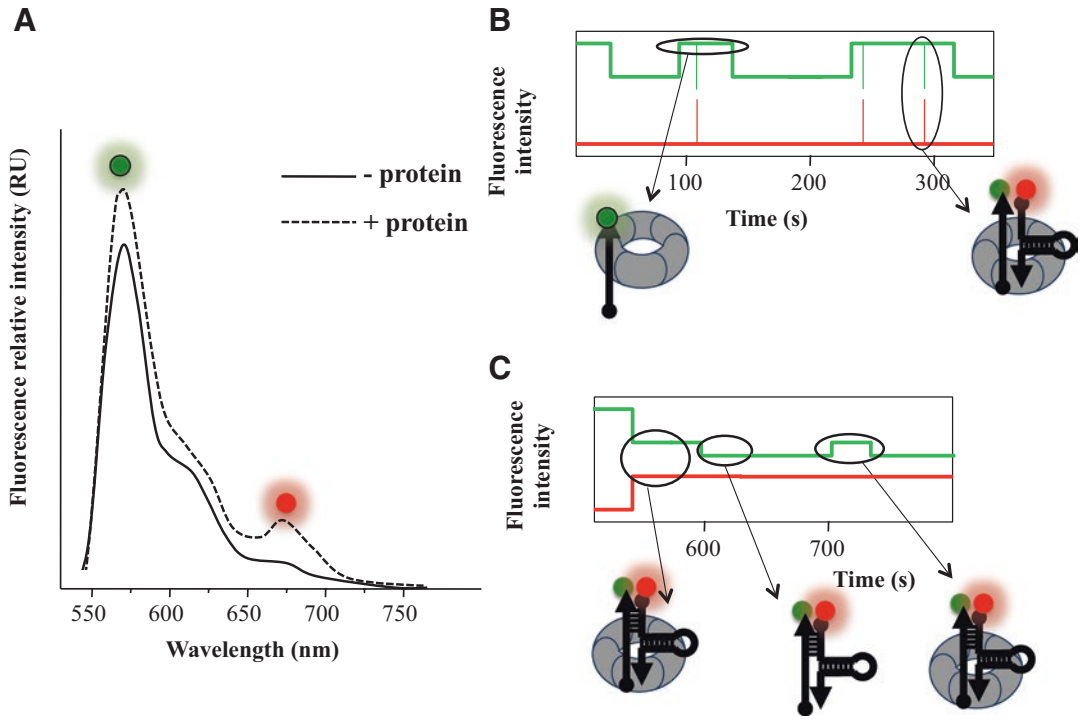


Fig. 7 Bulk and smFRET observation of RNA annealing. FRET is observed in the presence of a protein. In this case the protein is not fluorescent, while sRNA is labeled with Cy5 and mRNA with Cy3. Cy3-mRNA is immobilized while Cy5-sRNA and protein are in solution. (**a**) Bulk analysis. Increase of Cy3 emission (~565 nm green label) reflects the interaction of the protein with the cyanine dye (PIFE), while FRET is indicated by an increase in fluorescence emission at ~670 nm (red label). (**b**) The same at the single-molecule level. Increase of Cy3 emission (green trace) reflects the interaction of mRNA with the protein. Simultaneously, transient increases of Cy5 emission (red bursts) occur, reflecting brief FRET between Cy3 and Cy5 (i.e., sRNA and mRNA interaction). As seen, sRNA/mRNA complex is not stable in the first part of the kinetics (0–350 s) and is probably just due to protein simultaneous binding of sRNA and mRNA, without pairing between RNAs. (**c**) Then after 500 s, stable increase of Cy5 emission reflects annealing and sRNA/mRNA base pairing. At 600 s, decrease of Cy3 emission reflects the dissociation of the protein from the complex. At 700 s, new increase of Cy3 emission reflects the transient re-binding of the protein to the complex

3.4.6 Measurement of k_{-1} Dissociation Kinetic Constant by Dwell Time Analysis from smPIFE Experiment

Statistics on immobilized cyanine-RNA with increased emission due to PIFE also allows measurement of amounts of individual protein-bound RNAs as a function of time (Fig. 3c). From these data, a dissociation-time histogram can be obtained, allowing the measurement of k_{-1} dissociation kinetic constant (Fig. 8). k_{-1} can be obtained by simple exponential curve fitting ($F = F_0 e^{-k_{-1} \cdot t}$).

3.4.7 Observation of sRNA/mRNA Annealing

In the setup used for Fig. 5, Cy3-mRNA is immobilized while Cy5-sRNA and cofactor are in solution. Single-molecule imaging allows visualization of sRNA/mRNA interaction in real time. If sRNA and mRNA interact transiently, one will observe increases of Cy5 emission as spikes; if they form a stable duplex, then one will observe stable increases of Cy5 emission.

In this chapter, we have focused on a qualitative use of smFRET to evaluate the proximity (or not) of fluorescent dyes—a very useful way to unravel conformational changes/binding of the macromolecules labeled with the dyes (Fig. 9). It should be noted that, in principle, smFRET can also be used in a more quantitative way: that is, to evaluate precisely (in a certain range) the distances between the dyes, which allows to quantify parameters of the three-dimensional

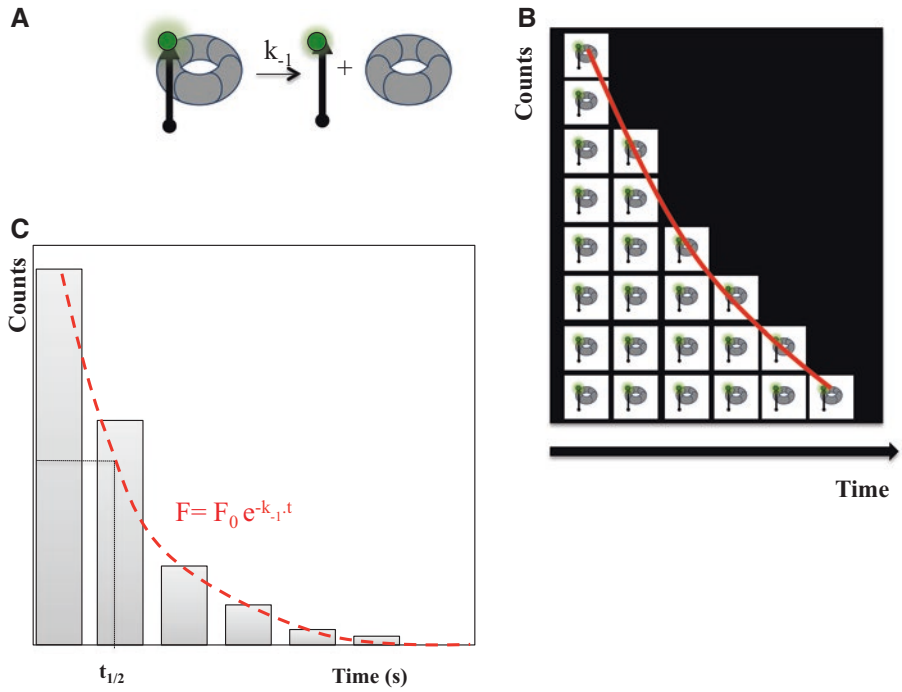


Fig. 8 Dissociation-time histogram. Statistics of protein-bound RNAs in a population as a function of time allows obtaining dissociation-time histograms, in order to measure k_{-1} dissociation kinetic constant of the complex (fitting on simple exponential $F = F_0 e^{-k_{-1} \cdot t}$). From k_{-1}, half-life of the complex can be calculated as $t_{1/2} = \ln 2 / k_{-1}$

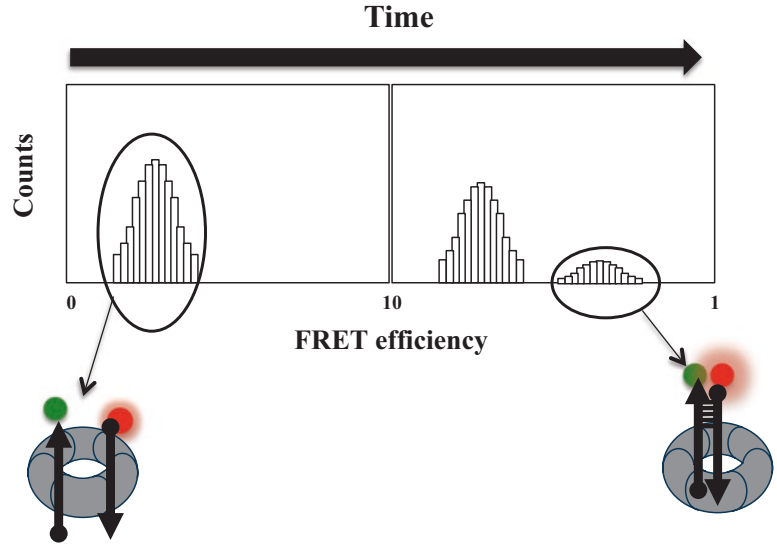


Fig. 9 smFRET efficiency and distance: Higher FRET efficiency means Cy3 and Cy5 are closer. FRET efficiency allows to estimate the proximity between RNAs and thus to evaluate the efficiency of base pairing

structures of the macromolecules. To achieve such an objective, the technique described in this chapter is not optimal [12] and the interested reader is referred to adequate methodological improvements of the smFRET technique which allow precise measurements of distances—such as, for example, the "ALEX" methodology, which uses alternated laser excitation of dyes [13, 14].

3.4.8 Observation of Strand Exchange

Finally three-color FRET allows observing strand exchange (*see* **Note 5**). This three-color FRET procedure can be particularly useful to follow the replacement of one sRNA by another (Fig. 10a), or in order to follow opening of a stem loop by an sRNA (Fig. 10b).

4 Notes

1. These RNA concentrations are usually adequate to obtain a good density of fluorescent single molecules: typically, 100–200 single-molecule spots are observed on the visualization area (~1000 μm² in our case). If single-molecule density is too high or too low, consider first adjusting the RNA concentration.
2. smFRET measurements usually cannot exceed ca. 30 min with the same visualization solution—this is because the oxygen scavenging system renders it too acidic with time. To avoid this problem, one can simply reinject fresh visualization solution after that lapse of time.

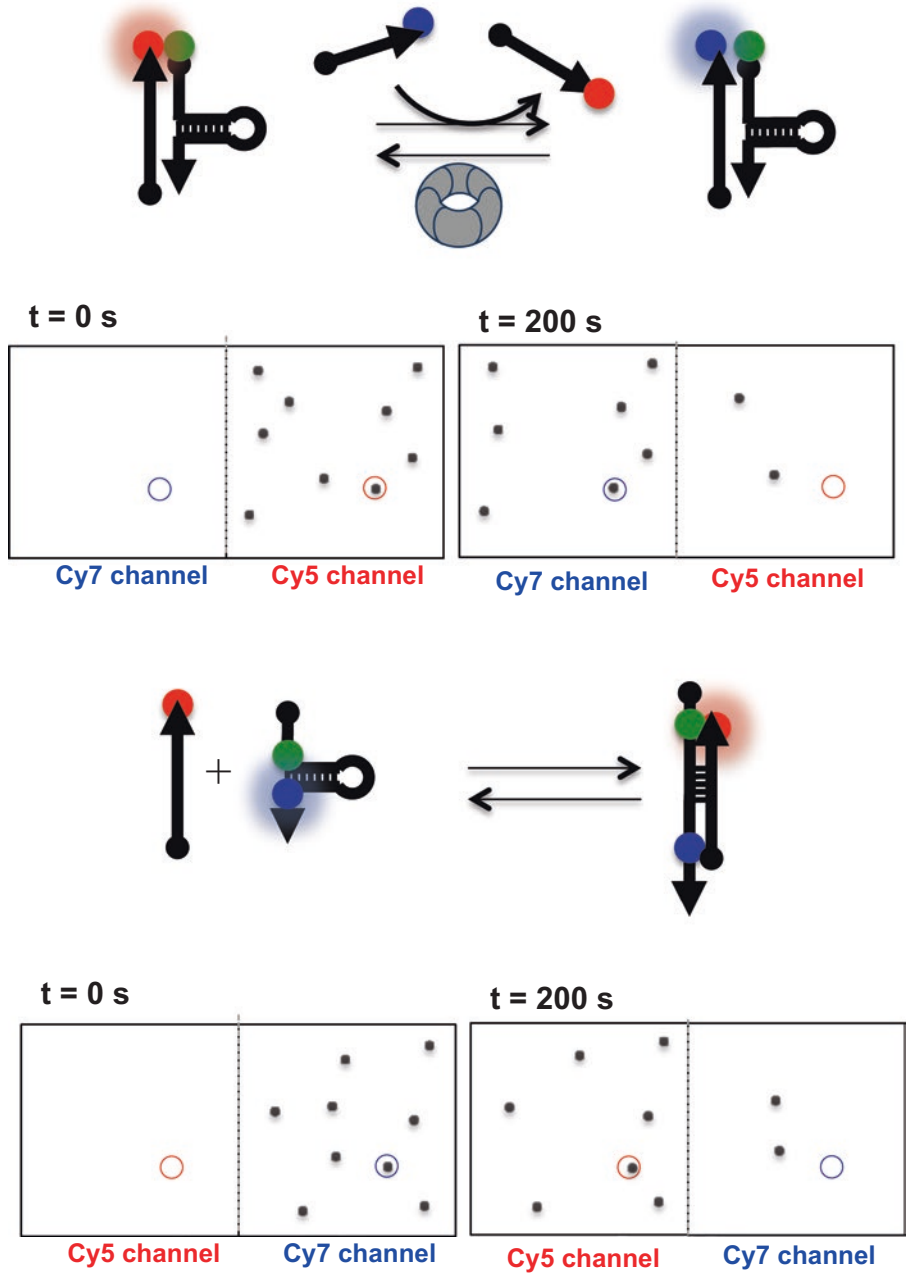


Fig. 10 Kinetics of RNA strand exchange. Labeling of RNA with FRET trio allows observing strand exchange (**a**) or opening of a stem loop by an sRNA (**b**).

3. When most fluorescent spots in the area directly excited by the laser are bleached—after 1 or 2 min—it is recommended to move to a neighboring "fresh" area in order to go on with the fluorescence measurements.

4. Some controls are mandatory to validate the quality of the smFRET experiments. At the minimum, we routinely perform two of them:
   (a) Check background fluorescence of the coverslips lot—i.e., all steps described in Subheading 3.4.2 are applied to a test coverslip except that the biotinylated fluorescent oligonucleotide is omitted; in these conditions, one should see less than 10 fluorescent spots in the excited area. If this is not the case, a new lot of passivated coverslips should be prepared with fresh chemicals.
   (b) Check nonspecific binding (adsorption) of fluorescent molecules present in solution—typically, fluorescently labeled proteins if they are used. The same procedure is applied than above; this time, the fluorescent molecule in solution is added at its maximum investigated concentration. And, once again, less than 10 fluorescent spots in the excited area should be observed. If not, consider changing buffer conditions (adding BSA, detergents, etc.) and/or coverslip treatment (cf. [7] for a good discussion on this point) to minimize unspecific adsorption.
5. Three-color FRET can sometimes be considered as pseudo three-color FRET. In some cases, the same donor can potentially transfer to both acceptors simultaneously—which requires controls and calibration—while in other cases (pseudo three-color), the two acceptors are mutually exclusive, so at a given time, one indeed observes two-color FRET.

Acknowledgments

This work was supported by the CNRS, CEA, ESPCI, and University Paris Diderot. Ulrich Bockelmann acknowledges support by the Human Frontier Science Program [RGP008/2014]. We are particularly grateful to Emmanuel Margeat and Erwin Peterman for help with setup design and fruitful discussions, to Rahul Roy, Sungchul Hohng, and Wonseok Hwang for many fruitful discussions and to Jingyi Fei for critical reading of the manuscript.

References

1. Tan YW, Hanson JA, Chu JW, Yang H (2014) Confocal single-molecule FRET for protein conformational dynamics. Methods Mol Biol 1084:51–62
2. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Mol Cell 9:11–22
3. Smirnov A, Wang C, Drewry LL, Vogel J (2017) Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. EMBO J 36:1029–1045
4. Arluison V, Hohng S, Roy R, Pellegrini O, Regnier P, Ha T (2007) Spectroscopic observation of RNA chaperone activities of Hfq in post-transcriptional regulation by a small

non-coding RNA. Nucleic Acids Res 35: 999–1006

5. Hwang W, Arluison V, Hohng S (2011) Dynamic competition of DsrA and rpoS fragments for the proximal binding site of Hfq as a means for efficient annealing. Nucleic Acids Res 39:5131–5139
6. Bronson JE, Fei J, Hofman JM, Gonzalez RL Jr, Wiggins CH (2009) Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. Biophys J 97:3196–3205
7. Roy R, Hohng S, Ha T (2008) A practical guide to single-molecule FRET. Nat Methods 5:507–516
8. Rabhi M, Espeli O, Schwartz A, Cayrol B, Rahmouni AR, Arluison V, Boudvillain M (2011) The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. EMBO J 30: 2805–2816
9. Hopkins JF, Panja S, Woodson SA (2011) Rapid binding and release of Hfq from ternary complexes during RNA annealing. Nucleic Acids Res 39:5193–5202
10. Taghbalout A, Yang Q, Arluison V (2014) The Escherichia coli RNA processing and degradation machinery is compartmentalized within an organized cellular network. Biochem J 458: 11–22
11. Hwang H, Myong S (2014) Protein induced fluorescence enhancement (PIFE) for probing protein-nucleic acid interactions. Chem Soc Rev 43:1221–1229
12. Lee NK, Kapanidis AN, Wang Y, Michalet X, Mukhopadhyay J, Ebright RH, Weiss S (2005) Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. Biophys J 88:2939–2953
13. Kapanidis AN, Lee NK, Laurence TA, Doose S, Margeat E, Weiss S (2004) Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. Proc Natl Acad Sci U S A 101:8936–8941
14. Margeat E, Kapanidis AN, Tinnefeld P, Wang Y, Mukhopadhyay J, Ebright RH, Weiss S (2006) Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. Biophys J 90: 1419–1431

Chapter 18

Techniques to Analyze sRNA Protein Cofactor Self-Assembly In Vitro

David Partouche, Antoine Malabirade, Thomas Bizien, Marisela Velez, Sylvain Trépout, Sergio Marco, Valeria Militello, Christophe Sandt, Frank Wien, and Véronique Arluison

Abstract

Post-transcriptional control of gene expression by small regulatory noncoding RNA (sRNA) needs protein accomplices to occur. Past research mainly focused on the RNA chaperone Hfq as cofactor. Nevertheless, recent studies indicated that other proteins might be involved in sRNA-based regulations. As some of these proteins have been shown to self-assemble, we describe in this chapter protocols to analyze the nano-assemblies formed. Precisely, we focus our analysis on *Escherichia coli* Hfq as a model, but the protocols presented here can be applied to analyze any polymer of proteins. This chapter thus provides a guideline to develop commonly used approaches to detect prokaryotic protein self-assembly, with a special focus on the detection of amyloidogenic polymers.

Key words Protein self-assembly, Noncoding RNA cofactor, Functional amyloid

Abbreviations

AFM	Atomic force microscopy
CF	Curve fitting
FTIR	Fourier transform infrared spectroscopy
SAXS	Small angle X-ray scattering
SRCD	Synchrotron radiation circular dichroism
TEM	Transmission electron microscopy
ThT	Thioflavin T

1 Introduction

In prokaryotes, the components of RNA metabolism self-assemble into complex structures, resulting in functional compartmentalization within the cell [1–3]. The RNA maturation components

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_18,

include proteins involved in post-transcriptional genetic regulation such as ribonucleases (RNAses) or RNA chaperone, but also small regulatory noncoding RNAs (sRNA) [2, 4, 5]. The best-characterized protein involved in such pathways using regulatory sRNA is probably the RNA chaperone Hfq [6]. Nevertheless, some others have also been identified such as ProQ and the archaeal Sm proteins [7, 8]. Our recent studies demonstrated that specific sequences in the *E. coli* Hfq C-terminus region (CTR) are able to polymerize, resulting in the formation of amyloid fibers [9]. Indeed, increasing evidence shows that amyloids are not only the result of protein misfolding associated with neurodegenerative diseases, but also found in cells for useful reasons [10]. In the later case, amyloids contribute to the physiology of the cell and are referred as *"functional"* [10]. Indeed, these functional amyloids have been widely reported in different organisms, including bacteria. In prokaryotes, diverse functions have been described for these amyloids, such as a role in biofilm development [11], formation of curli and pili [12], or creation of pores in the host membrane [13]. In this chapter we intend to provide a guideline to develop approaches that can be used to detect the fibers of sRNA protein cofactors, precisely molecular imaging, such as atomic force microscopy and transmission electron microscopy, infrared spectroscopy, synchrotron radiation circular dichroism, and small angle X-ray scattering. Our chapter will focus on Hfq and its amyloid nature as an example, but it could be applied to any self-assembling prokaryotic proteins.

2 Materials

2.1 Preparation of Samples

1. Synthetic peptide: 20 mg/mL in deionized RNA-grade water (*see* **Note 1**).
2. Purified Hfq/Sm proteins, prepared as described in the Chapter 15 by Stanek and Mura. Depending on buffer used and protein concentration reached during the purification process, the same limits apply as those described in **Note 1**.
3. UV spectrophotometer suitable for the precise determination of protein concentration measurements (micro-volumes).

2.2 Thioflavin Staining

1. ThT stock solution: 0.8 mg/mL in 10 mM phosphate buffer pH 7 containing 150 mM NaCl.
2. Fluorescence microscope equipped with a digital camera and emission/excitation filter sets to permit visualization of ThT staining.

2.3 AFM Imaging

1. AFM microscope with liquid cell.
2. AFM tips with different force constants: 2–75 N/m and resonant frequencies in the range of 100–320 kHz for imaging in

air or in the range of 0.01–0.8 N/m and resonant frequencies in the range of 70–100 kHz for imaging in solution.

3. AFM-grade mica.
4. AFM-free image analysis software such as WsXM (http://www.wsxmsolutions.com/) and Gwyddion (http://gwyddion.net/).

2.4 TEM Imaging

1. Cryo-transmission electron microscope equipped with a slow scan CCD camera.
2. Carbon-coated electron microscopy copper grid, CF200-Cu (200 mesh) or CF300-Cu (300 mesh), EMS.
3. Carbon-coated electron microscopy copper grid, ref. 01881-F (200 mesh) or 01883-F (300 mesh), Ted Pella.
4. Uranyl acetate solution: 2% (w/v).
5. Uranyl-less solution containing Gadolinium salts, Delta Microscopies.
6. Plasma cleaner.
7. Leica EM-CPC plunge-freezing device for cryo-fixation.

2.5 FTIR Spectroscopy

1. Interferometer-based spectrophotometer coupled with an IR source.
2. For transmission measurements, liquid cell equipped with CaF_2 windows and a 6 μm-thick polytetrafluoroethylene (Teflon) spacer (for example, Omni-Cell transmission Cell from Specac Company or Demountable liquid Cell from Piketech).
3. For reflection measurements, attenuated total reflection (ATR) sampling setup.

2.6 SRCD Spectroscopy

1. SRCD endstation such as that of DISCO beamline at SOLEIL synchrotron [14, 15].
2. Manually loaded circular demountable CaF_2 cells of 30 μm path length with a loading volume of 2 μL [16].
3. CDtool software for data acquisition and treatment, including averaging, smoothing, subtraction of buffer baselines from sample-spectra, calibration ((+)−camphor-10-sulfonic acid CSA), and normalization [17].
4. BestSel open access software (http://bestsel.elte.hu/) [18].

2.7 Small Angle X-Ray Scattering (SAXS)

1. SAXS beamline such as SWING at SOLEIL.
2. Quartz capillaries of 1.5 mm diameter and 0.01 mm thickness.
3. Foxtrot software (http://www.synchrotron-soleil.fr/Recherche/LignesLumiere/SWING).

3 Techniques

3.1 Analysis of Hfq Self-Assembly by Molecular Imaging

Characterization of Hfq self-assembly can be performed by microscopy techniques, including fluorescence microscopy, transmission electron microscopy (TEM), and atomic force microscopy (AFM).

3.1.1 Thioflavin Staining

Thioflavin T (ThT) is a benzothiazole salt commonly used to visualize amyloids. When bound to aggregated β-sheets, such as those found in amyloids, the dye displays an enhanced red-shifted fluorescence emission spectrum [19]. Even if the dye fluorescence is not completely specific and may bind to other structures (such as double-stranded DNA), it is however a good indicator that amyloid structures are formed in the sample. Practically, for ThT staining:

1. Dilute protein samples to 0.2 mg/mL.
2. Prepare a stock solution of ThT. This solution can be filtered and stored in the dark for 1 week. Before use, the stock solution is diluted in 10 mM phosphate buffer pH 7 containing 150 mM NaCl (dilution 1/50 = working solution); this must be done extemporaneously.
3. For a spectroscopic assay, add a 10 µL aliquot of protein to 1 mL of working solution, incubate for 1 min, and measure the fluorescence intensity with excitation set at 440 nm and emission detected in the range 450–500 nm (*see* **Note 2**).
4. For slide staining and visualization under a microscope, place 10 µL aliquots of stained samples and image using a fluorescence microscope with a 60× oil-objective (NA = 1.4). The laser excitation can be set at 415 or 450 nm and emission is detected at 482 nm.

3.1.2 AFM Imaging

Atomic force microscopy (AFM) has become a common tool for the analysis of protein aggregation and fibrilization [20]. It provides height information that complements structural information from two-dimensional projection images obtained using electron microscopy (*see* Subheading 3.1.3). In essence, an AFM setup consists of a microtip attached to a cantilever held on a piezo scanner that moves along the surface of the sample with high precision over a defined distance within the micrometer range. The deflection of the cantilever follows the surface topography and is monitored following the position of a laser spot reflecting from its surface on a photodiode. The motion of the cantilever can also be used to interrogate the mechanical properties of the sample. AFM imaging thus provides information about biological objects dimensions at nanometer resolution [21]. As amyloid fibrils are usually in the µm range in length with a lateral dimension between 3 and 30 nm, AFM is thus perfectly adapted to probe fibrils structure.

1. Substrate preparation. In order to study fibrillar protein structures, it is essential to first adsorb them to a flat surface.

This step is crucial for obtaining high-quality data. The fibers formed in solution have to be at an adequate density and firmly attached to allow the tip to scan over them. One of the surfaces that is most frequently used for AFM studies is mica, although clean glass or annealed gold can also be used. Here we will concentrate on describing the adsorption of fibrils to mica surfaces. Mica is convenient because it is atomically flat and can be easily cleaved immediately before use to obtain a clean surface. AFM-grade mica can be easily obtained from several AFM product suppliers. The freshly cleaved mica is glued to the sample holder using, for example, double-sided tape before incubating the protein sample.

2. Protein adsorption. A protein solution is incubated on the mica surface to allow for fibril adsorption to the surface (*see* **Note 3**). The size, charge, and mechanical stability of the fibers determine the details of the preparation protocol. Basically this protocol consists of two steps: first, conditioning the surface where the fibers will be deposited and second, incubating the protein solution to allow for filament deposition. As the surface/protein interaction will be affected by the surface charge of the protein and surface, the pH of the solution has to be selected to optimize this interaction. Precisely, mica is constituted of sheets of octahedral hydroxyl-aluminum lying between two silicon tetrahedral layers. The most favorable cleavage plane is located in the layer of unbound K^+ that lies in the interlayer space between neighboring silicon layers. When immersed in water, the hydrated potassium ions can be dissociated from the mica surface, leaving negatively charged aluminum tetrahedra (AlO_2^-) in the first outer layer [22]. Thus, the solution pH and ionic salt concentration should be adjusted to allow for the protein to be positively charged to promote its interaction with the negatively charged mica surface.

3. Imaging in air. If the protein binding to the surface is not strong enough to allow imaging with the sample immersed in solution, the best alternative is to image in air. Incubate a small amount of diluted protein (typically 10 μL of 5–10 μg/mL fibrils) on the mica surface and allow to air dry at room temperature. All the fibers in the solution will now be adsorbed on the mica surface. Before imaging, the sample must be extensively washed with Milli-Q water to remove the salt from the buffer and then carefully air-dried again. AFM can operate in different modes, and the ones more conveniently used to image amyloid self-assemblies are contact mode and tapping mode, depending on the characteristics of the fibrils of interest [21]. Contact mode images are obtained when the tip is in contact with the sample, so it requires that the fibers are firmly attached to the surface and rigid enough to withstand the tip

pressure as it scans along the surface without deformation or sliding. In tapping mode, the cantilever is oscillated at its resonant frequency during the scan, so that it does not drag the sample along as it moves over the surface. Resonant frequencies of the soft cantilevers used for biological materials are, typically, within the 70–100 KHz range, depending on whether the microscope is operated in air or in solution, so that the interaction time between the tip and sample is reduced dramatically to only a few microseconds. Fibers can be imaged either in tapping or in contact mode, depending on the characteristics of the sample.

4. Imaging in solution. If the fibers are sufficiently long and their interaction is strong enough to allow for a stable adsorption to the surface, the fibers will remain firmly attached during imaging in solution. After allowing 10–60 min of incubation, the sample is extensively rinsed with buffer to remove excess protein. The imaging buffer is selected to optimize the interaction between the tip and the sample and could differ from the one selected for fiber adsorption to the mica. The pH or salt concentration can be adjusted again so that the tip–protein interaction is minimized to reduce damaging the fibers during imaging. Operating in solution offers the possibility of imaging fiber growth. In such a case, imaging should be done in the presence of small amounts of protein in the solution to avoid interfering with the imaging process. Images of the same area taken at different times can be used to follow filament growth kinetics.
5. AFM imaging procedure. Once the sample is properly prepared, the AFM is operated selecting the most appropriate tip for the imaging mode. If imaging in air, stiffer tips with force constants in the range of 2–75 N/m and resonant frequencies in the range of 100–320 kHz can be more convenient to reduce the snap-to-contact region. For imaging in solution, softer tips, with force constants in the range of 0.01–0.8 N/m and resonant frequencies in the range of 70–100kHz can be used. In brief, AFM operation includes the following steps: placing the cantilever in the holder and then both to the scanner; aligning the laser to impinge the cantilever tip; aligning the laser reflection on the center of the photodiode; placing the sample in the sample holder; if operating in tapping mode, the cantilever resonance frequency has to be selected; approaching the tip and sample. Once the tip is touching the sample, the scanning parameters, gains, scanning speed, and setpoint, are adjusted to obtain reproducible trace and retrace images.
6. AFM topographic analysis. AFM obtains direct topographical images of the sample surface. Each AFM provider includes software to analyze the images, but there are also free specialized

software available to process scanning probe microscopy images (WsXM or Gwyddion).

Figure 1a–f shows Hfq peptide fibers imaged in air or in solution.

3.1.3 TEM Imaging

A quick way to check the presence and structural characteristics of fibers (length, diameter, interweaving), as well as the homogeneity of the sample, is negative staining and observation by transmission electron microscopy.

Protocol for Negative Staining

1. Sample preparation. Deposit a 5 μL drop of Hfq sample at 1 mg/mL in 50 mM phosphate buffer pH 7 (i.e., stock solution diluted in water) on a glow-discharged carbon-coated electron microscopy copper grid (200 mesh square grid, EMS). In order to optimize the adsorption of the sample to the grid it is recommended, but not strictly required, to realize a glow-discharge (gas ionization process giving the carbon film hydrophilic).

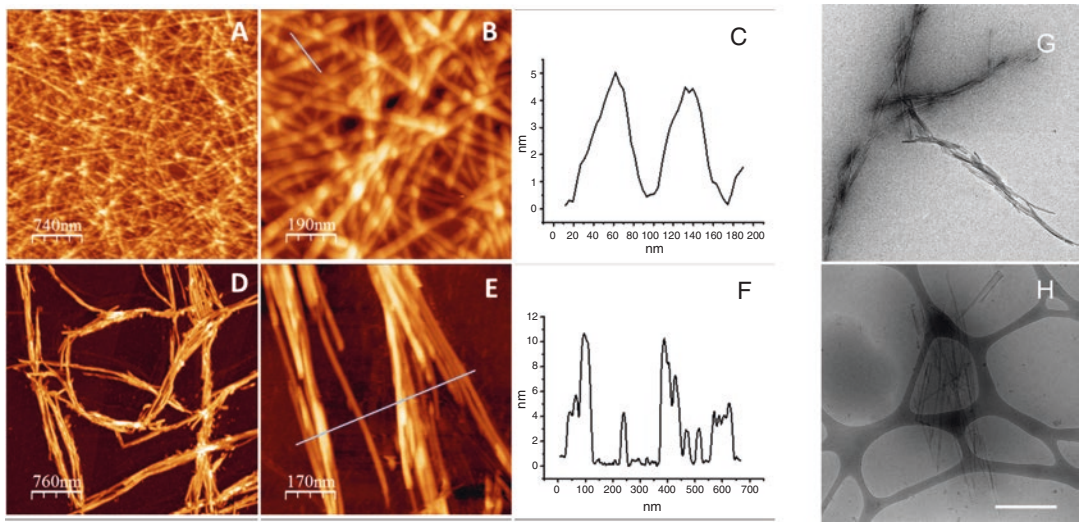


Fig. 1 Visualization of Hfq self-assembly by molecular microscopy. Transmission electron microscopy (TEM) and atomic force microscopy (AFM) are efficient tools to investigate protein nanostructures. (**a**–**f**) AFM images of Hfq peptide fibers in air and under liquid. Upper panels (**a**, **b**, **c**) are fibers imaged in air. A drop of the protein solution was allowed to dry on the mica before imaging in contact mode. Lower panels (**c**, **d**, **f**) show filaments observed under buffer at pH 4 in tapping mode, after incubating the fibers for 5 min on mica at pH 4 at a concentration of 0.5 mg/mL. Panels C and F show the height profile under the blue lines shown, respectively, in panels **b** and **e**. (**g**–**h**) In TEM experiments, 5 μL of Hfq solution is deposited on a glow-discharged electron microscopy grid covered by a continuous carbon film for negative staining (**g**) or a holey carbon film for cryo-TEM (**f**). (**g)** The sample is dried and subsequently stained with a solution of either uranyl acetate or uranyl-less (filling the protein cavities with heavy atoms). (**h**) The excess of sample is blotted and subsequently frozen in ethane at liquid nitrogen temperature (embedding the sample in thin amorphous ice layers). Both methods show filaments that seem more aggregated in negative staining. *Scale bar:* 500 nm

2. Sample adsorption. After 5 min of interaction, blot out the excess sample using a Whatman filter paper (1–3 min is usually enough to have efficient adsorption of the sample onto the grid).
3. Contrasting of the sample. The adsorbed sample can then be put in contact with the contrasting agent. To perform negative staining, 5 μL of uranyl acetate solution (2%) is applied onto the grid-containing sample. After 30 s incubation, the excess of uranyl acetate is blotted out and then the grids can be kept in a dry dark dust-free environment until observation at the electron microscope. Gadolinium salt (uranyl-less) is also used for negative staining. It is used instead of uranyl acetate because it is safer. However, it may create artifacts such as crystals. Store it at cold temperature and use it at room temperature to avoid salt deposition as much as possible.
4. The grids can be stored at room temperature for further observation on any standard transmission electron microscope.
5. Sample observation. The electron microscopy grid is then mounted onto a room temperature-equilibrated holder and subsequently introduced into an electron microscope (JEOL 2200FS, JEOL, Tokyo, Japan).
6. Image acquisition. 2 k by 2 k images are acquired using a Gatan Ultrascan 894 US1000 slow scan CCD Camera (Gatan, Pleasanton, CA, USA) at 40,000× nominal magnification (corresponding pixel size was 0.32 nm).

Figure 1g shows Hfq peptide fibers imaged by TEM after negative staining.

While it is true that negative staining is the tool of choice for quick characterization, it implies the dehydration of the sample, which can lead to artifactual aggregations of fibers or to structural modifications. Therefore, when high-resolution ultrastructural studies are required, samples should be observed in hydrated conditions, which is possible by cryo-transmission electron microscopy (cryo-TEM). Cryo-TEM is based on fast-freezing of the electron microscopy grid containing a thin layer (100–200 nm or less) of the adsorbed sample by quick immersion in liquid ethane cooled-down at liquid nitrogen temperature. This can be achieved by using a lab-made or commercial apparatus. Some commercial apparatuses can be fully automatized so that repetitiveness can be easily achieved when freezing conditions have been set up.

Protocol for Cryo-Electron Transmission Microscopy (Cryo-TEM)

1. Sample preparation. Deposit a 5 μL drop of Hfq suspension on a glow-discharged lacey carbon film grid (300 mesh).
2. Sample freezing. Blot the suspension through the grid for about 1 s using a Whatman filter paper to create thin suspended films in the carbon holes of the grid. Using a Leica EM-CPC, plunge

the grid into liquid ethane cooled-down with liquid nitrogen. Prior to sample freezing, the grid is kept in a liquid nitrogen storage unit until observation at the electron microscope.

3. Transfer to the electron microscope. The grid with the frozen hydrated sample is transferred to a cryo-TEM for observation. During this transfer process, and during observation, the sample should remain under liquid nitrogen to prevent water crystallization, which impedes the observation of the sample. Since the observed sample is fully hydrated, the obtained images reflect the near native structure of proteins and fibers. A simple but detailed description of negative staining and cryo-TEM techniques is described in [23]. Grids are transferred into a JEOL JEM 2200FS cryo-electron microscope equipped with a Ω-energy filter by using a Gatan 914 cryo-transfer system.
4. Image acquisition. Hfq peptide cryo-images can be acquired using a 2 k by 2 k Gatan Ultrascan 894 US1000 slow scan CCD Camera (Gatan, Pleasanton, CA, USA) at 40,000× nominal magnification (corresponding pixel size was 0.32 nm) with an energy window of 20 eV.
   Figure 1h shows Hfq peptide fibers imaged by cryoTEM.

3.2 Analysis of Hfq Amyloid-like Self-Assembly Secondary and Super-Secondary Structure

3.2.1 Infrared (FTIR) Spectroscopy

Fourier transform infrared (FTIR) spectroscopy can be used to analyze protein secondary and super-secondary structures such as amyloid. The peptide bond is an amide group (CONH) that gives rise to seven active absorption bands in the infrared. This method can thus give information on secondary structure (three bands carry information on the secondary structure, but the Amide I band is the most used, see below) and tertiary structure (through the Amide II band). Indeed, the Amide II spectral region studied in the presence of D_2O solution (i.e., Amide II′, *see* **Note 4**) is a marker of the partial opening of the protein and thus of the tertiary structure changes, during which the hydrogens remained within the protein core undergo H-D exchange [24]. The Amide I is most often used for secondary structure elucidation since it has the smallest contribution from peptide side chain absorptions. The Amide I band absorption is mainly due to the C=O stretching of the amide group vibration but this vibration is strongly delocalized over the amide C–N and C–H bonds giving rise to a strong coupling: the trans dipole coupling. The exact frequency of the absorption peak is thus sensitive to the angle of the amide bond, which is dependent on the local structure of the protein domain. α-Helices, β-sheets, β-turns, random coils, and 3–10 helices are each restricted to a different set of angles as seen in a Ramachandran plot, giving rise to amide peaks located at different frequencies. This gives rise to the sensitivity of infrared spectroscopy to the secondary structure

of proteins. Here, we thus focus on the infrared zone in the region of the Amide I.

1. Hfq fibril samples (from 1 to 20 mg/mL) are loaded into liquid cell equipped with CaF_2 windows and a 6 μm thick spacer. This configuration allows studying a sample in solution for several hours (8 on average), and therefore allows for following the aggregation kinetics.
2. FTIR absorption measurements are performed between 1000 and 4000 cm^{-1} through the use of an interferometer-based spectrophotometer coupled with an IR source and a Triglycine sulfate (DTGS) detector. The best way to have good transmission spectra is by performing a measurement in which 256 scans are collected and at least 5 independent measurements are averaged. Spectra at a nominal resolution of 2 cm^{-1} are baseline corrected and background subtracted before analysis. A spectrum of the buffer must be collected using the same experimental parameters for water subtraction. Buffer subtraction is performed by an iterative procedure to remove the signal originating from the water. The association band of water between 1800 and 2400 cm^{-1} is used to compute the subtraction factor. Most commercial software packages propose an integrated subtraction routine.
3. Alternatively, an attenuated total reflection (ATR) sampling setup can be used. With an ATR setup, the infrared beam light is reflected inside an infrared-transparent, high refractive index crystal. The crystal can be single bounce or have a multiple bounce geometry to increase sensitivity. An evanescent wave is created at the crystal–solution interface, which extends into the sample with a penetration depth typically around 1 μm (depending on the wavelength and on the crystal's refractive index). A background spectrum without sample is first recorded. A total of 64 scans are usually averaged. Then, ~ 5–20 μL of fibril sample (1–20 mg/mL) is deposited directly on the crystal and covered with a cap to reduce evaporation. A spectrum of the buffer must also be recorded in the same conditions for latter subtraction.
4. The analysis consists of qualitative evaluation of secondary structure through the combination of 2nd-derivative, Fourier self-deconvolution (FSD), and curve-fitting (CF) analysis [25]. FSD and the 2nd-derivative can be used to find the number and exact positions of the overlapping peaks composing the Amide I band. CF involves fitting a series of model peaks to the experimental data. The best function for peak shape in liquid samples is usually a convolution of Lorentzian and Gaussian bands, the Voigt profile. Overlapping peaks composing the Amide I band are fitted with Voigt profiles with a bandwidth

varying between 15 and 25 cm^{-1}. The goodness-of-fit is generally evaluated by comparing the rest with the noise. The CF, being an iterative algorithm, is pursued until the rest is less than 5 times the noise. The integration of the area of the component bands used to generate the final curve fit can provide an estimate of the relative percentages of secondary structure present. Both qualitative and quantitative analyses can be achieved with dedicated software. Most software packages such as OMNIC™ (Thermo Fisher Scientific) and Opus (Bruker) provide curve-fitting algorithms.

5. The assignment of a peak at a given wavenumber to a given secondary structure should be done according to Table 1. Briefly, a peak near 1645 cm^{-1} is indicative of random coil, 1655 cm^{-1} of α-helix, and 1630 cm^{-1} and 1690 cm^{-1} of β-sheet (*see* **Note 5**). Stronger hydrogen bonding results in a shift to lower wavenumbers and, therefore, amyloid fibrils often have β-sheet peaks below 1630 cm^{-1} [26].

Figure 2a shows Hfq peptide fibers analyzed by FTIR spectroscopy.

3.2.2 Synchrotron Radiation Circular Dichroism SRCD

Circular Dichroism (CD) is a sensitive absorption spectroscopy technique for studying biological samples such as proteins and polynucleotides (DNA and RNA) as well as sugars. Circular left and right polarized light is differentially absorbed, e.g., by proteins due to the excitation of the n-π and π-π* electronic transition of the peptide bonds. Additionally the aromatic side chains produce characteristic absorptions between 210 and 275 nm. Standardized CD spectra can be deconvoluted and used for protein secondary

Table 1
Empirical assignments for Amide I infrared (IR) bands characteristic of proteins [25]

Wavenumber (cm^{-1})	Assignment
1615–1625	Intermolecular β-sheet (in particular vibrations of strongly bound intermolecular aggregated β-strands, including amyloids)
1630–1640	Intramolecular native β-sheet
1640–1650	Disordered random coil
1650–1660	α-Helix
1660–1695	β-turn
1675–1695	Intermolecular β-sheet (in particular antiparallel aggregated β-sheets)

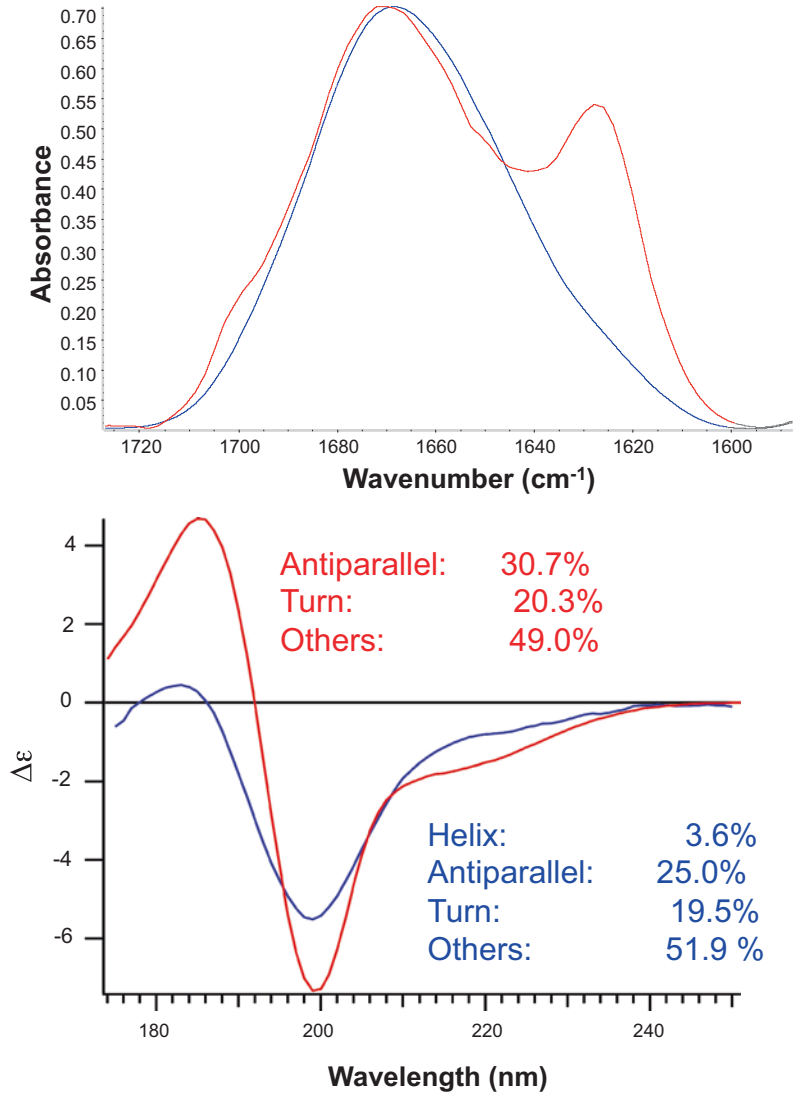


Fig. 2 Secondary and supersecondary structure determination. (**a**) FTIR spectra of Hfq peptide immediately after diluting in water (blue) and after incubation for 1 day at 20 °C (red). The peaks shown in (**a**) are due to intermolecular aggregated β-sheet at 1620 cm^{-1} and to the presence of native β-turn at 1670 cm^{-1} and intermolecular aggregated β-sheet at 1690 cm^{-1}, respectively. In addition, in a lot of proteins studied, the contemporary increase of the components at about 1620 and 1680 cm^{-1} indicates the building of antiparallel β-aggregated structures; the low frequency β-sheet moiety at approximately 1620 cm^{-1} are attributed to multistranded intermolecular β-sheet and resembles the typical fingerprint of the cross-β motif found in amyloid-like fibrils [37, 38]. (**b**) SRCD spectra of Hfq peptide at 20 mg/mL in water (blue) or after incubation at 20 °C for a few days at 100 mg/mL to ensure the formation of fibrils (red). In this case, sonication procedure to fragment fibrils into shorter forms or their monomeric form was not necessary. Secondary structure determination by BESTSEL (inset) revealed a structural increase of 5% in β-sheet content for the fibrils as opposed to the monomeric Hfq peptide

structure determination. Synchrotron radiation circular dichroism (SRCD) allows for the extension of the spectral range down to 168 nm in aqueous solution and 120 nm in hydrated films, with very good signal-to-noise ratios (typically ~20 mdeg/0.2 mdeg). The extension of the spectral range down to 168 nm has improved the information content obtainable from a SRCD spectrum. Especially for the weak CD signals of proteins with high β-sheet or disordered content, SRCD has improved the spectral analysis substantially [27]. Numerous algorithms exist for the estimation of the secondary structure composition from CD spectra. For the special case of β-sheet-rich proteins, with a broad range of protein folds, a recent publicly accessible algorithm BestSel [18] allows the determination of secondary structure contents including parallel and antiparallel β-sheets and potential fold recognition. The novel reference dataset with spectra that significantly differ from present reference sets extends the information content for secondary structure determination.

1. Sample preparation. Homogenized Hfq peptide fibrils were prepared by diluted peptide in water at 20 mg/mL to 100 mg/mL. Observation of fibrils can be made directly after few days (*see* **Note 1**) as described in **step 5**. Nevertheless, in some cases long and associated fibrils need to be fragmented to shorter and individual pieces to be analyzed, as described below (2–4).
2. Seeds can be prepared by a sonication procedure [28]. The sonication strength and time had to be carefully chosen. At low force, homogenization of the fibril solution will just decrease the viscosity by breaking the fibrils into shorter pieces. A microtip sonicator set up in the cold room needs to be regulated to minimal strength (force, amplitude), with a 50% duty cycle (e.g., 1 s pulses interrupted by 1 s silence). Sample volume should be at least 100 μL in microtubes. Sonication shall be carried out with the tubes containing sample cooled in ice-cold solution (for better contact), with the tip plunged right to the bottom of the tube without touching the tube wall. For homogenization of peptides and small proteins four cycles of five pulses with a 10 s rest between cycles produces reliable and reproducible results. During the rest period, mix well the solution containing the peptide by lifting and plunging the tip up and down. Droplet formation on the tube wall should be avoided. Gloves, mouth mask, and glasses are recommended for safety.
3. SRCD spectra are taken before and after sonication allowing the spectral distinction between fibril seeds and amyloid fibrils.
4. Ultracentrifugation in Beckmann TLA-100 rotor run at 53500-96500 Relative Centrifugal Field (RCF) × g for 15–30 min may be used to distinguish eventual monomers floating in the supernatant and the fibrils accumulating in the

pellet. The baseline spectrum for the fibril seeds should be taken from the supernatant. Once the sonication and centrifugation are finished, keep samples at 4 °C.

5. Sample concentration, buffer composition, and pathlength choices for SRCD data acquisition should be handled with care. In general, concentrations should be for α-helix rich proteins 1–2 g/L, for β-sheet rich proteins 3–4 g/L, and for unordered structures above 5 g/L in a 20 μm pathlength cell. Loading volumes are 2 microliters in special 20 μm CaF_2 cells [16] (Fig. 3). The choice of buffer is a function of how much chloride will be considered essential (10–50 mM NaCl allows obtaining spectra down to 185 nm). Chloride is readily replaced by fluoride allowing twice the initial chloride concentration. Ideally phosphate buffer is used being more temperature stable than Tris-Cl and being deep-UV transparent.

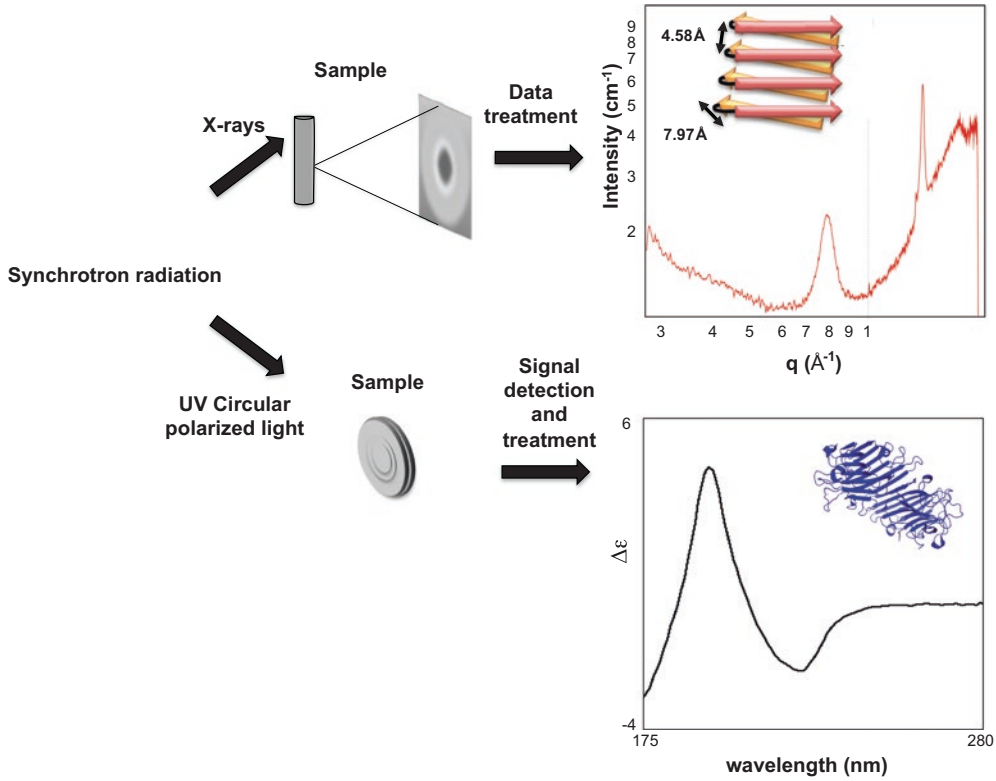


Fig. 3 Use of synchrotron radiation for the analysis of amyloids self-assemblies. Top: Basic sketch of data acquisition and treatment at a SAXS beamline. The sample scattering pattern recorded by a 2D detector is radially averaged to and background is subtracted to give the following SAXS curve. The CTR peptide of the Hfq protein exhibits typical diffraction peaks corresponding to a cross-β structure sketch in overlay. Bottom: Data acquisition and treatment at a SRCD beamline. A rich **β**-strand containing protein is shown as an example (Concanavalin A, PDB ID 4PF5 [39])

6. All spectra are normalized to the mean residue weight ellipticity (Θ_{MRW}, deg.cm^2.dmole^{-1}).
7. For secondary structure determination with BestSel, normalized root-mean-square deviation (NRMSD) indicates the most accurate fit for each spectrum; values of <0.15 are considered significant.
   Figure 2b shows Hfq peptide analyzed by SRCD spectroscopy.

3.3 Characterization of Cross-β Structure

Small angle X-ray scattering (SAXS) is a simple and powerful technique that directly provides structural information at scales ranging typically from 1 nm to 1 μm [29–31]. The physical principle is the same as X-ray crystallography; it is based on elastic scattering of an incoming X-ray wave by the electrons of the sample. The resulting "signal" is a scattering pattern most frequently collected using a bi-dimensional detector positioned behind the sample and perpendicular to the incoming beam. Practically, SAXS experiment can bring information for either peptides in solutions or assembled materials made of peptides. In this section, we will focus on how to prepare and analyze amyloid peptides and particularly Hfq cross-β structure.

3.3.1 Sample Preparation

1. Different solutions of peptide are prepared (25–100 μL), ranging from 0.5% w/v to the highest concentration by 2.5% steps by dissolving the peptide into an appropriate buffer. The buffer must be the same for the previously described experiments (FTIR, SRCD…) in order to compare the results; buffer containing phosphorous should be avoided and low salt concentration is better.
2. When all the solutions are ready, directly fill quartz capillaries of 1.5 mm diameter and 0.01 mm thickness. Then, capillaries are sealed using wax or nail polish. The capillaries containing the peptides are left at room temperature until cross-β strands are obtained (this is usually determined by other methods).
3. If the cross-β strand formation kinetics are unknown, FTIR will be a better method than SAXS to determine the kinetics of assembly.

3.3.2 SAXS Setup and Recording

The SAXS experiment can bring information for both peptides in solutions or assembled materials made of peptides. The intensity collected is proportional to the number of objects within the sample, *N*. It directly results from their size, shape, and internal structure, given by the Fourier Transformation modulus of their electronic density, showed here for an isotropic sample:

$$I(q) = r_e^2 \sum_{p=1}^{N} \left\langle \left| \iiint_{V_p} \left(\rho_e(\vec{r}) - \rho_0 \right) e^{-i\vec{q}\cdot\vec{r}} d^3r \right|^2 \right\rangle_\Omega$$

where the triple integral is performed within the volume V_p of each particle and the average performed over all possible orientations Ω. r_e is the classical electron radius, $\rho_e(\vec{r})$ is the electronic density at position $\vec{r}$ within the particle, ρ_0 is the average electronic density of the surrounding buffer, $q = \frac{4\pi \sin\theta}{\lambda}$ is the momentum transfer, θ is the scattering angle, typically below a few degrees, and λ is the X-ray wavelength. The rotational average severely limits the structural information content of the scattering pattern for a non-assembled material obtained in a SAXS experiment. When isolated macromolecules assemble into regular structures, new correlations appear in the electronic density of the whole structure, giving rise to specific modulations in the scattering intensity. When motifs are repeated along a defined axis with a repetition distance d, the scattered photon waves undergo strong constructive interferences at q values given by $q = n\frac{\pi}{d}$, where n is an integer (Bragg law). The phenomenon is called diffraction, but is not different in nature from solution scattering. When identical units, e.g., a peptide or a protein, are regularly spaced, with periodicities typically on the order of or larger than 1 nm, diffraction is observed in the Small angle range (SAXD) [32].

1. Cross-β structure is investigated here. The sample-to-detector distance needs to be optimized. The typical signals resulting from β-sheet structure correspond to inter-distances of around 10 and 4.7 Å; the available q-range should thus be from 0.3 to 2 $Å^{-1}$.
2. Capillaries containing the samples and buffer (control) are mounted onto capillary holder system and positioned into the beam path.
3. Depending on the sample homogeneity, several positions within the sample capillaries can be chosen. At least 3 positions are enough: one at the top of the sample, one in the middle, and the last at the bottom.
4. One important step here is the detection of radiation damage. When exposed to X-rays, organic molecules tend to form free radicals, leading to aggregation of the sample onto the capillary wall. Radiation damage does not occur for all samples. This phenomenon is observable after data treatment by an increase of the curve intensity at small q. In order to avoid this problem, samples should be exposed to X-rays and then let to rest for a few seconds before a new exposition at the same spot.
5. As samples on the same capillary holder can vary greatly, the exposure time must be adjusted, in order to get enough signal and to avoid saturation on the detector (depending of the beamline detector).

6. Due to radiation damage, acquisition time should not be higher than 3 s with at least 5 s of pause. It is better to make 10 acquisitions of 3 s than 1 of 30 s.
7. Record the scattering pattern with short and identifiable sample and buffer names.

3.3.3 Data Treatment

1. When the entire scattering patterns are recorded as 2D images, radial integration is performed in order to obtain SAXS curves for each position inside all the samples and buffers. Such operation can be done using the Foxtrot software or equivalent.
2. For curves corresponding to the same position within the same sample, averaging is suggested in order to obtain better statistics. This must be done carefully in order to avoid averaging data with radiation damage.
3. As samples are analyzed within the buffer, the buffer signal needs to be subtracted. Data subtraction is provided by most of the SAXS data reduction software.
4. When the resulting final curve (averaging + subtraction of buffer) is obtained, diffraction peaks corresponding to cross-β sheets should be observed. The example of Hfq C-terminus peptide is shown on Fig. 3 (top), where the anisotropic reflections are indicative of a partially aligned fiber, perpendicular to the X-ray beam. The reflections at d = 4.58 Å and d = 7.97 Å correspond, respectively, to the inter-strand and inter-sheet spacing.

4 Concluding Remarks

The self-assembly of biological macromolecules constitutes a key process in all living organisms. We recently reported that sRNA cofactor Hfq belongs to the family of functional amyloids [9]. The various experimental approaches described in this chapter aim at further investigating the self-assembling properties of sRNA cofactors. This could indeed represent a versatile means to regulate sRNA-related processes in vivo. Furthermore, this would also enable the formation of mixed synthetic sRNA:protein synthetic self-assemblies, with future perspectives for nanotechnologies.

5 Notes

1. The main obstacle for the examination of peptides and proteins that polymerize, and in particular those that form amyloids, is sample preparation. Indeed, the buffer used for the synthesis/purification processes and its composition (counter-ions, mainly chloride ions or trifluoroacetic acid TFA^- ions used for

peptide chemical synthesis) may affect fibrillization [33]. For this reason, batch-to-batch variability of synthetic peptides is frequently reported, resulting in a poor reproducibility of experiments in term of kinetics of self-assembling. Proper storage conditions are also important. Typically, peptides are most stable when stored lyophilized at −20 °C. Nevertheless, even under this condition, hydration occurs. This may affect the presence of pre-formed aggregates in samples. Therefore, in order to ensure comparison between various experiments, we do not recommend storing peptides and proteins for a long time and rather suggest ordering or preparing a fresh batch and performing experiments in a short timeframe.

2. Use of ThT staining is a widely used method and relatively well accepted as an indicator of the presence of amyloid fibrils. Nevertheless, it has to be taken with precaution as ThT can also be an inhibitor of fibrillization [34].
3. A frequently used surface modification protocol for AFM involves fusing unilamelar lipid vesicles on glass or mica to form a supported lipid bilayer that can then be exposed to the protein fibrils. Indeed, interaction of amyloids with membrane is commonly reported and has been observed for Hfq [35, 36].
4. D_2O solutions may be used in samples to avoid the spectral overlaps between the Amide I band and strong absorption band of water at 1640 cm^{-1}. Note that Amide I wavenumbers are lowered in D_20 environment (5–10 cm^{-1}). In this case Amide I/II bands are referred to as Amide I′/II′.
5. The interpretation of an amyloid FTIR spectrum should usually begin with the examination of the primary sequence of the protein/peptide. Indeed, side chains such as asparagine and glutamine, which are very common in amyloid-forming proteins, have IR vibrations that overlap in the Amide I band.

Acknowledgments

This work was supported by Université Paris Diderot, CNRS, CEA, and Synchrotron SOLEIL. We gratefully acknowledge help to MV from the French Embassy for their program for scientific and university cooperation. We are indebted to F. Gobeaux (CEA Saclay, Gif-sur-Yvette, France) and A. Deniset-Besseau (LCP, Université Paris-Sud, France) for many fruitful discussions. We thank Kimberly Stanek (University of Virginia) for her careful and critical reading of our manuscript.

References

1. Taghbalout A, Yang Q, Arluison V (2014) The Escherichia coli RNA processing and degradation machinery is compartmentalized within an organized cellular network. Biochem J 458: 11–22
2. Arluison V, Taghbalout A (2015) Cellular localization of RNA degradation and processing components in Escherichia coli. Methods Mol Biol 1259:87–101
3. Lavelle C, Busi F, Arluison V (2015) Multiple approaches for the investigation of bacterial small regulatory RNAs self-assembly. Methods Mol Biol 1297:21–42
4. Cayrol B, Geinguenaud F, Lacoste J, Busi F, Le Derout J, Pietrement O, Le Cam E, Regnier P, Lavelle C, Arluison V (2009) Auto-assembly of E. coli DsrA small noncoding RNA: molecular characteristics and functional consequences. RNA Biol 6:434–445
5. Busi F, Cayrol B, Lavelle C, LeDerout J, Pietrement O, Le Cam E, Geinguenaud F, Lacoste J, Regnier P, Arluison V (2009) Auto-assembly as a new regulatory mechanism of noncoding RNA. Cell Cycle 8:952–954
6. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. Nat Rev Microbiol 9:578–589
7. Fischer S, Benz J, Spath B, Maier LK, Straub J, Granzow M, Raabe M, Urlaub H, Hoffmann J, Brutschy B, Allers T, Soppa J, Marchfelder A (2010) The archaeal Lsm protein binds to small RNAs. J Biol Chem 285:34429–34438
8. Smirnov A, Wang C, Drewry LL, Vogel J (2017) Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. EMBO J 36:1029–1045
9. Fortas E, Piccirilli F, Malabirade A, Militello V, Trepout S, Marco S, Taghbalout A, Arluison V (2015) New insight into the structure and function of Hfq C-terminus. Biosci Rep 35:e00190
10. Maury CP (2009) The emerging concept of functional amyloid. J Intern Med 265:329–334
11. Ostrowski A, Mehert A, Prescott A, Kiley TB, Stanley-Wall NR (2011) YuaB functions synergistically with the exopolysaccharide and TasA amyloid fibers to allow biofilm formation by Bacillus subtilis. J Bacteriol 193:4821–4831
12. Zhou Y, Blanco LP, Smith DR, Chapman MR (2012) Bacterial amyloids. Methods Mol Biol 849:303–320
13. Aguilera P, Marcoleta A, Lobos-Ruiz P, Arranz R, Valpuesta JM, Monasterio O, Lagos R (2016) Identification of key amino acid residues modulating intracellular and in vitro microcin E492 amyloid formation. Front Microbiol 7:35
14. Refregiers M, Wien F, Ta HP, Premvardhan L, Bac S, Jamme F, Rouam V, Lagarde B, Polack F, Giorgetta JL, Ricaud JP, Bordessoule M, Giuliani A (2012) DISCO synchrotron-radiation circular-dichroism endstation at SOLEIL. J Synchrotron Radiat 19:831–835
15. Giuliani A, Jamme F, Rouam V, Wien F, Giorgetta JL, Lagarde B, Chubar O, Bac S, Yao I, Rey S, Herbeaux C, Marlats JL, Zerbib D, Polack F, Refregiers M (2009) DISCO: a low-energy multipurpose beamline at synchrotron SOLEIL. J Synchrotron Radiat 16:835–841
16. Wien F, Wallace BA (2005) Calcium fluoride micro cells for synchrotron radiation circular dichroism spectroscopy. Appl Spectrosc 59:1109–1113
17. Lees JG, Smith BR, Wien F, Miles AJ, Wallace BA (2004) CDtool-an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. Anal Biochem 332:285–289
18. Micsonai A, Wien F, Kernya L, Lee YH, Goto Y, Refregiers M, Kardos J (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc Natl Acad Sci U S A 112:E3095–E3103
19. Reinke AA, Gestwicki JE (2011) Insight into amyloid structure using chemical probes. Chem Biol Drug Des 77:399–411
20. Rhiannon GC, Christopher TG, Nicolas HV (2012) Characterization of fiber-forming peptides and proteins by means of atomic force microscopy. Curr Protein Pept Sci 13:232–257
21. Dufrene YF, Ando T, Garcia R, Alsteens D, Martinez-Martin D, Engel A, Gerber C, Muller DJ (2017) Imaging modes of atomic force microscopy for application in molecular and cell biology. Nat Nanotechnol 12:295–307
22. Liberelle B, Banquy X, Giasson S (2008) Stability of silanols and grafted alkylsilane monolayers on plasma-activated mica surfaces. Langmuir 24:3280–3288
23. Thompson RF, Walker M, Siebert CA, Muench SP, Ranson NA (2016) An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. Methods 100:3–15
24. Militello V, Casarino C, Emanuele A, Giostra A, Pullara F, Leone M (2004) Aggregation kinetics of bovine serum albumin studied by FTIR spectroscopy and light scattering. Biophys Chem 107:175–187

25. Byler DM, Susi H (1986) Examination of the secondary structure of proteins by deconvolved FTIR spectra. Biopolymers 25:469–487
26. Zandomeneghi G, Krebs MR, McCammon MG, Fandrich M (2004) FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. Protein Sci 13:3314–3321
27. Wallace BA (2009) Protein characterisation by synchrotron radiation circular dichroism spectroscopy. Q Rev Biophys 42:317–370
28. Kardos J, Micsonai A, Pal-Gabor H, Petrik E, Graf L, Kovacs J, Lee YH, Naiki H, Goto Y (2011) Reversible heat-induced dissociation of beta2-microglobulin amyloid fibrils. Biochemistry 50:3211–3220
29. Guinier A (1955) Small-angle scattering of x-rays, Structure of matter series. Wiley, New York
30. Feigin LAS, Svergun DI, Taylor GW (1987) Structure analysis by small-angle X-ray and neutron scattering. Plenum Press, NY, USA
31. Feigin O, Kratky O (1982) Small angle x-ray scattering. Academic Press, New York/London
32. Guinier A (1994) X-ray diffraction: in crystals, imperfect crystals, and amorphous bodies. Courier Corporation, North Chelmsford
33. Kaneko I, Tutumi S (1997) Matters arising: conformations of β-amyloid in solution. J Neurochem 68:438–439
34. Stains CI, Mondal K, Ghosh I (2007) Molecules that target beta-amyloid. ChemMedChem 2:1674–1692
35. Herrera AI, Tomich JM, Prakash O (2016) Membrane interacting peptides: a review. Curr Protein Pept Sci 17:827–841
36. Malabirade A, Morgado-Brajones J, Marquez I, Seguin J, Trepout S, Wien F, Marco S, Velez M, Arluison V (2017) Membrane association of the bacterial riboregulator Hfq and functional perspectives. Sci Rep 7(1):10724
37. Vetri V, D'Amico M, Fodera V, Leone M, Ponzoni A, Sberveglieri G, Militello V (2011) Bovine Serum Albumin protofibril-like aggregates formation: solo but not simple mechanism. Arch Biochem Biophys 508:13–24
38. Vetri V, Militello V (2005) Thermal induced conformational changes involved in the aggregation pathways of beta-lactoglobulin. Biophys Chem 113:83–91
39. Francois-Heude M, Mendez-Ardoy A, Cendret V, Lafite P, Daniellou R, Ortiz Mellet C, Garcia Fernandez JM, Moreau V, Djedaini-Pilard F (2015) Synthesis of high-mannose oligosaccharide analogues through click chemistry: true functional mimics of their natural counterparts against lectins? Chemistry 21:1978–1991

Chapter 19

Sequence-Specific Affinity Chromatography of Bacterial Small Regulatory RNA-Binding Proteins from Bacterial Cells

Jonathan Gans, Jonathan Osborne, Juliet Cheng, Louise Djapgne, and Amanda G. Oglesby-Sherrouse

Abstract

Bacterial small RNA molecules (sRNAs) are increasingly recognized as central regulators of bacterial stress responses and pathogenesis. In many cases, RNA-binding proteins are critical for the stability and function of sRNAs. Previous studies have adopted strategies to genetically tag an sRNA of interest, allowing isolation of RNA–protein complexes from cells. Here we present a sequence-specific affinity purification protocol that requires no prior genetic manipulation of bacterial cells, allowing isolation of RNA-binding proteins bound to native RNA molecules.

Key words Bacterial small RNAs, Hfq, RNA-binding proteins, Affinity chromatography, In vivo crosslinking, Western blot

1 Introduction

Over the past 15 years, bacterial small RNA (sRNA) molecules have gained substantial recognition for their role in shaping bacterial physiology and pathogenesis. In spite of this appreciation, we still know very little about the mechanisms that drive sRNA regulation in most bacterial specie0073. A majority of mechanistic studies have been centered on sRNAs produced by *Escherichia coli* [1–6]. This work has demonstrated that the host factor Q (Hfq) protein is required for the stability and function of many *E. coli* sRNAs [3, 7–12]. Subsequent work indicates that the requirement for Hfq extends to sRNAs produced by other bacterial species [13–19]. However, many bacterial sRNAs have been shown to not require Hfq for their stability or regulation [15, 20], and the Hfq proteins of most Gram-positive bacteria do not appear to contribute to sRNA regulation [21–23]. This has spurred the development

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_19,

of methods aimed at identifying novel RNA-binding proteins that may contribute to sRNA biology.

Previous approaches relied on genetic manipulation of bacterial cells to produce sRNAs that are tagged with aptamers [24]. However, this method requires the bacterial species of interest to be amenable to genetic manipulation and relies on the assumption that the aptamer will not affect the expression or structure of the tagged sRNA. Here we present a method for the enrichment of sRNA-binding proteins from bacterial cells, previously developed for use in the opportunistic bacterial pathogen *Pseudomonas aeruginosa* [25]. The method was adapted from earlier protocols developed for use with eukaryotic organisms [26, 27]. It entails sequence-specific affinity chromatography (SSAC) of UV-crosslinked RNA–protein complexes from bacterial cells using a tagged complementary DNA (cDNA) oligonucleotide that is specific to the sRNA of interest. As such, this method requires no genetic manipulation of the organism, allowing adaptation to bacterial species that are not easily amenable to genetic manipulation. We previously combined this system with a discovery-based liquid chromatography tandem mass spectrometry (LC-MS/MS) proteomics approach to identify proteins that interact in vivo with the *P. aeruginosa* PrrF and PrrH sRNAs [25]. Here, we provide a detailed protocol of the in vivo crosslinking and SSAC methodology, and we discuss how this method may be used in conjunction with different downstream applications to identify or detect proteins bound to sRNAs in vivo.

2 Materials

Prepare all solutions in nuclease-free water with molecular biology-grade reagents. Our lab generally purchases nuclease-free water and buffers, but we have also used diethyl pyrocarbonate (DEPC) to render deionized water RNase-free (0.1% DEPC treatment for 2 h, followed by 15 min autoclaving to inactivate residual traces of DEPC). Do not use DEPC to treat solutions containing Tris. Molecular biology-grade reagents should be reserved for RNA work only, and powdered reagents should be measured without the use of a spatula. Reagents should be prepared and stored at room temperature, unless indicated otherwise. All work should be performed with filtered pipette tips that are certified nuclease-free.

2.1 Preparation of Crosslinked Bacterial RNA–Protein Complexes

1. Bacterial cultures, grown under conditions that allow for maximal expression of the sRNA of interest.
2. Qiagen RNAlater®.
3. Large Petri plates (150 mm × 15 mm).

4. UV Crosslinker (115 V).
5. Lysozyme solution (1 mg/mL) in 10 mM Tris–HCl- (pH 8.0), 1 mM EDTA.
6. Beta-mercaptoethanol.
7. Qiagen RNeasy® Buffers RLT, RW1, and RPE.
8. 100% ethanol, molecular biology grade.
9. Qiagen RNeasy® Mini Columns.
10. Qiagen on-column RNase-free DNase.
11. RNase-free water.
12. New England Biolabs (NEB) DNase I (RNase-free).
13. 3 M sodium acetate.
14. Nanodrop or similar instrumentation for quantifying nucleic acid concentrations.

2.2 Sequence-Specific Affinity Chromatography

1. Invitrogen streptavidin-coated M-270 Dynabeads®.
2. Magnetic stand.
3. 2× Binding and Washing Buffer (10 mM Tris–HCl pH 7.5, 1 mM EDTA, 2 M NaCl).
4. Solution A (0.1 M NaOH, 0.05 M NaCl).
5. Solution B (0.1 M NaCl).
6. Nuclease-free water.
7. 1 μg 5′ biotinylated cDNA "bait" that is complementary to the RNA of interest, prepared in RNase-free water at a concentration of 50 ng/μL (*see* **Note 1**).

3 Methods

All procedures are performed at room temperature, unless otherwise stated.

3.1 Preparation of In Vivo Crosslinked RNA–Protein Complexes

1. Grow *P. aeruginosa* strains in appropriate media to ensure maximal expression of sRNA of interest (*see* **Note 2**).
2. Add 1.5 mL Qiagen RNAlater® to 5 mL culture to stabilize RNAs during crosslinking.
3. Pour mixture into a large Petri dish and make sure culture is spread evenly across the bottom of the dish. Irradiate using a UV crosslinker for 2–3 min (*see* **Note 3**).
4. Transfer 1 mL of irradiated culture to a microcentrifuge tube and harvest by centrifugation at ≥13,000 × *g* for 1 min.

   ***Steps 5–12** are taken from the Qiagen RNeasy® Mini Kit manufacturer's protocol with some minor modifications:*

5. Resuspend harvested cells in 100 μL of lysozyme solution. Incubate for 5 min at room temperature.

6. Prepare Qiagen RNeasy® RLT Buffer for use by adding 10 μL of beta-mercaptoethanol (BME) per 1 mL of RLT. Add 350 μL RLT-BME solution to lysozyme-treated cells. Vortex and incubate at room temperature for 15 min.
7. Add 250 μL 100% ethanol to lysed cells and invert tube 3–4 times. Transfer mixture to Qiagen RNeasy Mini Column. Centrifuge for 1 min at ≥13,000 × *g*.
8. Discard flow-through and add 350 μL of Qiagen RW1 Buffer to RNeasy Mini Column. Centrifuge for 1 min at 13,000 × *g*.
9. Combine 10 μL of Qiagen on-column RNase-free DNase solution with 30 μL RDD Buffer. Apply to RNeasy Mini Column. Incubate for 1–2 h at room temperature.
10. Add 350 μL RW1 Buffer to RNeasy Mini Column. Centrifuge for 1 min at ≥13,000 × *g*.
11. Discard flow-through and add 500 μL of RPE Buffer to RNeasy Mini Column. Centrifuge for 1 min.
12. Discard flow-through and centrifuge RNeasy Mini Column for 5 min at ≥13,000 × *g* to dry column.
13. Add 40 μL RNase-free water to RNeasy Mini Column. Incubate for 5 min at room temperature. Elute by centrifugation for 1 min at 13,000 × *g*.
14. Add 5 μL NEB DNase and 5 μL 10× reaction buffer to supernatant. Incubate at room temperature for 2–3 h.
15. Add 150 μL 100% ethanol and 5 μL 3 M sodium acetate and vortex mixture. Incubate precipitation overnight at −20 °C.
16. Centrifuge sample at 16,000 × *g* for 20 min at 4 °C. Remove supernatant and add 1 mL ice-cold 100% ethanol.
17. Centrifuge sample at 16,000 × *g* for 10 min at 4 °C. Remove supernatant and let RNA pellet air-dry at room temperature for 5–10 min, with tube caps open slightly to allow evaporation of residual ethanol.
18. Add 15 μL of water to RNA pellet and incubate at 65 °C for 10–15 min to promote resuspension.
19. Measure the RNA using a nanodrop and adjust the concentration to 1 μg/μL (*see* **Note 4**). Keep RNA on ice until **Step 6** of Subheading 3.2.

3.2 Enrichment of RNA-Binding Proteins

1. Transfer 1 mg of M-270 streptavidin-coated Dynabeads® (100 μL) to an RNase-free eppendorf tube.
2. Place eppendorf tube on magnetic stand and allow the Dynabeads® to separate from the storage buffer. Remove storage buffer completely, and add 250 μL of 2× Binding and Washing Buffer. Resuspend Dynabeads® in buffer by inverting

tube 3–4 times. Repeat this process two more times with the 2× Binding and Washing Buffer.

3. Wash Dynabeads® two times as in **Step 2** with 250 µL of Solution A.
4. Wash Dynabeads® two times as in **Step 2** with 250 µL of Solution B.
5. Resuspend Dynabeads® in 100 µL 2× Binding and Washing Buffer. Keep Dynabeads® on ice until **Step 7**.
6. Combine 1 µg of 5′ biotinylated cDNA-bait with 10 µg (10 µL) of the SSAC supernatant from **Step 19** of Subheading 3.1. Bring total volume up to 100 µL using nuclease-free water, and incubate the SSAC supernatant and cDNA bait mixture at 70 °C for 30 min. Transfer mixture to 37 °C and incubate for 15 min with gentle rotation at 150 rpm.
7. Combine the SSAC supernatant and cDNA bait mixture with 1 mg of M-270 Dynabeads®. Incubate at 37 °C for 45 min with gentle shaking. Separate beads from solution using a magnetic stand and discard the supernatant.
8. Resuspend the beads in 100 µL of nuclease-free water and invert tube 3–4 times. Incubate beads at 65 °C for 10 min to disrupt streptavidin–biotin interaction and remove RNA-bait mixture from the beads. Separate beads from solution using a magnetic stand and collect the supernatant.
9. Enriched RNA–protein complexes in the SSAC supernatant are now ready for analysis by any number of proteomic approaches for a discovery-based identification of RNA-binding proteins (*see* **Note 5**). We previously performed LC-MS/MS analysis of wild-type and Δ*prrF1,2 P. aeruginosa* SSAC supernatants ("SSAC-MS/MS"), allowing the identification of several potential PrrF1- and PrrH-binding proteins (Table 1). Alternatively, western blot can be used to determine if a specific protein is present in the enriched RNA–protein complexes, as is shown for the PrrF1 sRNA and *P. aeruginosa* Hfq protein in Fig. 1.

4 Notes

1. The cDNA bait should be long enough to provide specificity in a complex RNA mixture isolated from bacterial cells, as would be warranted for a northern blot. The PrrF1 cDNA bait that was previously used for such studies was 27 nucleotides long, and the PrrH probe was 60 nucleotides long. The PrrF1 and PrrH cDNA baits were each shown by northern blot to be specific for the PrrF1 and PrrH sRNAs, respectively (Fig. 2a, b) [25].

Table 1
Summary of SSAC-MS/MS results

	Positive samples[a]			
	PrrF1 bait		PrrH bait	
Protein	WT–Fe $n = 15$	Δ*prrF1,2* $n = 8$	WT–Fe $n = 8$	Δ*prrF1,2* $n = 9$
Hfq (PA4944)	11**	1	3	1
PvdL (PA2424)	5*	0	3	1
LysR-type transcriptional regulator (PA1128)	3	0	2	0
Shikimate biosynthesis (PA1750)	3	0	2	1
Putative oxidoreductase (PA3106)	3	0	2	1
HemB (PA5243)	3	0	0	0
HscA (PA3810)	3	1	1	2

[a]Number of samples with at least one peptide corresponding to the indicated protein
Asterisks indicate a significant difference in the frequency of Hfq ($P < 0.005$) and PvdL ($P < 0.05$) positive samples when comparing PrrF1-enriched PAO1 and Δ*prrF1,2* samples, as determined by a two-tailed Student's *t*-test. SSAC-MS/MS, sequence-specific affinity chromatography and tandem mass spectrometry

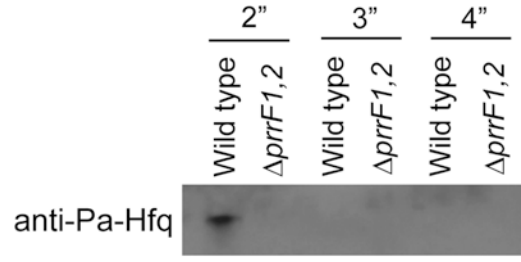


Fig. 1 Visualization of Hfq from SSAC supernatants by western blot. SSAC supernatants from wild-type and Δ*prrF1,2* mutant cultures, grown in M9 minimal media, supplemented with 50 nM $FeCl_3$ added to support low-iron growth for 8 h, were separated by SDS-polyacrylamide gel electrophoresis (PAGE), transferred to a PVDF membrane, and blotted with *P. aeruginosa* Hfq antibody that was generated as previously described [25]

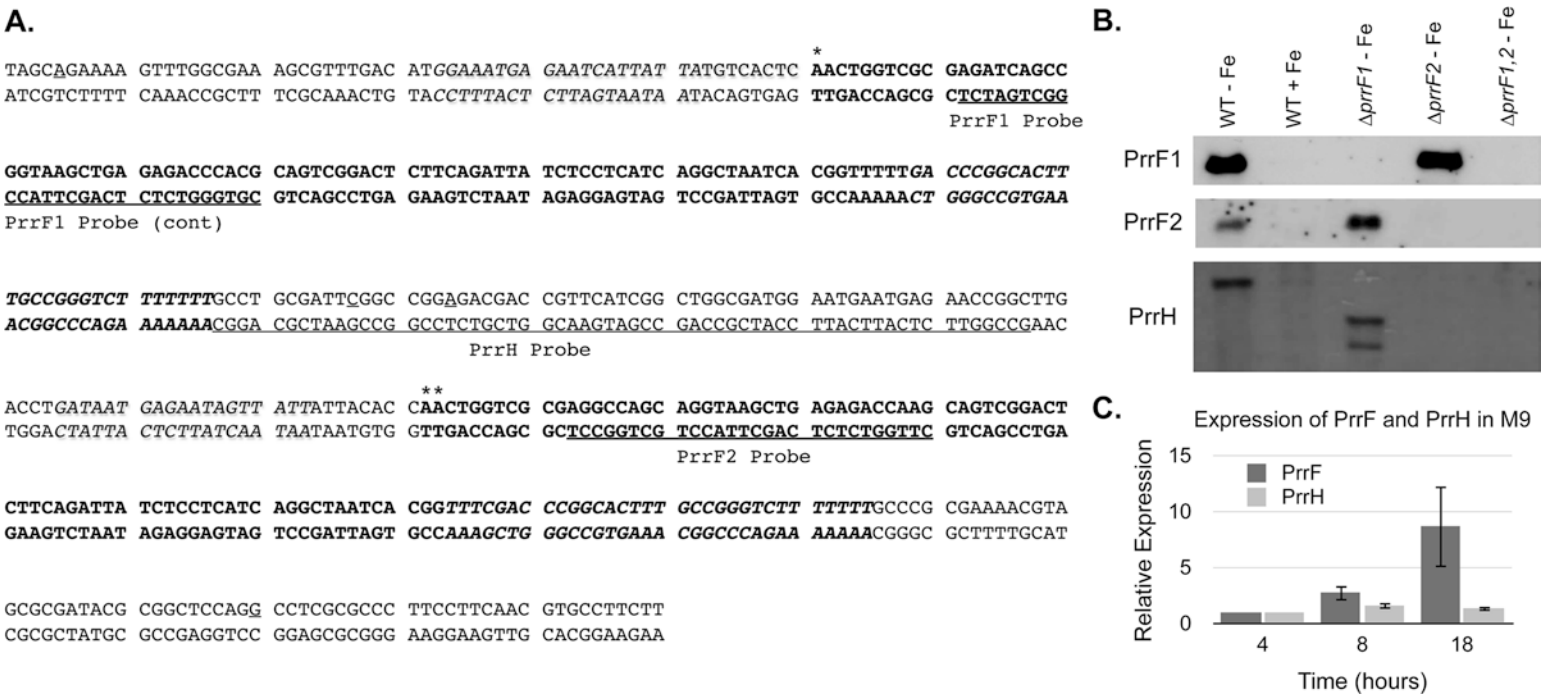


Fig. 2 Development of PrrF and PrrH probes for SSAC-MS/MS. (**a**) Sequence of the *prrF* locus with the PrrF1, PrrF2, and PrrH northern blot probe locations underlined. Asterisks indicate the transcriptional start site of PrrF1 and PrrH (*) and PrrF2 (**). The PrrF1 and PrrF2 coding sequences are in bold, and the PrrF1 and PrrF2 Rho-independent terminators are italicized. (**b**) Northern blots of PAO1 and the indicated *prrF* mutants grown in M9 minimal media for 18 h, with or without supplementation of 100 μM $FeCl_3$ as indicated, using the PrrF1, PrrF2, and PrrH probes shown in (**a**). (**c**) Expression of the PrrF and PrrH sRNAs was determined by qPCR after 4, 8, and 18 h of growth in M9 minimal media with no iron supplementation. Relative expression of each RNA was determined using a standard curve, and values for each time point were normalized to the 4 h time point. Error bars indicate the standard deviation of three independent experiments. (This figure was reproduced from reference [25] under the Creative Commons Attribution license, copyright © 2014 Osborne, et al.)

2. The culture conditions used for PrrF1 and PrrH analysis (growth in M9 minimal medium) were chosen because they allowed for robust expression of the PrrF and PrrH sRNAs (*see* Fig. 2c). We also opted to purchase M9 minimal medium to ensure consistency across experiments when conducting discovery-based analysis of PrrF-binding proteins.
3. When first developing this protocol for discovery of novel PrrF- and PrrH-binding proteins, we attempted various crosslinking times ranging from 30 s to 6 min and evaluated the amount of protein pulled down by gel electrophoresis. Three minutes was the longest irradiation time that still allowed efficient separation of proteins by electrophoresis and was sufficient for identifying Hfq in wild-type SSAC supernatants but not in the Δ*prrF1,2* mutant SSAC supernatants (Table 1) [25]. However, we found that crosslinking times greater than 2 min precluded visualization of the Hfq-sRNA complex by western blot (Fig. 1), indicating this method is more sensitive to aggregation of UV-crosslinked macromolecules. The time for crosslinking may also need to be optimized for different organisms, sRNAs, or crosslinking apparatus. We recommend preparing SSAC-enriched RNA samples from cultures exposed to a range (1–5 min) of crosslinking times, and analyzing lysates from these cultures by PAGE and Coommassie staining to determine which allows for the best separation of crosslinked proteins.
4. We routinely use the Qiagen RNeasy® RNA Isolation Mini Kit, which is sufficient to yield 15 μL of RNA at a concentration of >100 ng/μL with a 260/280 ratio of 2.0 ± 0.1. This concentration was sufficient for detecting Hfq bound to PrrF1 and PrrH by LC-MS/MS (Table 1) [25] and to PrrF1 by Western blot (Fig. 1). Alternative methods for RNA isolation should be sufficient, so long as they yield a high-quality RNA sample at a sufficient concentration. The concentration of RNA that is necessary for detection, however, will depend on the expression level of the specific sRNA being targeted, the amount of protein that is bound to that sRNA, and the sensitivity of the method used to detect the bound protein.
5. Discovery of novel sRNA-binding proteins is dependent upon analysis of an isogenic mutant lacking expression of the sRNA of interest. As shown in Table 1, only two proteins identified using the PrrF1 cDNA bait showed statistically significant differences in pull-down rates between the wild-type and Δ*prrF1,2* RNA samples. Also note that the rate of Hfq pull-down by the PrrH bait was much lower than that for the PrrF1 bait. It remains unclear if this is due to decreased interaction of PrrH with Hfq, or the smaller amount of PrrH as compared to PrrF1 in *P. aeruginosa* cells [28].

Acknowledgments

This work was supported by NIH grant AI123320 and funding from the University of Maryland School of Pharmacy (to A.G.O.-S.).

References

1. Panja S, Schu DJ, Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. Nucleic Acids Res 41:7536–7546
2. Panja S, Santiago-Frangos A, Schu DJ, Gottesman S, Woodson SA (2015) Acidic residues in the Hfq chaperone increase the selectivity of sRNA binding and annealing. J Mol Biol 427:3491–3500
3. Dimastrogiovanni D, Frohlich KS, Bandyra KJ, Bruce HA, Hohensee S, Vogel J, Luisi BF (2014) Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. elife 3. https://doi.org/10.7554/eLife.05375
4. Afonyushkin T, Vecerek B, Moll I, Blasi U, Kaberdin VR (2005) Both RNase E and RNase III control the stability of *sodB* mRNA upon translational inhibition by the small regulatory RNA RyhB. Nucleic Acids Res 33: 1678–1689
5. De Lay N, Gottesman S (2011) Role of polynucleotide phosphorylase in sRNA function in Escherichia coli. RNA 17:1172–1189
6. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Mol Cell 9:11–22
7. Schu DJ, Zhang A, Gottesman S, Storz G (2015) Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. EMBO J 34:2557–2573
8. Soper TJ, Doxzen K, Woodson SA (2011) Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. RNA 17:1544–1550
9. Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S (2003) Global analysis of small RNA and mRNA targets of Hfq. Mol Microbiol 50:1111–1124
10. Masse E, Escorcia FE, Gottesman S (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. Genes Dev 17:2374–2383
11. Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, Feig AL (2004) Escherichia coli Hfq has distinct interaction surfaces for DsrA, *rpoS* and poly(A) RNAs. Nat Struct Mol Biol 11:1206–1214
12. Lease RA, Woodson SA (2004) Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. J Mol Biol 344:1211–1223
13. Sorger-Domenigg T, Sonnleitner E, Kaberdin VR, Blasi U (2007) Distinct and overlapping binding sites of *Pseudomonas aeruginosa* Hfq and RsmA proteins on the non-coding RNA RsmY. Biochem Biophys Res Commun 352:769–773
14. Hunter GA, Keener JP (2014) Mechanisms underlying the additive and redundant Qrr phenotypes in *Vibrio harveyi* and *Vibrio cholerae*. J Theor Biol 340:38–49
15. Deng Z, Meng X, Su S, Liu Z, Ji X, Zhang Y, Zhao X, Wang X, Yang R, Han Y (2012) Two sRNA RyhB homologs from *Yersinia pestis* biovar *microtus* expressed *in vivo* have differential Hfq-dependent stability. Res Microbiol 163:413–418
16. Nielsen JS, Lei LK, Ebersbach T, Olsen AS, Klitgaard JK, Valentin-Hansen P, Kallipolitis BH (2010) Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in *Listeria monocytogenes*. Nucleic Acids Res 38:907–919
17. Troxell B, Fink RC, Porwollik S, McClelland M, Hassan HM (2011) The Fur regulon in anaerobically grown *Salmonella enterica* sv. Typhimurium: identification of new Fur targets. BMC Microbiol 11:236
18. Metruccio MM, Fantappie L, Serruto D, Muzzi A, Roncarati D, Donati C, Scarlato V, Delany I (2009) The Hfq-dependent small noncoding RNA NrrF directly mediates Fur-dependent positive regulation of succinate dehydrogenase in *Neisseria meningitidis*. J Bacteriol 191:1330–1342
19. Davis BM, Quinones M, Pratt J, Ding Y, Waldor MK (2005) Characterization of the small untranslated RNA RyhB and its regulon in *Vibrio cholerae*. J Bacteriol 187:4005–4014
20. Gaballa A, Antelmann H, Aguilar C, Khakh SK, Song KB, Smaldone GT, Helmann JD (2008) The *Bacillus subtilis* iron-sparing

response is mediated by a Fur-regulated small RNA and three small, basic proteins. Proc Natl Acad Sci U S A 105:11927–11932

21. Zheng A, Panja S, Woodson SA (2016) Arginine patch predicts the RNA annealing activity of Hfq from gram-negative and gram-positive bacteria. J Mol Biol 428:2259–2264
22. Silvaggi JM, Perkins JB, Losick R (2005) Small untranslated RNA antitoxin in *Bacillus subtilis*. J Bacteriol 187:6641–6650
23. Boisset S, Geissmann T, Huntzinger E, Fechter P, Bendridi N, Possedko M, Chevalier C, Helfer AC, Benito Y, Jacquier A et al (2007) *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. Genes Dev 21:1353–1366
24. Said N, Rieder R, Hurwitz R, Deckert J, Urlaub H, Vogel J (2009) *In vivo* expression and purification of aptamer-tagged small RNA regulators. Nucleic Acids Res 37:e133
25. Osborne J, Djapgne L, Tran BQ, Goo YA, Oglesby-Sherrouse AG (2014) A method for *in vivo* identification of bacterial small RNA-binding proteins. Microbiology 3:950–960
26. Blencowe BJ, Sproat BS, Ryder U, Barabino S, Lamond AI (1989) Antisense probing of the human U4/U6 snRNP with biotinylated 2′-OMe RNA oligonucleotides. Cell 59:531–539
27. Lingner J, Cech TR (1996) Purification of telomerase from *Euplotes aediculatus*: requirement of a primer 3′ overhang. Proc Natl Acad Sci U S A 93:10712–10717
28. Oglesby-Sherrouse AG, Vasil ML (2010) Characterization of a heme-regulated non-coding RNA encoded by the *prrF* locus of *Pseudomonas aeruginosa*. PLoS One 5:e9930

Chapter 20

Identification of Small RNA–Protein Partners in Plant Symbiotic Bacteria

Marta Robledo, Ana M. Matia-González, Natalia I. García-Tomsig, and José I. Jiménez-Zurdo

Abstract

The identification of the protein partners of bacterial small noncoding RNAs (sRNAs) is essential to understand the mechanistic principles and functions of riboregulation in prokaryotic cells. Here, we describe an optimized affinity chromatography protocol that enables purification of in vivo formed sRNA–protein complexes in *Sinorhizobium meliloti*, a genetically tractable nitrogen-fixing plant symbiotic bacterium. The procedure requires the tagging of the desired sRNA with the MS2 aptamer, which is affinity-captured by the MS2-MBP protein conjugated to an amylose resin. As proof of principle, we show recovery of the RNA chaperone Hfq associated to the strictly Hfq-dependent AbcR2 *trans*-sRNA. This method can be applied for the investigation of sRNA–protein interactions on a broad range of genetically tractable α-proteobacteria.

Key words Rhizobia, *trans*-sRNA, MS2, Maltose-binding protein

1 Introduction

Posttranscriptional regulation of gene expression by small noncoding RNAs (sRNAs) is ubiquitous in bacteria. The vast majority of sRNAs rely on antisense interactions with one or multiple mRNAs to control translation and/or stability of the targeted transcripts. However, sequence complementarity between sRNAs and *trans*-encoded target mRNAs is typically limited to short and discontinuous nucleotide stretches, and therefore these interactions require the assistance of proteins. Previous work on classical model enterobacteria identified Hfq and, more recently, ProQ as RNA chaperones acting as global sRNA stabilizers and matchmakers in riboregulation [1, 2]. However, these proteins are not widely distributed across bacterial kingdom. Even in bacteria encoding a functional Hfq homolog, this protein has a limited impact or is fully dispensable in riboregulation, anticipating that other yet undiscovered proteins may fulfill similar chaperone roles in sRNA

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_20,

regulatory networks. It is also well known that the enterobacterial Hfq establishes a higher-order protein complex with the RNA degradosome consisting of the major single-strand RNase E endoribonuclease, the 3′–5′ exoribonuclease polynucleotide phosphorylase (PNPase), RNA helicase B, and enolase [3]. The degradosome promotes target mRNA decay upon base-pairing interaction with its cognate partner Hfq-dependent sRNAs. On the other hand, other classes of bacterial sRNAs act by target mimicry rather than through base-pairing interactions to counteract the activity of specific proteins functionally related to the flow of genetic information. Well-characterized examples of this class of riboregulators are the 6S and CsrB sRNA families, which antagonize the activity of the σ^{70} RNA polymerase holoenzyme or the translational repressors of the CsrA/RmsA family, respectively [4, 5].

Understanding the function and activity mechanisms of bacterial sRNAs therefore requires the identification of their protein partners. In this regard, research on phylogenetically distant bacteria exhibiting complex lifestyles is expected to add new paradigms to what is known about protein-assisted riboregulation in classical model enterobacteria. We have optimized an affinity chromatography protocol to capture proteins interacting with sRNAs identified in the nitrogen-fixing symbiotic α-proteobacterium *Sinorhizobium meliloti*. The procedure relies on 5′-tagging of bait sRNAs with the RNA aptamer recognized by the MS2 coat protein, which is fused to a maltose-binding protein (MBP) to allow its immobilization on an affinity matrix. As a proof of principle, we show reliable recovery of Hfq with its known *S. meliloti* AbcR2 sRNA partner.

2 Materials

Standard equipment in molecular biology (e.g., incubators, gel electrophoresis devices, refrigerated centrifuges) is required for the following protocols. To avoid RNA degradation, special attention should be kept on glassware and equipment cleanness. Working solutions were prepared in ultrafiltered sterile water. Commercial RNase-free water and plasticware were used for in vitro assays. Below, the specific material needed to carry out the methods described in Subheading 3 is listed.

2.1 Culture and Harvest of Bacteria

1. Rhizobial strains: wild-type and derivatives expressing a FLAG-tagged Hfq (e.g., Sm*hfq*^{FLAG} in Sm2B3001) [6] or carrying a deletion of the corresponding sRNA *locus* (e.g., Sm2B3001Δ*abcR2*) [7].
2. Media: LB medium was routinely used to grow *E. coli* at 37 °C and complex tryptone yeast (TY) [8] or defined minimal medium (MM) [9] for growing rhizobia at 30 °C.

3. Antibiotics: streptomycin (600 mg/mL), tetracycline (10 mg/mL), or kanamycin (50 mg/mL for *E. coli* and 180 mg/mL for *Sinorhizobium* strains).
4. Cultures (usually 200 mL cultures in 1 L flasks) were incubated at 30 °C and 180 rpm.
5. Although most of the bacterial sRNAs are upregulated in early stationary phase [10], optimal expression conditions for the sRNA of interest must be assessed prior the study.
6. Cells are harvested by centrifugation (10 min at 5500 × *g* and 4 °C). Cell pellets are washed once with 0.1% sarcosyl in Tris-EDTA pH: 8.0 (TE Buffer) to facilitate subsequent cell disruption and once again with the working buffer (e.g., buffer A for affinity purification) prior to storage at −80 °C.

2.2 sRNA Aptamer Tagging

1. Vector pSRK-C, which is an engineered pSRKKm [11], lacking the LacIQ operator is used as backbone for aptamer-tagged sRNA constructions (confers Km resistance).
2. DNA oligonucleotides (100 pmol/mL) for direct annealing of aptamer and terminator sequences and for PCR amplification of the sRNA (see details for primer design in Subheading 3.1).
3. Conventional PCR: Phusion High-Fidelity DNA Polymerase (Finnzymes), 5× Phusion HF Buffer, 10 mM dNTPs.
4. DNA electrophoresis: 6× DNA loading dye, 10× TAE buffer (0.4 M Tris, 17.4 M acetic acid, 0.02 M EDTA pH: 8.2), and agarose.
5. PCR and plasmid DNA purification kits.
6. Cloning: Restriction enzymes, T4 DNA ligase (5000 U/mL), and 10× T4 DNA ligation buffer.
7. Rubidium chloride competent *E. coli* DH5α (for cloning) and S17.1 cells (for conjugation by biparental mating).

2.3 Northern Hybridization

Materials listed below are required to purify RNA, including sRNAs, without the use of columns, although commercial kits are available (*see* **Note 1**).

1. RNA isolation and DNA digestion: lysis solution (1.4% SDS, 4 mM EDTA, 50 μg proteinase K), 5 M sodium chloride (NaCl), ethanol (EtOH), RNase-free DNase I.
2. RNA extraction and precipitation: phenol (pH: 4.5):chloroform:isoamyl alcohol solution (25:24:1, v/v), EtOH and sodium acetate 3 M.
3. Acrylamide gel: 10× TBE (0.89 M Tris, 0.89 M boric acid, 0.02 M EDTA pH: 8.0); 40% acrylamide/bisacrylamide solution, 7 M urea, 10% ammonium persulfate (APS, prepare fresh or store aliquots at −20 °C), and tetramethylethylenediamine (TEMED).

4. Electrophoresis: 2× RNA loading buffer (97.5% formamide, 10 mM EDTA pH 7.5, 0.3% xylene cyanol, 0.3% bromophenol blue); RNA molecular weight marker.
5. Membrane blotting: 3 mm Whatman paper, positively charged nylon membrane, semidry electroblot transfer apparatus.
6. Hybridization: Buffer (0.5 M sodium phosphate buffer pH 7.2, 7% SDS, 10 mM EDTA), hybridization tubes, and oven.
7. Probe labeling: T4 phosphonucleotidekinase (PNK) provided with 10× reaction buffer, 10 mCi/mL ɣ-[^{32}P] ATP, 20-25mer oligonucleotide probes (50 pmol/μL; complementary to the sRNA under study, the aptamer, and to the 5S rRNA), and Sephadex G-25 spin column.
8. Membrane washes: 20× SSC (3 M NaCl, 300 mM trisodium citrate pH: 7.0) and 1% SDS.
9. Detection: phosphorimager cassette, screen, scanner, and image analysis software.

2.4 Affinity Purification of Aptamer-Tagged sRNAs

1. Purification: Buffer A (20 mM Tris–HCl pH: 8.0, 150 mM KCl, 1 mM $MgCl_2$, 1 mM DTT), sonicator, MS2-MBP, amylose resin (NEB), disposable chromatography columns, and maltose.
2. RNA extraction and precipitation (*see* **item 2** in Subheading 2.3).
3. Protein precipitation: acetone.

2.5 RT-PCR Analysis

1. Reverse transcription: random hexamers, SuperScript™ II RT, 5× First-Strand Buffer (Invitrogen), 0.1 M DTT, and RNaseOUT™ (Invitrogen).
2. PCR: specific primers to amplify the sRNAs under study and conventional reagents (*see* **item 3** in Subheading 2.2).

2.6 Protein Electrophoresis and Western Blotting

1. Acrylamide gel: Tris–HCl buffer 1.5/1 M pH: 8.8/6.8 resolving/stacking gel; 40% acrylamide/bisacrylamide solution, 10% APS, 10% SDS, TEMED, and butanol.
2. Electrophoresis: 10× SDS running buffer (250 mM Tris base, 1.92 M glycine and 1% SDS in 1 L); protein loading buffer containing β-mercaptoethanol; protein size marker.
3. Gel staining: silver stain kit.
4. Membrane blotting: 3 mm Whatman paper, polyvinylidene difluoride membrane (P 0.45 PVDF, Amersham), methanol, transfer buffer (25 mM Tris pH: 8.3, 192 mM glycine, and 20% methanol).
5. Immunoassay: TBST20 buffer (20 mM Tris–HCl pH: 8.0, 0.18 M NaCl, 0.1% Tween 20), blocking reagent (Amersham),

monoclonal anti-FLAG antibody (1:5000 in TBST20), and anti-mouse antibody conjugated to horseradish peroxidase (1:100,000).

6. Detection: blotting detection reagent (ECL, Amersham) and imaging system.

3 Methods

The experimental strategy to engineer, analyze, and purify aptamer-tagged sRNAs and in vivo bound proteins is overviewed in Fig. 1a. The *S. meliloti* Hfq-dependent *trans*-sRNA AbcR2 [6, 12], which is widely conserved in α-proteobacteria, has been used to successfully implement and optimize the following protocol. Figures 1 and 2 illustrate the detailed protocol using AbcR2 as a model sRNA.

3.1 Plasmid-Based MS2-Aptamer Tagging of sRNAs

sRNAs can be theoretically tagged with a wide variety of aptamer tags (e.g., MS2 tandem repeats, boxB, elF4A, or streptomycin) at any position between its transcription start site (TSS) and the beginning of the terminator [13]. However, experiments in *E. coli* have revealed that MS2 tagging at the sRNA 5′-end yields the most homogenous transcripts and efficient target regulation [13] (*see* **Note 2**). This section describes the 5′-MS2-aptamer tagging of the full-length AbcR2 sRNA (Fig. 1b) under the control of the P_{Lac} promoter in the engineered plasmid pSRK-C, which has been successfully used for constitutive overexpression of sRNAs in rhizobia.

1. The MS2 aptamer is generated by annealing of complementary DNA oligonucleotides that leave *Bam*HI and *Xba*I compatible overhangs. 1 μL of FMS2 and RMS2 oligonucleotides (Table 1) are mixed in a final volume of 10 μL, heated for 5 min at 90 °C, and let cool down to enable primer annealing.
2. The obtained 48 bp fragment is directly ligated into pSRK-C digested with *Bam*HI and *Xba*I enzymes, generating the pSRK-C derivative harboring the MS2 aptamer sequence.
3. *E. coli* DH5α competent cells are used for transformation with ligation reaction and plated in kanamycin-containing LB plates (*see* **Note 3**).
4. Successful integration of the insert is verified by colony PCR using flanking reverse primer PCR1 (CGGGCCTCTTCG CTATT) and forward PCR2 (TTAGCTCACTCATTAGG).
5. Plasmid DNA from colonies showing the correct size is isolated and sequenced with PCR2.

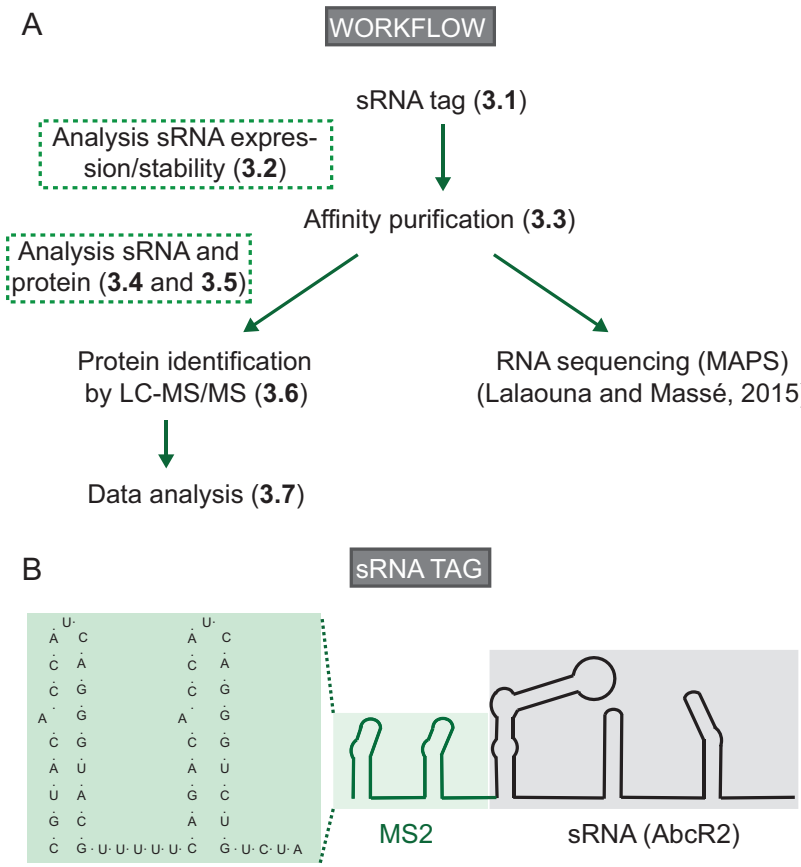


Fig. 1 (**a**) Schematic view of the workflow. (**b**) Schematic representation of the secondary structure of the sRNA AbcR2 tagged with the MS2 epitope. Left panel shows the sequence of the MS2 tandem tag

6. To subsequently generate vectors expressing the aptamer-sRNA fusion (e.g., MS2-AbcR2), primer pairs incorporating *Xba*I and *Hind*III sites to the 5′- and 3′-ends of the sRNA of interest, respectively, are designed.
7. PCR amplification is performed using genomic DNA as template and obtained products are digested with the aforementioned enzymes.
8. In parallel, to construct a control vector with the MS2 aptamer sequence followed by a transcription terminator (e.g., T1), the fragment is generated as in **step 1** by oligonucleotide hybridization using primers T1X/H and T1cX/H listed in Table 1, which leave *Xba*I and *Hind*III compatible overhangs.
9. Purified fragments obtained in **steps 7** and **8** are ligated into *Xba*I-*Hind*III digested pSRK-C containing the MS2 sequence.
10. **Steps 3**–**5** are repeated.
11. Correct plasmids are purified, sequenced, and subsequently transformed into *E. coli* S17.1 cells for biparental mating with the corresponding rhizobial strain (*see* **Note 4**).

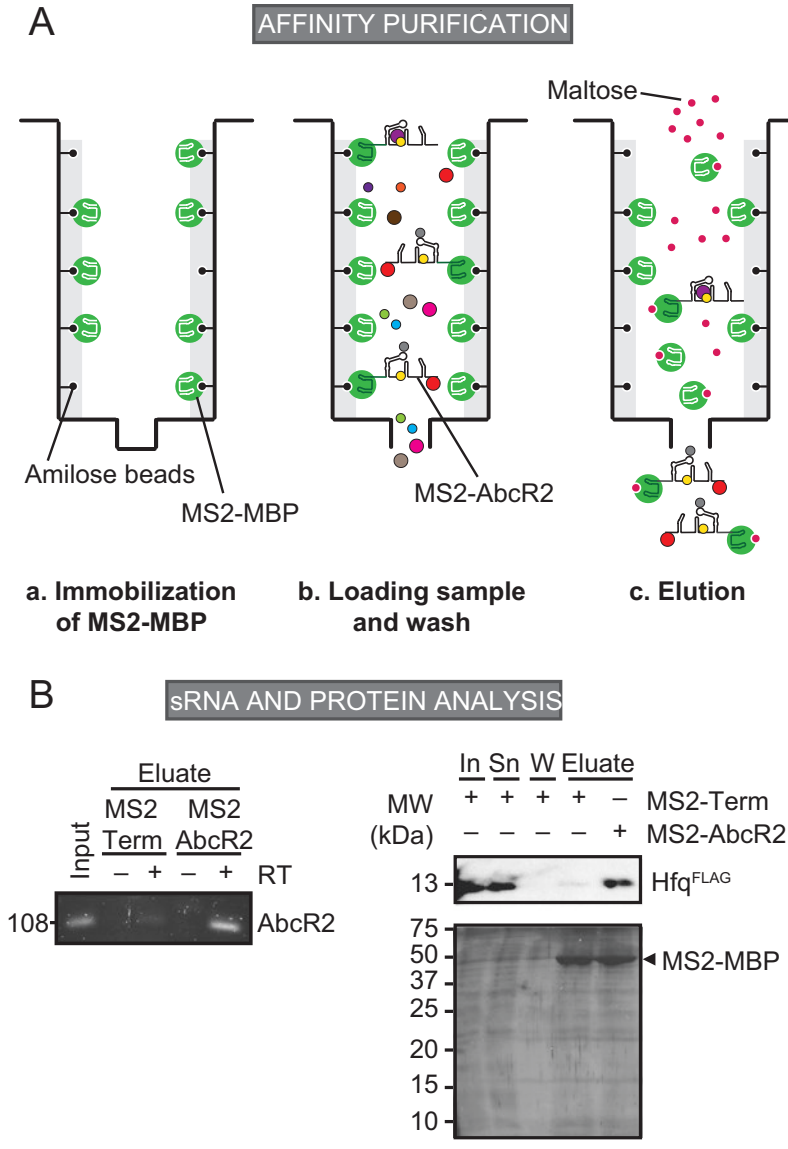


Fig. 2 (**a**) Schematic view of the affinity purification procedure. MS2-MBP protein is immobilized in an amylose column. Cell lysate containing the tagged MS2-AbcR2, previously pre-incubated with MS2-MBP is applied to the amylose column. After several column washes, addition of a maltose enables the elution of both the MS2-AbcR2- and the AbcR2-binding proteins. (**b**) sRNA and protein analysis. Agarose gel (left panel) showing products from RT-PCR reactions for detection of AbcR2 within the eluates from MS2-Term and MS2-AbcR2. Input: total RNA. RT: reverse transcription. Immunoblot and silver-stained polyacrylamide gel (right panel) to monitor specific proteins (Hfq^{FLAG}) and protein pattern across the procedure when using cells expressing MS2-AbcR2 or MS2-Term as control. Input, total lysate; Sn, supernatant after incubation with amylose column; W, wash fraction. MS2-MBP in the silver-stained PAA gel is indicated with an arrow

Table 1
DNA oligonucleotides encoding the MS2 and T1 that leave *Bam*HI/*Xba*I or *Xba*I/*Hind*III compatible overhangs for construction of pSRK-MS2-(Term)

Oligo	Sequence
FMS2	GATCCGTACACCATCAGGGTACGTTTTTCAGACACCATCAGGGTCTGT
RMS2	CTAGACAGACCCTGATGGTGTCTGAAAAACGTACCCTGATGGTGTACG
T1X/H	CTAGATGAAAAAACGACAAAGCAGCACTGATTACAGTGCTGCTTTT TTTATCCCTGTA
T1cX/H	AGCTTACAGGGATAAAAAAAGCAGCACTGTAATCAGTGCTGCTTTG TCGTTTTTTCAT

3.2 Analysis of Aptamer-Tagged sRNAs by Northern Hybridization

Before affinity purification, the impact of tagging on secondary structure, expression, stability, and, if possible, functionality of the sRNA should be assessed (i.e., the Hfq-binding, regulatory potential or susceptibility to RNase degradation). Northern blot hybridization was used to assess the stability and the transcript size of aptamer-tagged sRNAs, which should not be more than 48 nt longer than the wild-type version (size of the tandem MS2 aptamer and the restriction site). As control vector, a construct in which the MS2 aptamer is followed by a transcription terminator (MS2-Term) can be included. An oligonucleotide probe corresponding to the aptamer region is generally used to ensure specificity, but membranes can also be probed with oligonucleotides designed to detect the sRNA under study.

1. Bacterial pellets (*see* details in Subheading 2.1) are split into 1.5 mL microtubes containing cells equivalent to OD_{600} ~3 (e.g., 5 mL culture of OD_{600} 0.6), gently resuspended in 300 μL of lysis solution and incubated for 10 min at 65 °C with regular mixing.
2. Lysates are chilled on ice and 125 μL 5 M NaCl is added to each 1.5 mL microtube.
3. After 10 min on ice, samples are centrifuged (15 min, 16,000 × *g*, 4 °C).
4. The aqueous (upper) phase is transferred to a new 1.5 mL microtube with 1.35 mL of cold 100% EtOH.
5. Tubes are mixed by inversion and stored at −80 °C at least 1 h prior to centrifugation (30 min, 16,000 × *g*, 4 °C).
6. EtOH is completely removed and pellets are resuspended in 42.5 μL of water and pooled together for DNAse I treatment according to the manufacturer's instructions.
7. After incubation, 1× vol of cold phenol:chloroform:isoamyl alcohol is added. Samples are mixed by vortex and the organic

and inorganic phases are separated by centrifugation (15 min, 16,000 × *g*, 4 °C).

8. The aqueous (upper) phase is transferred to a new microtube containing 20 µL 3 M NaAc (pH: 5.2) and 600 µL EtOH and mixed by inversion. **Step 5** is repeated.
9. Supernatant is removed and RNA pellet is washed with 700 µL cold 70% EtOH, avoiding pipetting or vortex.
10. Precipitated RNA is pelleted by centrifugation (30 min, 16,000 × *g*, 4 °C) and supernatant is carefully removed.
11. Samples are air-dried at room temperature with open lids for 10 min.
12. RNA pellets are resuspended in 25 µL of RNase-free water, and RNA concentration is determined by measuring optical density at 260 nm with a Nanodrop device (~1 µg/µL).
13. A 6% polyacrylamide 7 M urea solution is prepared (*see* **Note 5**).
14. Per mL of gel solution, 10 µL of 10% APS and 1 µL of TEMED are added.
15. Gel mixture is immediately poured in between two glass plates, separated by 1-mm spacers, and a comb is inserted (5 mm-depth into the glass plates is sufficient and allows better resolution) avoiding air bubbles (*see* **Note 6**).
16. Electrophoresis device is casted and the required volume of 1× TBE is added.
17. Prior to samples loading, a pre-electrophoresis step is performed (current set at ~30 mA) to warm up the gel.
18. To adjust sample volumes, equal amounts of total RNA (5–15 µg) are dried in a vacuum concentrator and 10–20 µL 1× RNA loading buffer are added.
19. Samples are denatured by 5 min-boiling at 95 °C and cooled on ice.
20. After stopping pre-electrophoresis, gel wells are flushed with 1× TBE with a syringe to remove the urea and unpolymerized acrylamide.
21. Samples are spin down and loaded into the wells (loading volume should not exceed 25 µL). As a size reference, a RNA molecular weight marker (MWM) is loaded.
22. Electrophoresis is set at ~30 mA until bromophenol blue reaches the bottom part of the gel.
23. After the electrophoresis, the lane containing the MWM is excised and stained with ethidium bromide (EtBr) or GelRed.
24. RNA is transferred onto a nylon membrane according to the manufacturer's instructions (*see* **Note 7**) which is subsequently

exposed to UV light at 254 nm for 5 min to enable a covalent link between the RNA and the membrane.

25. The nylon membrane can be then stored or directly placed in a glass hybridization tube, RNA facing inside, with 10–20 mL hybridization buffer and incubated at 42 °C with rotation for 30–120 min in a hybridization oven.
26. Oligonucleotide labeling reaction is set as follows (*see* **Note 8**): 5 μL of RNase-free H_2O, 1 μL of 50 pmol/mL oligonucleotide, 1 μL of 10× reaction buffer, 1 μL of T4-PNK, and 2 μL of γ-[^{32}P]ATP. Labeling reactions are incubated for 1 h at 37 °C.
27. RNAse-free H_2O is added to the reaction (final volume 25 μL) and the mixture is applied to the center of a Sephadex G-25 spin column in which the storage buffer has been already removed by centrifugation.
28. The column is transferred to a new 1.5 mL microtube and centrifuged (2 min, 3500 × *g*) to purify the oligonucleotide probe by eliminating the unincorporated γ-[^{32}P]ATP.
29. Eluted purified RNA probe (25 μL) is heated at 95 °C for 5 min and added into the hybridization bottle (*see* **Note 9**).
30. After overnight incubation at 42 °C, the hybridization solution is discarded and the membrane is washed twice for 5 min with 2× SSC-0.1% SDS solution and twice for 15 min with 1× SSC-0.1% SDS solution. All the washes are performed at the hybridization temperature.
31. Membrane is dried with Whatman paper, wrapped in plastic, and exposed onto a phosphorimager screen overnight.
32. After scanning, the same membrane is stripped by boiling at 95 °C in 0.1% SDS twice for 15 min with shaking. The protocol is repeated from **step 23** using a 5S rRNA probe (TACTCTCCCGCGTCTTAAGACGAA) as loading control.
33. For quantitative comparison of samples, an image analysis software is used.

The functionality of the tagged sRNA, i.e., the ability to regulate its targets or to trigger a specific phenotype upon overexpression, should also be checked in comparison with the untagged sRNA. Detailed methods for in vivo verification of sRNA–mRNA interactions in *S. meliloti* have been already described [14]. If tagging impairs sRNA function, other alternative approaches should be undertaken, e.g., MS2 incorporation into different positions within the sRNA, as detailed previously [13], or the use of biotinylated probes.

3.3 Affinity Purification of Aptamer-Tagged sRNAs

The previously established affinity chromatography protocol for *E. coli* [13, 15] was adapted for *S. meliloti*, assuming that FLAG-tagged Hfq should be detected in the eluates from bacteria harboring MS2-tagged AbcR2 sRNA. The bait protein MS2-MBP was

firstly purified by FPLC over amylose and heparin columns (*see* **Note 10**). A schematic view of the experimental approach to purify cognate proteins bound to MS2-tagged AbcR2 in vivo is represented in Fig. 2a. Briefly, *S. meliloti* expressing MS2-AbcR2 were harvested and disrupted. The whole-cell cleared lysate was then incubated with the MS2-MBP protein that binds the MS2-AbcR2 and applied to an amylose column, which interacts non-covalently with the MBP moiety. After removing unspecific-bound by several column washes, addition of a maltose-buffer disrupts the interactions between the MBP and the amylose from the column, facilitating the elution of both the MS2-AbcR2 and the AbcR2-binding proteins. The detailed protocol is described below. All the steps were performed at 4 °C or on ice.

1. Cells equivalent to 240 OD_{600} (e.g., 100 mL of a culture with OD_{600} ~2.4) were harvested as described in Subheading 2.1 and stored at −80 °C.
2. Bacterial cells were thawed on ice, resuspended in 8 mL buffer A (*see* **Note 11**), and split into 2 mL RNAse-free tubes (6–10 tubes are used to facilitate sonication). An analytical sample is collected at this point for subsequent monitoring of sRNA and protein within this fraction by RT-PCR and Western blot, respectively (*see* Subheadings 3.4 and 3.5).
3. Cells are broken using a sonicator with a microprobe by three rounds of 10 s bursts at 32 W. Lysates are chilled on ice between sonication rounds (*see* **Note 12**).
4. Cell lysates are cleared by centrifugation (15 min, 16,000 × *g*, 4 °C) to remove cell debris.
5. During centrifugation, affinity column is prepared. Columns are washed three times with 800 μL buffer A prior to resin application. On the other hand, amylose beads are briefly centrifuged to remove the storage solution from the affinity medium in a 1.5 mL microtube (100 μL per sample), washed with 800 μL buffer A and resuspended again in 800 μL buffer A to be loaded into the column together with 200 pmol of MS2-MBP (*see* **Note 13**).
6. After centrifugation of lysates, soluble cell fractions are transferred to a new microtube. An analytical sample (input-In) is kept as reference for further analysis. MS2-MBP (200 pmol) is added and the mixture is incubated for 5 min with soft soaking (*see* **Note 14**).
7. Cell lysates containing half of the bait protein mixture were then applied into the amylose column to interact with the MBP moiety (*see* **Note 15**).
8. An aliquot of the flow-through fraction (supernatant–Sn) is collected for reference. Buffer A (600 μL) is added three times to remove unspecific-bound to the column.

9. An aliquot from the wash fraction (wash-W) is collected prior to loading 600 μL of buffer A containing 12 mM maltose to enable the elution of MS2-AbcR2- and AbcR2-binding proteins (*see* **Note 16**).
10. An aliquot of the eluate fraction (eluate-E) is stored separately. For protein and RNA dissociation, 1× vol. phenol:chloroform:isoamylalcohol [25:24:1 (v/v)] is added and mixed by vortex for 20 s.
11. The mixture is centrifuged (30 min, 16,000 × *g*, 4 °C) enabling separation of water and organic phases.
12. RNA precipitation: the aqueous upper phase is transferred to a new microtube. 3× vol. EtOH and 20 μL 3 M NaAc pH: 5.2 are added.
13. Protein precipitation: 3× vol. acetone are added to the organic lower phase.
14. Samples from **steps 12** and **13** are mixed by inversion and stored overnight at −20 °C.
15. RNA and proteins are precipitated by centrifugation (30 min, 16,000 × *g*, 4 °C).
16. RNA and protein pellets are washed with 500 μL of cold 70% EtOH and 500 μL of acetone, respectively.
17. Samples from **step 16** are mixed by tube inversion and centrifuged (10 min, 16,000 × *g*, 4 °C).
18. EtOH or acetone is carefully removed in two steps and pellets are air-dried at room temperature.
19. RNA is resuspended in 10 μL of RNase-free water and stored at −80 °C (*see* **Note 17**), whereas the protein pellet is dissolved in 50 μL 1× protein loading buffer and stored at −20 °C.

3.4 RT-PCR Analysis of (AbcR2) sRNA Co-purification

To address whether the MS2 tag allows enrichment of MS2-sRNA during the purification process compared to control, fractions eluted upon affinity chromatography can be analyzed by RNA reverse transcription followed by PCR to monitor the presence of the sRNA of interest. Figure 2b shows the RT-PCR product for AbcR2 obtained when using the control (MS2-Term) and MS2-AbcR2 eluates, confirming that aptamer-tagged AbcR2 is successfully recovered from the affinity chromatography assay. This assay is performed as follows:

1. After treatment with DNAse I (*see* **Note 16**), eluted RNA samples are diluted 1:10 and 1 μL (2.4 OD_{600} equivalents) is PCR amplified with sRNA specific primers and conventional reagents to rule out the presence of genomic DNA.
2. The remaining 9 μL are subjected to first-strand cDNA synthesis using SuperScript™ II reverse transcriptase. The reaction is set mixing the RNA with 1 μL random hexamers

(100 ng/μL), 1 μL dNTP mix (10 mM each), and water (up to 12.5 μL).

3. The mixture is heated to 65 °C for 5 min and quickly chilled on ice.
4. The following reagents are subsequently added: 4 μL 5× first-strand buffer, 2 μL 0.1 M DTT, and 1 μL RNaseOUT™.
5. The contents are gently mixed and incubated at 42 °C for 2 min (*see* **Note 18**).
6. Finally, 0.5 μL (100 units) of RT are added to half of the reactions, mixed by pipetting, and incubated at 42 °C for 50 min.
7. The reaction is inactivated by heating at 70 °C for 15 min.
8. The first-strand reaction (1 μL) is now used as template for PCR with the same sRNA-amplifying primers and conventional reagents.
9. PCR products are electrophoresed in 2–2.5% agarose gel in TBE to allow for visualization, and PCR fragments of correct size are sequenced.

The remaining RNA sample from the soluble fraction can be subjected to RNA-seq (*see* **Note 17**) to decipher the sRNA "targetome," i.e., the array of mRNAs that are targeted by a sRNA under specific conditions. This approach, known as MAPS (MS2-affinity purification with RNA sequencing) has been already described elsewhere [16].

3.5 Western Blotting Analysis of Co-purified (Hfq) Proteins

To further evaluate the output of the affinity chromatography, protein aliquots collected across the experimental procedure are resolved by SDS-PAGE and visualized by silver staining and/or Western blot as described below (Fig. 2c). Gel profiling allows direct comparison between eluate fractions. Distinct protein patterns across control and experimental samples can be observed anticipating different binding proteomes that must be subsequently analyzed by LC-MS/MS. Of note, MS2-MBP is clearly enriched in all eluate fractions indicating that the affinity purification is working with the same efficiency for control and tagged-sRNA samples.

AbcR2 binds tightly to Hfq both in vivo and in vitro and this strong association has allowed enrichment of AbcR2 by co-immunoprecipitation with Hfq from *S. meliloti* lysates [6]. Given that the affinity chromatography approach recovered stable MS2-AbcR2 (Fig. 2b), Hfq is expected to be present within the MS2-AbcR2 eluate. Figure 2c shows enrichment of Hfq in the MS2-AbcR2 when compared to MS2-term, confirming a successful affinity purification procedure. As it has been already reported in *E. coli* [13], Hfq can weakly associate with MS2 alone, but the significant higher recovery with MS2-AbcR2 indicates that Hfq was purified specifically through the interaction with AbcR2.

1. Aliquots of collected proteins to be resolved by SDS-PAGE are thawed.
2. Two 15% polyacrylamide gels are prepared.
3. Immediately after addition of 10 μL of 10% APS, 10 μL of 10% SDS, and 0.4 μL of TEMED per mL of gel solution, the mixture is poured in between the glass plates up to ~2 cm of the top.
4. Separation gels are overlaid with butanol.
5. After polymerization of the resolving gel, the butanol is removed with filter or Whatman paper.
6. A 5% polyacrylamide stacking gel is prepared with the same proportion of APS, SDS, and TEMED per mL as for the separating gel (*see* **step 3**), mixed to homogeneity, and poured on top of the resolving gel.
7. Gel combs are immediately placed over the gel avoiding bubbles formation.
8. Two sets of protein samples are prepared: one for subsequent gel staining with silver containing aliquots equivalent to 0.05 OD of the lysate, flow-through and wash fractions and other set to 0.2 OD for Western blot. Half of the elution fractions can be loaded into each gel (120 OD).
9. Samples are resuspended in 2× protein loading buffer and, together with half of the elution fractions (120 OD), denatured by heating at 95 °C for 5 min.
10. Electrophoresis device is casted and the necessary 1× SDS-running buffer is added.
11. After gel polymerization, combs are removed and wells are carefully flushed with running buffer prior to sample and protein marker loading.
12. Electrophoresis is set at ~30 mA until desired protein separation, monitored by the pre-stained markers (1.30–2 h).
13. One gel is subsequently stained with silver for protein visualization.
14. For immunoblot analysis, prior to protein transfer, the PVDF membrane must be shortly activated in methanol, washed with ddH_2O, and finally immersed in 1× transfer buffer, together with the gel and the Whatman paper (*see* **Note 7**).
15. Proteins are blotted onto the PVDF typically for 50 min at 0.8 mA/cm^2. At this point, the membrane can be stored at 4 °C (*see* **Note 19**) or directly subjected to immunoassay at room temperature with shaking.
16. Membrane is incubated 1 h in TBST20 with 1.5% blocking reagent.

17. After a short rinse with TBST20, membrane is incubated for 1 h with a monoclonal anti-FLAG antibody (1:5000 in TBST20) to detect FLAG-tagged Hfq.
18. Membrane is then rinsed and washed in TBST20 six times for 5 min each.
19. After removal of the primary antibody, membrane is incubated with an anti-mouse antibody conjugated to horseradish peroxidase (1:100,000 in TBST20).
20. Membrane is then rinsed and washed again in TBST20 six times for 5 min each.
21. Proteins are detected by incubation with blotting detection reagent (5 min, room temperature in the dark) and visualized with an image documentation system.

3.6 Protein Identification by Mass Spectrometry

The protein complexity of the eluted fractions can be examined by mass spectrometry with the aim of identifying novel protein partners for specific sRNAs. This analysis will help to gain insights into the unknown cellular functions of the sRNA of interest. The combination of this technique together with the RNA-seq can shed some light on the role of sRNAs in *S. meliloti*. The protocol detailed below can be adapted depending on the proteomics facility available (e.g., mass spectrometer can vary and therefore the settings described below).

1. Protein samples equivalent to 120 OD_{600} were run 10 min in a 4% SDS-PAGE.
2. Each gel lane corresponding to the different samples was cut into 10 slices.
3. Gel slices are subjected to in-gel tryptic manual digestion.
4. The resulting peptides are fractionated using an Easy n-LC II chromatography system (Proxeon) in line with an Amazon Speed ETD mass spectrometer (Bruker Daltonics).
5. CaptiveSpray Ion Source (Bruker Daltonics) set at 1300 V is used to ionize the peptides.
6. Amazon Speed ETD Ion Trap mass spectrometer controlled by TrapControl software v7.2 (Bruker-Daltonics) and operated in AutoMS(2) acquisition mode is used to acquire tandem mass spectra.
7. Ion Trap set to analyze the survey scans in the mass range m/z 400–1400 in Enhanced Resolution MS mode and the top ten multiply charged ions in each duty cycle selected for MS/MS in UltraScan MS/MS mode.
8. DataAnalysis software v4.3 (Bruker-Daltonics) and search against the UniProtTrembl database using Mascot 2.4 (Matrix Science) integrated together with ProteinScape v4.0 (Bruker-Daltonics) was used to process the raw data files.

9. Peptide precursor mass tolerance was set at 0.5 Da, and MS/MS tolerance was set at 0.5 Da. Search criteria included carbamidomethylation of cysteine (+57.02 Da) as a fixed modification and oxidation of methionine (+15.99 Da) as a variable modification.
10. A maximum of 2 missed cleavages for tryptic digestion was set to perform the search.
11. The reverse database search option was enabled and all peptide data was filtered to satisfy false discovery rate (FDR) of ≤2%.

Detected proteins associated in vivo with the sRNAs identified using the above-described protocol may be unspecific or form part of complexes. Therefore, a set of protein candidates should be validated, e.g., by immunoprecipitation followed by sRNA co-purification detection as described before [6]. Furthermore, functionality of the sRNA–protein complex should be tested, ideally comparing wild-type and derivative strains carrying a deletion in the protein coding gene. Anyway, a suitable assay depending on the expected protein function should be designed (Northern blot to check sRNA stability in the absence of the protein, sRNA degradation assays, or double plasmid assay to assess protein involvement in target regulation).

3.7 Data Analysis

Once obtained the list of proteins identified by mass spectrometry, it is essential to demarcate the range of nonspecific binding proteins. The nonspecific proteins identified by affinity purification experiments can vary considerably and depend on the experimental conditions and also the construction used to assess the contamination, i.e., different proteins will associate with the MS2 moiety itself or with the type of matrix used for the purification. Thus, all proteins associated with negative controls, either containing only the MS2 aptamer or the sRNAs should be discarded from the analysis. To ensure that proteins cataloged as putative contaminants are indeed indirect partners, it is important to analyze the correlation between samples, calculating the Pearson correlation coefficient. When different proteins are identified across samples and corresponding controls, it is expected to have a bad linear correlation between samples ($r < 0.4$). However, when samples present a similar protein pattern or when biological replicates are performed, a higher correlation is expected ($r > 0.75$). This criterion can be used to remove potential contaminants although is commonly accepted to remove only those proteins not significantly enriched in the samples when compared to the corresponding control after a label-free quantification analysis. The major limitation of this quantification is the number of biological and technical replicates required. Therefore, when limited samples are available for the analysis and quantification cannot be performed, proteins identified in control samples should be excluded to avoid false positives.

Stringent thresholds can then be applied to redefine the protein list. One commonly applied threshold is the exclusion of those proteins identified by less than two peptides. However, the length of the protein must be considered when applying this criterion. Shorter proteins, such as Hfq, will produce fewer unique peptides, and thus might be artificially excluded using the two peptide cutoff.

Gene Ontology (GO) enrichment analysis is performed to explore protein properties. Several online tools are available for this purpose; some examples are DAVID [17] or comparative GO [18]. The main difference between these two tools is the statistical method applied to calculate the enrichment level. DAVID calculates the p-value of the enrichment by the Fischer exact test and allows for p-value correction. Conversely, comparative GO calculates the p-value by a hypergeometric test, resulting in a more relaxed analysis. Regardless the online tool chosen for the enrichment analysis, it is essential that the number of genes annotated for the corresponding strain and species, in this case *S. meliloti* 1021 would be complete, otherwise the results obtained will be biased. In other words, the selected tool must have the majority of the proteins, if not all, annotated for the organism of interest to be used as the reference proteome. Moreover, DAVID allows for exploring other protein properties such as protein domains or cellular pathways where these proteins are involved.

4 Notes

1. The majority of the commercial kits for RNA isolation are based on columns that do not retain RNA molecules smaller than ~200 nt and, therefore, to ensure that isolation of the small RNA fraction, RNA purification should be performed with specifically designed kits like the miRNeasy Mini Kit from QIAGEN [19].
2. The existence of alternative 5′ processed ends in sRNAs (e.g., EcpR1 sRNA) [19], which can be anticipated in RNAseq data and confirmed by Northern blot analysis, should be taken into account to implement this strategy. Cloning of the functional stable sRNA version starting at the processed end could be the best strategy to handle with these molecules, avoiding loss of the aptamer sequence by ribonucleolytic activity on the full-length version.
3. Blue/white selection of transformants with IPTG and X-gal is not possible in pSRK-C derivatives.
4. Preferably, a markerless deletion mutant of the sRNA under study should be used as recipient strain.

5. The mix should be heated to completely dissolve urea, filtered, and stored at 4 °C if not used immediately.
6. Polymerization takes approximately 1 h, but at this point the cast gel can be stored overnight at 4 °C.
7. For electrophoretic transfer during Northern and Western blotting, membrane and Whatman papers are cut slightly larger than the size of the acrylamide gel. A dry Whatman paper is placed onto the gel to facilitate gel removal and prevent introduction of air bubbles. The gel and the membrane are then surrounded by 3 Whatman papers, previously soaked in the corresponding transfer buffer and placed on the electroblot transfer device following the right order, considering that RNA and proteins migrate to the positively charged pole.
8. If detection using radioactivity is not possible, digoxigenin-labeled RNA (riboprobes) or DNA probes can also be used. Even though that detection may not be as sensitive as when using radioactive-labeled oligonucleotides, the fact that riboprobes are homolog to the (full) sRNA coding sequence and internally labeled, make these methods reliable for sRNA detection by Northern hybridization. Membranes hybridized with either nonradioactive labeled riboprobes (synthesized with the Maxiscript kit from Ambion) or DNA probes obtained by PCR can be subjected to chemiluminescence detection (DIG Luminescent Detection Kit, Roche) as described before [19, 20].
9. To avoid signal background, do not apply the probe directly on the membrane but on the hybridization buffer.
10. Recombinant MS2-MBP is composed by the MS2 coat protein N-terminally fused to maltose-binding protein. The fusion protein purification is described in [21]. Protein purity is checked by SDS-PAGE and Commassie blue gel staining and concentration is determined with Bradford dye-binding method (Bio-Rad).
11. Establishment and maintenance of RNA–protein complexes considerably depends on salt concentration and Hfq has been reported to co-purify better with InvR-MS2 RNA increasing concentration of KCl in Buffer A [13, 22]. We also compared MS2-AbcR2 eluates obtained with different KCl concentrations (150 mM, 500 mM and 1 M). However, we did not observe better binding of FLAG-tagged Hfq or other proteins with KCl concentrations higher than 150 mM, but an excess of salt precipitates in the eluates that hampered subsequent steps.
12. Alternatively, rhizobial cells can also be disrupted in a French press.
13. Considering the starting amount of cells and the conditions used here, a total of 400–500 pmol was the optimal amount of

MS2-MBP to recover FLAG-tagged Hfq with MS2-AbcR2 and, therefore, is recommended in this protocol. Lower amounts (100–200 pmol) of MS2-MBP did not recover Hfq so efficiently and increasing MS2-MBP concentration (1000 pmol) did result in more purified Hfq yield, but seemed to lead to unspecific protein binding.

14. Brief pre-incubation of MS2-MBP protein with the lysate prior to loading into the column had a positive effect on Hfq recovery with MS2-AbcR2, disregarding the incubation time. Considering that some MS2-MBP may not bind efficiently to the tagged-sRNA during incubation with the lysate or once into the column, the effect of additive application in two steps (200 pmol MS2-MBP each) was also tested. This experimental set-up showed the best Hfq recovery rates when MS2-AbcR2 was used as a bait and thus it is recommended.
15. If the column flow rate is high, the lysate can be applied twice to the resin, increasing the chance for complex binding.
16. DNAse I digestion of total or part of the eluates can be performed at this step to enable subsequent RT-PCR analysis of this fractions without further phenol-chloroform treatment.
17. If the samples are going to be either stored for long periods or shipped, RNA sample pellets can be stored dry.
18. A conventional PCR thermocycler can be set to follow the RT steps to avoid using multiple incubators.
19. If the PVDF membrane gets dry during storage, it must be shortly activated in methanol again prior to proceeding with immunoassay.

References

1. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. Nat Rev Microbiol 9(8):578–589. https://doi.org/10.1038/nrmicro2615
2. Smirnov A, Förstner KU, Holmqvist E, Otto A, Günster R, Becher D, Reinhardt R, Vogel J (2016) Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. Proc Natl Acad Sci U S A 113(41):11591–11596. https://doi.org/10.1073/pnas.1609981113
3. Sobrero P, Valverde C (2012) The bacterial protein Hfq: much more than a mere RNA-binding factor. Crit Rev Microbiol 38(4):276–299. https://doi.org/10.3109/1040841X.2012.664540
4. Cavanagh AT, Wassarman KM (2014) 6S RNA, a global regulator of transcription in *Escherichia coli*, *Bacillus subtilis*, and beyond. Annu Rev Microbiol 68(1):45–60. https://doi.org/10.1146/annurev-micro-092611-150135
5. Babitzke P, Romeo T (2007) CsrB sRNA family: sequestration of RNA-binding regulatory proteins. Curr Opin Microbiol 10(2):156–163. https://doi.org/10.1016/j.mib.2007.03.007
6. Torres-Quesada O, Reinkensmeier J, Schlüter JP, Robledo M, Peregrina A, Giegerich R, Toro N, Becker A, Jiménez-Zurdo JI (2014) Genome-wide profiling of Hfq-binding RNAs uncovers extensive post-transcriptional rewiring of major stress response and symbiotic regulons in *Sinorhizobium meliloti*. RNA Biol 11(5):563–579
7. Torres-Quesada O, Millán V, Nisa-Martínez R, Bardou F, Crespi M, Toro N, Jiménez-Zurdo JI (2013) Independent activity of the homologous small regulatory RNAs AbcR1 and AbcR2 in the legume symbiont *Sinorhizobium meliloti*. PLoS One 8(7):e68147. https://doi.org/10.1371/journal.pone.0068147

8. Beringer JE (1974) R factor transfer in *Rhizobium leguminosarum*. J Gen Microbiol 84(1):188–198
9. Robertsen BK, Aman P, Darvill AG, McNeil M, Albersheim P (1981) Host-symbiont interactions: V. The structure of acidic extracellular polysaccharides secreted by *Rhizobium leguminosarum* and *Rhizobium trifolii*. Plant Physiol 67(3):389–400
10. Waters LS, Storz G (2009) Regulatory RNAs in bacteria. Cell 136(4):615–628. https://doi.org/10.1016/j.cell.2009.01.043
11. Khan SR, Gaines J, Roop RM, Farrand SK (2008) Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid *quorum sensing*. Appl Environ Microbiol 74(16):5053–5062. https://doi.org/10.1128/AEM.01098-08
12. del Val C, Rivas E, Torres-Quesada O, Toro N, Jiménez-Zurdo JI (2007) Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. Mol Microbiol 66(5):1080–1091. https://doi.org/10.1111/j.1365-2958.2007.05978.x
13. Said N, Rieder R, Hurwitz R, Deckert J, Urlaub H, Vogel J (2009) *In vivo* expression and purification of aptamer-tagged small RNA regulators. Nucleic Acids Res 37(20):e133. https://doi.org/10.1093/nar/gkp719
14. Jiménez-Zurdo JI, Robledo M (2015) Unraveling the universe of small RNA regulators in the legume symbiont *Sinorhizobium meliloti*. Symbiosis 67(1–3):43–54. https://doi.org/10.1007/s13199-015-0345-z
15. Corcoran CP, Rieder R, Podkaminski D, Hofmann B, Vogel J (2012) Use of aptamer tagging to identify *in vivo* protein binding partners of small regulatory RNAs. Methods Mol Biol 905:177–200. https://doi.org/10.1007/978-1-61779-949-5_11
16. Lalaouna D, Massé E (2015) Identification of sRNA interacting with a transcript of interest using MS2-affinity purification coupled with RNA sequencing (MAPS) technology. Genom Data 5:136–138. https://doi.org/10.1016/j.gdata.2015.05.033
17. Huang d W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4(1):44–57. https://doi.org/10.1038/nprot.2008.211
18. Fruzangohar M, Ebrahimie E, Ogunniyi AD, Mahdi LK, Paton JC, Adelson DL (2013) Comparative GO: a web application for comparative gene ontology and gene ontology-based gene selection in bacteria. PLoS One 8(3):e58759. https://doi.org/10.1371/journal.pone.0058759
19. Robledo M, Frage B, Wright PR, Becker A (2015) A stress-induced small RNA modulates alpha-rhizobial cell cycle progression. PLoS Genet 11(4):e1005153. https://doi.org/10.1371/journal.pgen.1005153
20. Robledo M, Jiménez-Zurdo JI, Becker A (2015) Antisense transcription of symbiotic genes in *Sinorhizobium meliloti*. Symbiosis 67(1):55–67. https://doi.org/10.1007/s13199-015-0358-7
21. Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA 8(4):426–439
22. Lohman TM, Overman LB, Ferrari ME, Kozlov AG (1996) A highly salt-dependent enthalpy change for *Escherichia coli* SSB protein-nucleic acid binding due to ion-protein interactions. Biochemistry 35(16):5272–5279. https://doi.org/10.1021/bi9527606

Part V

Applications in Synthetic Biology

Chapter 21

A Modular Genetic System for High-Throughput Profiling and Engineering of Multi-Target Small RNAs

Samuel D. Stimple, Ashwin Lahiry, Joseph E. Taris, David W. Wood, and Richard A. Lease

Abstract

RNA biology and RNA engineering are subjects of growing interest due to recent advances in our understanding of the diverse cellular functions of RNAs, including their roles as genetic regulators. The noncoding small RNAs (sRNAs) of bacteria are a fundamental basis of regulatory control that can regulate gene expression via antisense base-pairing to one or more target mRNAs. The sRNAs can be customized to generate a range of mRNA translation rates and stabilities. The sRNAs can be applied as a platform for metabolic engineering, to control expression of genes of interest by following relatively straightforward design rules (Kushwaha et al., ACS Synth Biol 5:795–809, 2016). However, the ab initio design of functional sRNAs to precise specifications of gene control is not yet possible. Consequently, there is a need for tools to rapidly profile uncharacterized sRNAs in vivo, to screen sRNAs against "new/novel" targets, and (in the case of metabolic engineering) to develop engineered sRNAs for regulatory function against multiple desired mRNA targets. To address this unmet need, we previously constructed a modular genetic system for assaying sRNA activity in vivo against specifiable mRNA sequences, using microtiter plate assays for high-throughput productivity. This sRNA design platform consists of three modular plasmids: one plasmid contains an inducible sRNA and the RNA chaperone Hfq; the second contains an inducible fluorescent reporter protein and a LacY mutant transporter protein for inducer molecules; and the third plasmid contains a second inducible fluorescent reporter protein. The second reporter gene makes it possible to screen for sRNA regulators that have activity against multiple mRNAs. We describe the protocol for engineering sRNAs with novel regulatory activity using this system. This sRNA prototyping regimen could also be employed for validating predicted mRNA targets of uncharacterized, naturally occurring sRNAs or for testing hypotheses about the predicted roles of genes, including essential genes, in cellular metabolism and other processes, by using customized antisense sRNAs to knock down or tune down gene expression.

Key words Small regulatory sRNA, In vivo fluorescent reporter gene assay, High-throughput screen, Bacterial genetics, DsrA, Multi-target sRNA engineering, sRNA engineering platform, sRNA characterization

---

Samuel D. Stimple and Ashwin Lahiry contributed equally to this work.

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8_21,

1 Introduction

As part of a broader and deeper understanding of the roles of RNA in biology, there has been a tremendous increase in the number and relevance of sRNAs found in bacteria. However, despite their critical function in many bacterial stress and environmental responses, many naturally occurring small RNAs (sRNAs) remain poorly understood and inadequately characterized. Tools for rapidly validating putative sRNA:mRNA interactions could therefore have a significant impact on our understanding of RNA biology.

Although several noncoding sRNAs act by titrating effector proteins, many sRNAs regulate translation of their target mRNAs via antisense base-pairing interactions [1]. This type of regulation by sRNAs typically occurs through one of several routes: (1) sRNA binding can interfere with ribosome access to the mRNA, decreasing the translation rate; (2) binding can destabilize the mRNA typically by exposing or creating an RNase-sensitive site, enhancing the mRNA turnover rate; or (3) sRNA binding can lead to a structural rearrangement of the mRNA leader, exposing the ribosome-binding region and typically increasing the translation rate. These sRNAs show great potential as modular, flexible, and portable genetic regulators and can be readily tailored for specific applications [2] both in model organisms and in genetically intractable hosts [3]. By using a chimeric fluorescent reporter gene fused to the translation initiation region (TIR) for an mRNA of interest, we can study the effect of a given sRNA on the translation of the mRNA-encoded protein during cell growth. Others have used colorimetric assays for surveys of sRNA function at mRNAs [4, 5]. In many sRNA:mRNA studies that use reporter genes to quantify sRNA activity, it is common to see quantification of sRNA activity at a single fixed time point [6–9]. Others have followed a single fluorescent reporter activity during growth from lag phase through early log phase [10, 11].

We sought to increase the range of analysis of sRNA genetic function to include assays of multi-targeting sRNAs that coordinate regulation of several transcripts simultaneously. Accordingly, we created a second reporter gene with a distinct fluorescent reporter. The two fluorescent proteins that we chose, namely GFPuv and mCherry, have minimal or no overlap in their fluorescent excitation and emission spectra. To permit combinatorial and high-throughput analyses, we elected to put each reporter gene on a separate compatible plasmid, and the vector that produces the sRNA on a third compatible plasmid, each with its own drug resistance gene [3]. Each transcript (two reporters and one sRNA) is repressed by orthogonal repressor proteins (AraC, TetR, LacI) that

can be separately induced by their respective small molecule inducers: arabinose (ara), anhydrotetracycline (aTet), and isopropyl β-D-1-thiogalactopyranoside (IPTG). This *E. coli* system is amenable to analysis by flow-cytometry, but for high-throughput screening it is convenient to use a plate reader equipped with fluorescent readout mode. Because the assay is performed during the growth of the strain, the effects of the sRNA in real time, on growth rate, and on the simultaneous expression control of the two mRNAs can be captured by this method.

We constructed our 3-plasmid genetic system for engineering sRNA variants in *E. coli* with the goal of implementing these prototype sRNAs in other organisms. As a proof-of-principle we created retargeted sRNAs that are capable of simultaneously tuning translation in *E. coli* from two reporter genes with TIR sequences derived from mRNAs in the ABE (acetone, butanol, ethanol) fermentation pathway of *Clostridium acetobutylicum*. These two reporter genes (*hydA* and *buk*) and one of the prototype dual-acting retargeted sRNAs generated by this system (DsrA-*buk*'1.1-*hydA*'2.4.1) form the basis of the current generation of our modular three-plasmid system (Fig. 1).

Prototype sRNAs can be readily designed to target new mRNAs by semirational design using an antisense sequence-tiling method [3, 9], with the aid of free energy-based simulations and structural prediction software [12, 13], and can be rapidly tested for regulatory function in the 3-plasmid system. One innovation is the use of structured stem-loop antisense modules (fingerloops; Fig. 2) for production of antisense sequences in the sRNA [3]. For this approach, the antisense region is incorporated on one strand of the stem and into one loop (Fig. 2b, c, letter N symbolizes antisense nucleotides), and the complementary strand of that helix is reconstructed de novo (lowercase letter n reconstructs the helix). We use this second strand of the stem-loop RNA structure to create Watson-Crick base pairs, along with the judicious use of G•U base pairs or bulge mismatches to scale the predicted free energy of stem formation. Typically, we clone these variant sRNA libraries as annealed oligonucleotide pairs (Fig. 2b) that create compatible ends for existing restriction sites. We describe the methodology for creating these customized fingerloops and for using our platform to generate sRNAs that regulate the translation of up to two mRNA(s) of interest. The protocol could easily be exploited for validating predicted activity of naturally occurring sRNAs against putative mRNA targets, for in-depth characterization of poorly understood sRNAs, or for surveying gene-knockdown effects by targeting desired genes or pairs of genes simultaneously.

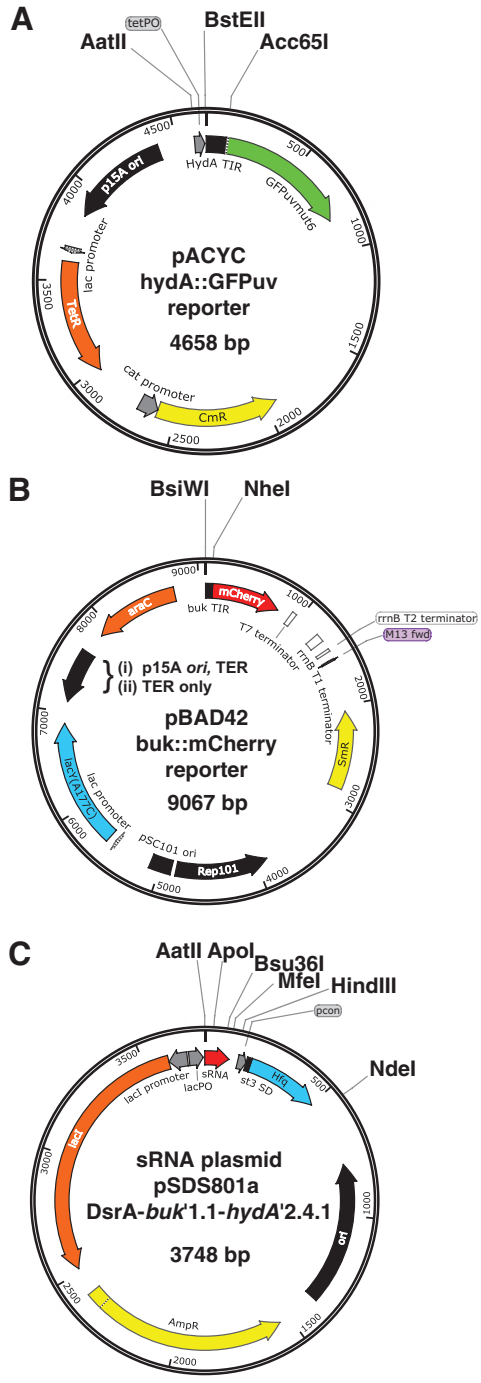


Fig. 1 Plasmid maps for the three-plasmid system. Arrows indicate gene orientations. (**a**) The GFPuv reporter plasmid on the pACYC backbone; (**b**) The mCherry reporter plasmid on the pSC101 backbone; (**c**) The sRNA plasmid on the pBR322 backbone. Note that there are two versions of the pBAD plasmid that contain either an *ori* p15A (and its internal terminator) or a synthetic terminator, in addition to the pSC101 origin (*see* **Note 1**)

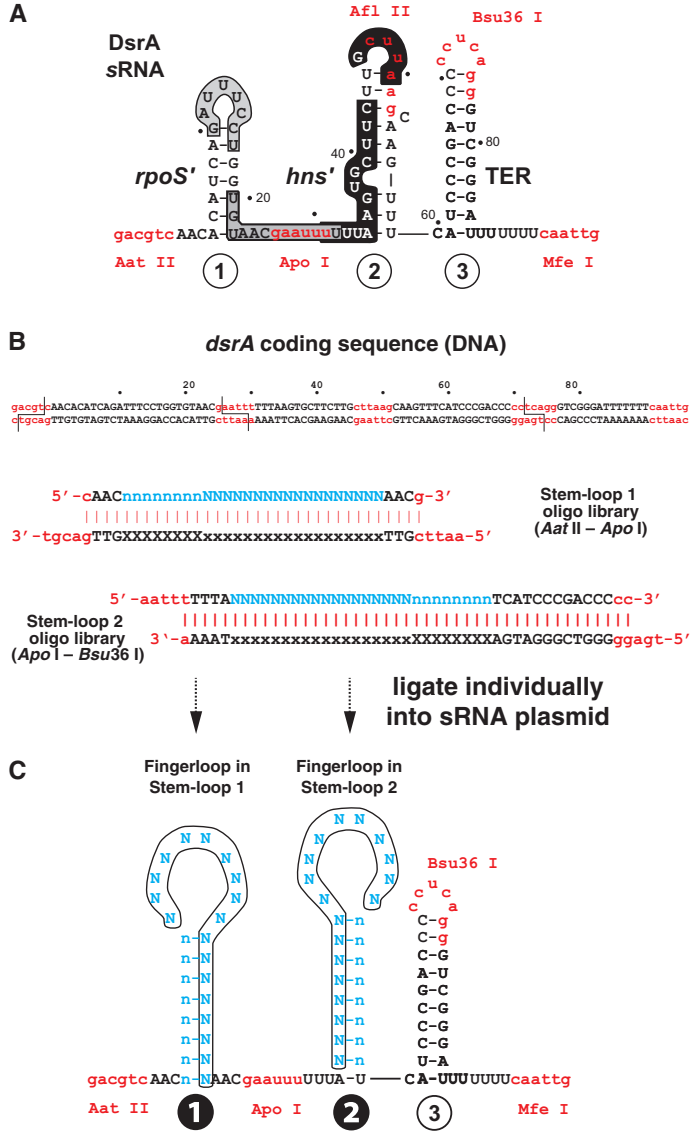


Fig. 2 Modular cloning of sRNA antisense domains. (**a**) The secondary structure model of DsrA [13, 21–24] in the sRNA plasmid context, with the location of corresponding restriction sites highlighted in red. The use of restriction sites Aat II and Apo I, or Apo I and Bsu36 I, for cloning, respectively, permit modular exchange of stem-loops 1 or 2. (**b**) The *dsrA* DNA gene sequence with cleavage patterns from Aat II, Apo I, and Bsu36 I digestion superposed on the restriction sites (lowercase and red letters). Pairs of annealed DNA oligonucleotides can thus be designed to create compatible ends that introduce synthetic fingerloops or unstructured antisense regions as desired. Uppercase N (in cyan) represents the 18-mer antisense sequence that targets a transcript; lowercase n (cyan) represents the sequence used to complete the fingerloop structure; lowercase and uppercase X sequences encode complementary Watson-Crick base pairs in the opposite oligo strand. This process creates modular dsDNA structured antisense fragments to be ligated for creation of fingerloops in sRNAs. (**c**) The orientation of fingerloops in a DsrA derivative that mimics the DsrA native structure. Mismatches can be engineered into stems by altering single "n" nucleotides. The lengths of loop and stem sequences could also be changed to vary regulatory outcomes (**Notes 9–11**)

2 Materials

2.1 Modular Vector Cloning

2.1.1 Plasmids and Strains (Available upon Request; Fig. 1)

1. Reporter plasmid (RP) 1: pACYC *hydA::gfp-uv*$_{mut6}$ *tet*^{r} *cam*^{r}.
2. RP 2: pBAD42 *buk::mCherry araC lacY*^{A177C} *spc*^{r}.
3. sRNA vector: pSDS801a *dsrA (buk'1.1 hydA'2.4.1) lacI*$^{+}$ *hfq*$^{+}$ *amp*^{r}.
4. Reporter 1 control: pACYC184 Δ*tetA cam*^{r}.
5. Reporter 2 control: pBAD42 *spc*^{r}.
6. sRNA vector control: pSDS1002 (=pBR322 Δ*tetA amp*^{r}*)*.
7. Delta-sRNA vector control: pSDS801a Δ*dsrA lacI*$^{+}$ *hfq*$^{+}$ *amp*^{r}.
8. Host strain*: E. coli* CM1000 (= MG1655 Δ*lacX*74 *dsrA*14) [14] (*see* **Note 1**).

2.1.2 Restriction Enzymes

1. BstE II-HF, Acc65 I, BsiW I-HF, Nhe I-HF, Hind III-HF, Nde I, Aat II, Apo I-HF, Afl II, Bsu36 I, Mfe I-HF restriction enzymes and their corresponding 10× NEB reaction buffers.
2. T4 DNA ligase and 10× ligation reaction buffer.
3. Phusion® High-fidelity DNA polymerase, dNTPs, and 10× Phusion reaction buffer.

2.1.3 Media and Antibiotic Stocks

1. Luria-Bertani lysogeny broth media (per L: 10 g Bacto Tryptone, 5 g Bacto Yeast Extract, 10 g NaCl). Sterilize by autoclaving. For plates, add 15 g/L Bacto Agar before autoclaving.
2. SOC media: per L: 20 g Bacto Tryptone, 5 g Bacto Yeast Extract, 2 mL 5 M NaCl, 2.5 mL 1 M KCl; supplement with 10 mL 1 M $MgCl_2$, 10 mL 1 M $MgSO_4$, 20 mL 1 M glucose. Sterilize by autoclaving. Separately filter-sterilize magnesium salts and glucose solutions to be added after the other ingredients cool to ~50 °C.
3. Ampicillin (amp): 200 mg/mL stock solution. Sterilize by filtration. 200 μg/mL final concentration in media.
4. Chloramphenicol (cam): 34 mg/mL stock solution (in EtOH). 25 μg/mL final concentration in media.
5. Spectinomycin (spec): 100 mg/mL stock solution. Sterilize by filtration. 75 μg/mL final concentration in liquid media; 175 μg/mL final concentration for plates.
6. DH5α chemically competent cells. (*see* **Note 2**).

2.2 In Vivo Fluorescence Assay

2.2.1 Growth Media

1. LB media: per L: 10 g Bacto Tryptone, 5 g Bacto Yeast Extract, 10 g NaCl. Sterilize by autoclaving. For plates, add 15 g/L Bacto Agar before autoclaving.

2.2.2 Assay Media

1. 5× M9 media salts solution: per L: 64 g Na_2HPO_4•$7H_2O$, 15 g KH_2PO_4, 5 g NH_4Cl, 2.5 g NaCl. Sterilize by autoclaving.
2. 1 M $MgSO_4$. Sterilize by filtration.
3. 1 M $CaCl_2$. Sterilize by filtration.
4. 10% (w/v) casamino acids solution. Sterilize by autoclaving.
5. 10,000× NBS trace metals solution [15]: per L: 5 g NaCl, 1 g $ZnSO_4$•$7H_2O$, 4 g $MnCl_2$•$4H_2O$, 4.75 g $FeCl_3$•$6H_2O$, 0.4 g $CuSO_4$•$5H_2O$, 0.575 g H_3BO_3, 0.5 g Na_2MoO_4•$2H_2O$, 12.5 mL 6 N•H_2SO_4. Sterilize by filtration.
6. 20% (w/v) glucose solution. Sterilize by filtration.
7. 5% (w/v) Bacto Tryptone solution. Sterilize by autoclaving.
8. M9 liquid media, supplemented with 0.2% (w/v) Casamino acids. For 1 L, combine:
   - 200 mL 5× M9 media salts solution.
   - 700 mL sterile distilled H_2O.
   - 2 mL 1 M $MgSO_4$.
   - 20 mL 10% (w/v) casamino acids.
   - 100 µL 1 M $CaCl_2$.
   - Adjust final volume to 1000 mL with sterile water.
9. M9 + Glucose + Metals M9GM: M9 liquid media supplemented with 4 g/L of glucose and 1× NBS trace metals. For 1 L, combine:
   - 200 mL 5× M9 media salts solution.
   - 700 mL sterile distilled H_2O.
   - 2 mL 1 M $MgSO_4$.
   - 20 mL 10% (w/v) casamino acids.
   - 100 µL 1 M $CaCl_2$.
   - 100 µL 10,000× NBS trace metals solution.
   - 20 mL 20% glucose.
   - Adjust final volume to 1000 mL.
10. M9 + GMT: M9 + GM supplemented with 1% (w/v) tryptone (*see* **Note 3**).
    - 200 mL 5× M9 media salts solution.
    - 500 mL sterile distilled H_2O.
    - 2 mL 1 M $MgSO_4$.
    - 20 mL 10% (w/v) casamino acids.

- 100 µL 1 M $CaCl_2$.
- 100 µL 10,000× NBS trace metals solution.
- 20 mL 20% glucose.
- 200 mL 5% Bacto Tryptone.
- Adjust final volume to 1000 mL.

11. Mineral oil.

*2.2.3 Inducers for Assays (see **Note 4**)*

1. Isopropyl β-D-1-thiogalactopyranoside (IPTG). 1 M stock solution, sterilize by filtration, store at −20 °C.
2. Anhydrotetracycline (aTet). 1 mg/mL stock solution in EtOH. Store under foil or in opaque tubes at −20 °C.
3. L-arabinose (ara). 20% (w/v) stock solution, sterilize by filtration.

2.2.4 Plate Reader

1. Biotek Synergy 2 Multi-Mode microplate reader (*see* **Note 5**).
2. 96-well Corning 3603 fluorescence microtiter plate (sterile, from manufacturer)
3. Plate reader filters and mirrors for $GFPuv_{mut6}$:
   - Excitation filter (395 ± 10 nm).
   - Emission filter (528 ± 10 nm) with a 435 nm-cutoff dichroic mirror.
4. Plate reader filters and mirrors for mCherry.
   - Excitation filter (585 ± 5 nm).
   - Emission filter (620 ± 7.5 nm) with a 595 nm-cutoff dichroic mirror.

2.3 Data Analysis Using Excel Template

1. Excel template for rapid data analysis (Supplementary Materials).

3 Methods

3.1 Modular Vector Cloning: Reporters (see Note 6)

3.1.1 Cloning into the GFPuv Reporter Plasmid

1. PCR-amplify the mRNA leader (translation initiation region or TIR) sequence for the gene of interest. Alternatively, prepare oligos that will generate a deliberate primer-dimer for the leader sequence. Use BstE II (upstream) and Acc65 I (downstream) as the flanking restriction sites (*see* **Note 7**).
2. Digest 1 µg pACYC *hydA-gfp-uv*$_{mut6}$ RP and the PCR product with BstE II-HF and Acc65 I for 1.5 h at 37 °C.
3. Run digested samples on an agarose gel and cut out relevant DNA bands.

4. Gel extract DNA from the agarose gel using standard methods.
5. Ligate vector and insert in a 10 μL reaction in a 1:3 molar ratio using T4-DNA ligase; incubate at RT for 30–45 min.
6. Use the ligation mix to transform competent DH5α cells (100 μL cells +5 μL ligation mix) and plate cells on LB + cam (25 μg/mL) agar plates.
7. Screen colonies for fragment of interest (digest check or colony PCR screen).
8. Sequence and re-transform verified clones into DH5α competent cells and store glycerol stocks at −80 °C.

3.1.2 Cloning into the mCherry Reporter Plasmid

1. PCR-amplify the mRNA leader sequence for the gene of interest or make a deliberate primer-dimer for the leader sequence with (upstream) BsiW I and (downstream) Nhe I as the flanking restriction sites (*see* **Note 7**).
2. Digest 1 μg pBAD42 Buk-mCherry RP and the PCR product with BsiW I-HF and Nhe I-HF for 1.5 h at 37 °C.
3. Run digested samples on an agarose gel and cut out relevant DNA bands
4. Gel extract DNA from the agarose gel using standard methods.
5. Ligate vector and insert in a 10 μL reaction in a 1:3 molar ratio using T4-DNA ligase; incubate at RT for 30–45 min.
6. Use the ligation mix to transform competent DH5α (100 μL cells +5 μL ligation mix) and plate cells on LB + spec (175 μg/mL) agar plates.
7. Screen colonies for fragment of interest (digest check or colony PCR screen).
8. Sequence and re-transform verified clones into DH5α competent cells and store glycerol stocks at −80 °C.

3.1.3 Autofluorescence Control Strains

1. pACYC184 Δ*tetA* and pBAD42 can be used as autofluorescence (background) control plasmids for the assay. Together with an sRNA control plasmid (below) these should be used to transform a mock reporter strain and grown in parallel in triplicate with each screen (*see* **Note 8**).

3.2 Cloning sRNA Variants into the Modular sRNA Expression Vector

3.2.1 Construction of sRNA Library

1. To retarget DsrA or other sRNAs to the TIR of interest, an sRNA library can be constructed by "tiling" antisense sequences to pair with the TIR of the target mRNA [9] (Fig. 2) (*see* **Note 9**).
2. For DsrA, antisense sequence tiling can be done in 2–3 bp increments.

3. 18 bp antisense sequences that span the target mRNA TIR can then be cloned into DsrA using the modular cloning sites in the pSDS801 *dsrA (buk'1.1 hydA'2.4.1)* vector. It is important that the fingerloop antisense motif is conserved during this process (Fig. 2b, c) (*see* **Notes 10** and **11**).

3.2.2 Cloning Retargeted sRNA Variants (Annealed Primers or Primer Dimers)

1. Annealed primers or deliberate primer dimer PCR reactions can be used to exchange the antisense regions within our DsrA variant, creating retargeted DsrA variants against the mRNA(s) of interest. Because the sRNA vector is modular, other sRNAs can be cloned in place of our DsrA variant and tested/retargeted in a similar fashion.
2. Antisense sequences/fingerloop motifs can be integrated into stem loop 1 of our DsrA variant by cloning annealed primers between the Aat II and Apo I restriction sites (Fig. 2b).
3. Antisense sequences/finger loop motifs can be integrated into stem loop 2 of our DsrA variant by cloning annealed primers between the Apo I and Bsu36 I restriction sites (Fig. 2b).
4. DsrA can be completely replaced with a different sRNA using the Aat II and Mfe I restriction sites (Fig. 2a, b). A new sRNA can be constructed using annealed primers, via deliberate primer dimer PCR, or can be PCR-amplified from the host chromosome. If desired, the gene encoding Hfq protein can be removed from the sRNA vector using the Hind III and Nde I restriction sites, with blunting of ends with T4 DNA polymerase (Fig. 1c) prior to ligation. (Such a plasmid encoding *lacI* and *dsrA*, but not *hfq*, is available upon request.)
5. Digest 1 μg pSDS801 *dsrA (buk'1.1 hydA'2.4.1)* with the relevant enzymes for the region of the sRNA you would like to modify.
   - To replace stem loop 1: digest with AatII and ApoI.
   - To replace stem loop 2: digest with ApoI-HF and Bsu36I.
   - To fully replace the sRNA scaffold: digest with AatII and MfeI-HF.
6. Run digested samples on an agarose gel and cut out relevant DNA bands.
7. Gel-extract DNA from the agarose gel.
8. Ligate vector with insert in a 10 μL reaction in a 1:10 molar ratio using T4 DNA ligase; incubate at RT for 30–45 min (*see* **Note** 7).
9. Use the ligation mix to transform competent DH5α (100 μL cells +5 μL ligation mix) and plate cells on LB + amp (200 μg/mL) agar plates.
10. Screen colonies for fragment of interest (digest check or colony PCR screen).

11. Sequence plasmids. To archive, re-transform verified clones into DH5α competent cells and store glycerol stocks at −80 °C. Retain sequenced DNA for transformations of assay strain.

3.2.3 Autofluorescence Control

1. pSDS801(Δ*dsrA*) can be used as the autofluorescence control plasmid for the assay.

3.3 High-Throughput In Vivo Assay

1. For DsrA variants, it is essential that fresh transformants are used for the assay (*see* **Note 12**).

3.3.1 Pre-assay Setup

2. The CM1000 strain with the two reporter plasmids is re-streaked from the −80 °C stock onto an LB (+cam^{25}, +spec^{175}) agar plate.
3. After 16 h, 5–10 colonies from the plate are resuspended in 400 μL of ice-cold, sterile 0.1 M $CaCl_2$ and placed on ice.
4. After 30 min, 100 ng of the sRNA plasmid is added to 50 μL of ice-cold cell suspension and the mixture is kept on ice.
5. After 10 min, the cells are heat shocked at 42 °C for 30–60 s.
6. Cells are placed on ice for 1 min and 150 μL of SOC medium is added to the cells.
7. Place the cells in a 37 °C water bath to recover for 1 h.
8. Cells can be plated on LB +(amp^{200}, cam^{25}, spec^{175}) triple antibiotic plates.
9. After 16 h, single colonies from the plate are used to inoculate 2 mL cultures in LB (+amp^{200}, cam^{25}, spec^{75}) liquid medium.
10. Over-day 3-plasmid assay cultures are grown in a 37 °C water bath for 12 h, shaking at 220 rpm. This can be done with test tube or microtiter plate cultures.

*3.3.2 Cell Growth and Fluorescence Assay (see **Note 13**)*

1. Add 150 μL of M9 + GMT liquid medium (+antibiotics, +2% (w/v) ara, +20 ng/mL aTet) to half of the wells of a sterile Corning 3603 microtiter plate.
2. Add 150 μL of M9 + GMT medium (+ antibiotics, +2% (w/v) ara, +20 ng/mL aTet, + 1 mM IPTG) to the other half of the wells of the plate.
3. For each over-day culture, dilute 1:100 (v/v) into one well from **step 1** and one well from **step 2**. IPTG will induce transcription of the sRNA.
4. For varying reporter gene or sRNA expression, inducer concentrations can be scaled from 0 to 2% ara, 0–20 ng/mL aTet, and 0–1 mM IPTG, respectively.
5. After the plate is set up and the wells are inoculated, overlay each well with 50 μL mineral oil to prevent evaporation during the assay. Note that the mineral oil should be added very carefully down one side of the plate wells to avoid introducing bubbles, which create artifacts in the plate reader data for a given well.

6. Controls for the assay include (1) an sRNA-deletion plasmid and (2) wild-type sRNA vector in the reporter strain (CM1000 + 1–2 relevant fluorescence reporter plasmids). An autofluorescence control strain is constructed by transforming the 3 empty vectors (pSDS1002Δ*dsrA*, pBAD42, and pACYC184 Δ*tetA*) into CM1000 to calculate autofluorescence from the cells under the assay growth conditions. During data analysis, this cellular autofluorescence is subtracted from the fluorescence signal observed in each well to yield the final fluorescence reading for each culture. (*see* **Note 8**).

3.3.3 Plate Reader Configuration for the Assay

1. Before the 96-well plate is set up, the plate reader is warmed up to 37 °C.
2. After plate setup is complete, the plate is placed inside the plate reader in the kinetic readout mode, with continuous shaking of the plate.
3. The plate reader is configured to read OD_{600}, $GFPuv_{mut6}$ fluorescence, and mCherry fluorescence at the start of the assay (t_0) and then every 30 min until the end of the run (12–16 h at 37 °C) (*see* **Notes 14** and **15**).

3.4 Data Analysis

Many plate readers come with built-in data analysis software. The software usually performs several automatic functions such as data transposition (per well for time series) and automated analysis tools to detect fluorescence significantly different from a chosen control. Many labs use these features to quickly identify interesting strains in high-throughput experiments. We developed an Excel sheet (*see* EMS Files) that takes the raw format plate reader data and performs many of these functions, with further analyses to facilitate data interpretation.

3.4.1 Exporting Data from the Biotek Plate Reader

1. To export data from the plate reader, select the "File export" option and export in the following order: $OD_{600nm}(t_0)$, $GFPuv_{mut6}(t_0)$, mCherry(t_0), followed by kinetic time points (1–25) for OD_{600nm}, $GFPuv_{mut6}$ fluorescence, and mCherry fluorescence.
2. This method exports data in a notepad format for the assay (31 total kinetic reads plus t_0.)
3. The data can then be copied and pasted into sheet 2 of the Excel data processing template for further analysis. The sheet assumes the same total number of readings, so user variations will need to take this into account when using the Excel sheet.

3.4.2 Excel Template

1. The Excel template (*see* EMS files) is designed for rapid data analysis at the end of an in vivo assay. It is divided into three parts, Sheet 1 (Experiment Summary), Sheet 2 (Raw Data), and Sheet 3 (Processed Data).

2. Sheet 1 records the experiment summary and has information about the date of the assay, aim of the experiment, codes/shorthand used for strain designations, plasmids in each strain, growth media used, inducer concentrations, and the layout of the 96-well plate.
3. Sheet 2 records the raw data from the plate reader assay. The data can be copied and pasted here from **step 3** of Subheading 3.4.1 above (starting in cell A1).
4. Sheet 3 has tables corresponding to every well in the plate from A1–A12 to H1–H12. Sheet 3 automatically processes the data from Sheet 2 and gives information about $GFPuv_{mut6}$ and mCherry fluorescence at mid-log by automatically plotting fluorescence versus OD_{600} and then interpolating the fluorescence value at an OD_{600} value of 0.5.
5. If the linear fit R^2 is <0.9, the spreadsheet highlights the particular column in orange for heuristic analysis of those data elements.
6. The autofluorescence signal can then be subtracted from each strain's fluorescence at mid-log to yield final values, and can be graphed in desired format (*see* **Note 16**).

4 Notes

1. To study variants of wild-type DsrA sRNA we worked in a markerless *dsrA* knockout strain [14]. It may be important to the study of any particular sRNA to work in a strain that does not produce that sRNA from the chromosome. For the three-plasmid system we used reporter plasmids with three distinct origins of replication (p15A, pSC101, pMB1), but also included a p15A *ori* in the modular pBAD-mCherry reporter plasmid. The effect of this *ori* is to increase the copy number of the pBAD reporter vector and decrease the copy number of the pACYC reporter vector in multi-reporter strains. Experiments performed with variants of pBAD that contain a rho-independent synthetic terminator (http://parts.igem.org/Part:BBa_B1006) [16] in place of this extra *ori* sequence show similar results (S. Stimple, A. Lahiry and R. Lease, unpublished). Different plasmids with equivalent origins can be selected and maintained under pressure of antibiotic resistance [17]. Copies of both plasmid versions will be made available upon request.
2. For straightforward preparation of cells with competency in the range of 10^7–10^8 transformants per μg DNA, we recommend preparing cells by the Inoue method [18, 19].
3. In our experience, many DsrA variants have exhibited a slow-growth phenotype in M9 + GM. This problem can be partially overcome by supplementing the media with 1% (w/v) tryptone.

4. Inducers and antibiotics must be added to the assay media (Subheadings 2.2.2 and 3.3.2). For antibiotics, aTet, and IPTG, the volumes that must be added are negligible. However, 10% of the final media volume will come from the 20% arabinose stocks for a final concentration of 2%. Consequently, to prevent dilution of the media components, it is suggested that the arabinose stock solution is added to the media ingredients before adjusting to the final volume. An *E. coli* strain with arabinose degradation genes knocked out can also be utilized to decrease the amount of arabinose required in the assay media, while maintaining a reasonably linear dosage-response induction profile [20].
5. We have specified a plate reader that uses filters and dichroic mirrors to achieve fluorescent readout, but other plate reader types may be suitable. The plate reader needs to be able to do the following: Incubate cells at desired temperature (typically 37 °C), shake cells continuously, read OD_{600}, and read in fluorescence mode at two specified wavelengths (with separate excitation frequencies).
6. Both reporter plasmids contain modular, in-frame 5′ mRNA leaders/TIRs fused to their respective reporter genes (*hydA*:*gfp-uv*$_{mut6}$ and *buk*::*mCherry*). The sRNA plasmid contains a *dsrA* gene variant that successfully represses translation of HydA-GFPuv$_{mut6}$ and Buk-mCherry, and can act as a positive control for these assays. TIRs/mRNA leaders from the genes of interest can be exchanged via cloning (described in Subheadings 3.1.1 and 3.1.2). For engineering sRNAs that target these TIRs, the sRNA sequence can be altered to incorporate new antisense regions, or a different sRNA can be used (Subheadings 3.2.1 and 3.2.2).
7. For cloning short patches of sequence, we favor two simple methods. In the first, for fragments up to about 100 base pairs, we synthesize a pair of primers which, when annealed in 1× ligase buffer, create a double-stranded DNA fragment with staggered, compatible ends for ligation (Fig. 2b). These fragments can be ligated directly into digested vectors without extra kinase treatment, as the vectors are phosphorylated from digestion with restriction enzymes. Force-cloning of annealed oligo pairs is extremely efficient if the proper molar rations of fragment to insert are used. We recommend using a large molar excess (~20–50-fold) of insert to vector in the ligation reactions. The second method, forced primer-dimer construction, is used for fragments between 100 and 200 base pairs. Here the desired sequence to be cloned is broken roughly into halves with an overlap region in the middle, and synthesized as two primers, such that the upper and lower strands of the desired sequence terminate at their 3′-ends in a large overlap with the partner complementary sequence (20–25 bp is usually sufficient). No template DNA is used; rather the two primers each

use the other primer as template. The dsDNA fragments resulting from deliberate primer-dimer design must be digested with appropriate restriction enzymes. Accordingly, it is useful to leave 3–5 bp extra "junk" DNA at the 5′-ends of the primers to facilitate digestion. See the NEB catalog document "Cleavage Close to the end of DNA fragments" (https://www.neb.com/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments) for guidance on particular enzymes.

8. The autofluorescence strain should be CM1000 transformed with the three empty vectors (pBAD42, pACYC *ΔtetA*, and pSDS1002 *ΔdsrA*). The negative controls should include the following strains: CM1000/pBAD42-TIR_1-mCherry//pACYC-TIR_2-*gfp-uv*$_{mut6}$///pSDS801 Δ*dsrA hfq*$^+$; CM1000/pBAD42-TIR_1-mCherry//pACYC-TIR_2-*gfp-uv*$_{mut6}$///pSDS801 *dsrA*$^+$ *hfq*$^+$; where the wild-type/parent sRNA (here, DsrA) is not expected to bind/regulate the novel TIRs of interest.
9. To yield functional regulatory sRNAs against the TIR of interest, a very effective method is to create a series of antisense sequence "tiles" against the region spanning from the Shine-Dalgarno sequence through the first 6 amino acids of the target gene coding sequence. Each construct must be tested separately and compared to the others to determine the optimum active sRNA variant.
10. The fingerloop motif can be used as a modular antisense element. Structured antisense regions can be exchanged in modular fashion [3], and likely stabilize the sRNA transcript against ssRNA-cleaving nucleases, thus improving sRNA stability relative to unstructured, single-stranded antisense RNA sequences. As a design guide for the size of the stem and loop regions of the fingerloop, the complex between the *hydA*'2.4.1 antisense sequence and the *hydA* mRNA leader has a predicted free energy of ~31 kcal/mol, as determined from NUPACK. Typically we see ~2–12-fold target mRNA repression with fingerloop stabilities in the range of −4 to −14 kcal/mol for this antisense sequence length. As mentioned above, G•U pairs and mismatches may be engineered into the opposite strand of the stem to modulate free energy parameters, and loop sizes can also be varied. If desired, unstructured sRNA antisense sequences could also be used with this system.
11. Our retargeting efforts have focused on antisense regions with a length of 18 nt. The length of the antisense sequence can be shortened to decrease the dynamic range of the translational repression, or lengthened to improve it. Na et al. [9] used MicC sRNA as a scaffold to incorporate tiled antisense sequence against their target genes using a 24 nt antisense sequence basis length, to improve dynamic range and as an effort to

minimize the number of off-target sequences that could match the sRNA antisense region.

12. CM1000 can be transformed with one of the reporter plasmids by preparing "Quickie" competent cells: Resuspend ~15 colonies from a plate still growing at 37 °C into 100 μL 0.1 M $CaCl_2$ in an eppendorf tube on ice. After 30 min, ~100 ng of the plasmid is added, the cell-DNA mixture is incubated on ice for >10 min, and heat-shocked at 42 °C for 30–60 s. SOC media is added, the cells are incubated at 37 °C for 1 h to recover, and the cells are plated. An individual colony is re-streaked, and if desired the process can be repeated with a second reporter plasmid. After re-streaking, CM1000 will now contain both reporter plasmids, and can be stored as a glycerol storage culture at −80 °C. Restreaks of this two-reporter strain can then be used to make Quickie competent cells for transformation with desired sRNA clones, scaling the 0.1 M $CaCl_2$ step to ~50 μL per transformant. (This "Quickie" transformation protocol was authored by Dr. Jill Salvo many years ago in Dr. Marlene Belfort's lab (Wadsworth/SUNY, Albany NY) and may be unpublished.)
13. For each strain used in the assay, the overday culture must be used to inoculate separate wells of the 96-well plate (one containing reporter inducer molecules but lacking IPTG, and one containing reporter inducers with IPTG added to 1 mM for sRNA induction). This control will allow direct comparison of the fluorescence signals when assessing whether the sRNA in question represses (or activates) its target TIR.
14. Some plate readers accurately read OD_{600}, but for others a "correction factor" must be used to yield accurate OD_{600} readings. For instance, in our plate reader a culture growing in 150 μL of media (overlayed with 50 μL mineral oil) outputs values that must be multiplied by 3.6 to reflect the true OD_{600} value of the culture (detected by an external spectrophotometer). This number should be determined for a given plate reader in advance of experiments, in accordance with the manufacturer's manual. The value should not change significantly across the plate.
15. During plate setup, the Excel spreadsheet assumes that well H11 is not inoculated, serving as a background control well for liquid growth media. In well H11, simply add 150 μL of media and overlay with mineral oil; this well will show the background contributions of the media to mCherry and $GFPuv_{mut6}$ fluorescence and OD_{600nm}. The mCherry and $GFPuv_{mut6}$ fluorescence of this well should be similar to or below that of the autofluorescence strain during growth at mid-exponential phase.
16. We recommend plotting OD_{600} (X-axis) versus fluorescence signals (Y-axis or double Y-axis for two fluorescent signals; Fig. 3). This method permits comparison of strains at their

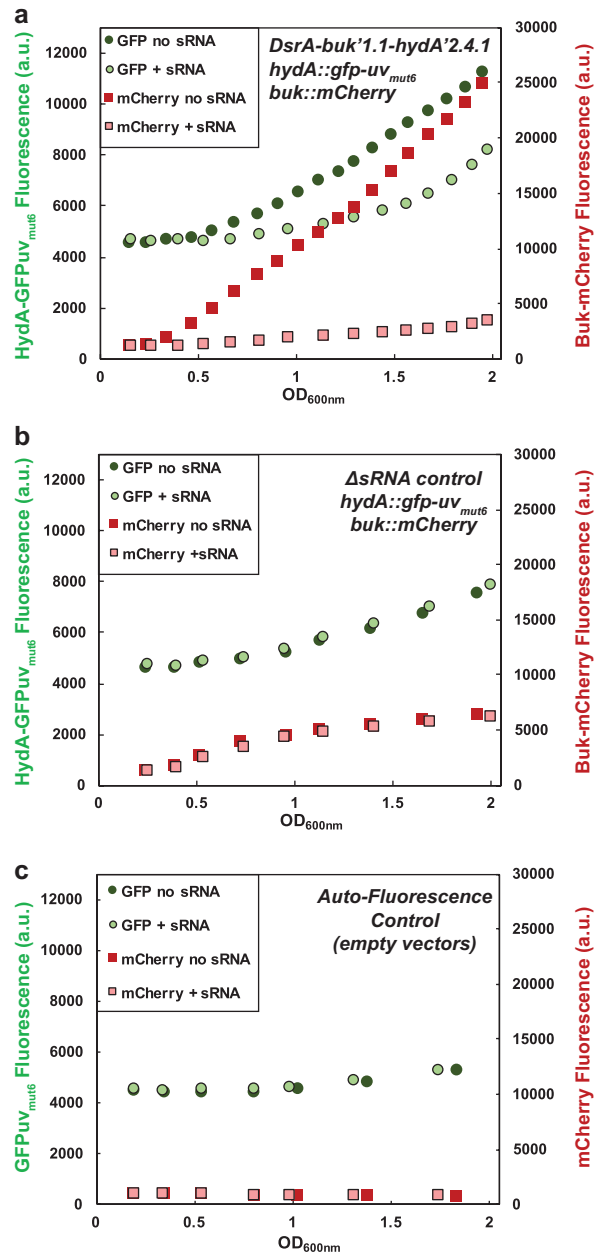


Fig. 3 Representative data from the assay. Data are plotted as fluorescence (mCherry, GFPuv) arbitrary units versus cell density (as OD_{600}) for comparison of sRNA-uninduced and sRNA-induced strains at equivalent cell density. (**a**) Plot of the activity of a DsrA-*buk*'1.1-*hydA*'2.4.1 retargeted sRNA variant at its two reporter genes (*buk::mCherry*, *hydA::gfp-uv*$_{mut6}$) across a range of cell densities. Solid and open green circles represent HydA::GFPuv whereas solid and open red squares represent Buk-mCherry expression in the absence and presence, respectively, of IPTG which induces the sRNA variant. The magnitude of the repression by the DsrA sRNA variants is larger at higher OD_{600} but may vary for different sRNA:mRNA pairs. (**b**) Control experiments using the sRNA plasmid deleted for the sRNA gene. (**c**) Autofluorescence control for cells harboring empty vectors (pACYC184Δ*tet*, pBAD42, pBR322Δ*tet*). The Excel spreadsheet also contains a function that subtracts autofluorescence values and interpolates the fluorescence values at OD_{600} = 0.5, but can be altered to accommodate different quantification needs (*see* **Note 16**)

equivalent OD_{600}. The OD_{600} is thus representative of time, although it underrepresents differences in growth rates between strains. The plate reader typically plots OD_{600} versus time during its run, and the Excel sheet represents the data in this format for determination of fluorescence at OD_{600}.

Acknowledgments

We would like to thank Dr. Susan Gottsman for the generous gift of *E. coli* strain CM1000.

References

1. Storz G, Vogel J, Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 43:880–891
2. Cho SH, Haning K, Contreras LM (2015) Strain engineering via regulatory noncoding RNAs: not a one-blueprint-fits-all. Curr Opin Chem Eng 10:25–34
3. Lahiry A, Stimple SD, Wood DW, Lease RA (2017) Retargeting a dual-acting sRNA for multiple mRNA transcript regulation. ACS Synth Biol 6:648–658
4. Mandin P, Gottesman S (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. EMBO J 29:3094–3107
5. Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ (2004) Engineered riboregulators enable post-transcriptional control of gene expression. Nat Biotechnol 22:841–847
6. Hoynes-O'Connor A, Moon TS (2016) Development of design rules for reliable antisense RNA behavior in *E. coli*. ACS Synth Biol 5(12):1441–1454
7. Hussein R, Lim HN (2012) Direct comparison of small RNA and transcription factor signaling. Nucleic Acids Res 40:7269–7279
8. Sakai Y, Abe K, Nakashima S, Yoshida W, Ferri S, Sode K et al (2014) Improving the gene-regulation ability of small RNAs by scaffold engineering in *Escherichia coli*. ACS Synth Biol 3:152–162
9. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY (2013) Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. Nat Biotechnol 31:170–174
10. Urban JH, Vogel J (2007) Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. Nucleic Acids Res 35:1018–1037
11. Levine E, Zhang Z, Kuhlman T, Hwa T (2007) Quantitative characteristics of gene regulation by small RNA. PLoS Biol 5:e229
12. Kushwaha M, Rostain W, Prakash S, Duncan JN, Jaramillo A (2016) Using RNA as molecular code for programming cellular function. ACS Synth Biol 5:795–809
13. Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR et al (2011) NUPACK: analysis and design of nucleic acid systems. J Comput Chem 32:170–173
14. McCullen CA, Benhammou JN, Majdalani N, Gottesman S (2010) Mechanism of positive regulation by DsrA and RprA small noncoding RNAs: pairing increases translation and protects *rpoS* mRNA from degradation. J Bacteriol 192:5559–5571
15. Fong BA, Wood DW (2010) Expression and purification of ELP-intein-tagged target proteins in high cell density *E. coli* fermentation. Microb Cell Factories 9:77
16. Cambray G, Guimaraes JC, Mutalik VK, Lam C, Mai QA, Thimmaiah T et al (2013) Measurement and modeling of intrinsic transcription terminators. Nucleic Acids Res 41:5139–5148
17. Chang AC, Cohen SN (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J Bacteriol 134:1141–1156
18. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
19. Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. Gene 96:23–28

20. Afroz T, Biliouris K, Kaznessis Y, Beisel CL (2014) Bacterial sugar utilization gives rise to distinct single-cell behaviours. Mol Microbiol 93:1093–1103
21. Lease RA, Belfort M (2000) A *trans*-acting RNA as a control switch in *Escherichia coli:* DsrA modulates function by forming alternative structures. Proc Natl Acad Sci U S A 97:9919–9924
22. Lease RA, Woodson SA (2004) Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. J Mol Biol 344:1211–1223
23. de Almeida Ribeiro E, Beich-Frandsen M, Konarev PV, Shang W, Večerek B, Kontaxis G et al (2012) Structural flexibility of RNA as molecular basis for Hfq chaperone function. Nucleic Acids Res 40:8072–8084
24. Lalaouna D, Morissette A, Carrier M-C, Massé E (2015) DsrA regulatory RNA represses both *hns* and *rbsD* mRNAs through distinct mechanisms in *Escherichia coli*. Mol Microbiol 98:357–369

INDEX

Véronique Arluison and Claudio Valverde (eds.), *Bacterial Regulatory RNA: Methods and Protocols*, Methods in Molecular Biology, vol. 1737, https://doi.org/10.1007/978-1-4939-7634-8, © Springer Science+Business Media, LLC 2018

T

W