Association for Women in Mathematics Series

Anita T. Layton
Laura A. Miller *Editors*

Women in Mathematical Biology

Research Collaboration Workshop, NIMBioS, Knoxville, June 2015

AWM
ASSOCIATION FOR WOMEN IN MATHEMATICS

Springer

Association for Women in Mathematics Series

Volume 8

Series Editor

Kristin Lauter
Microsoft Research
Redmond, Washington, USA

More information about this series at http://www.springer.com/series/13764

Association for Women in Mathematics Series

Focusing on the groundbreaking work of women in mathematics past, present, and future, Springer's Association for Women in Mathematics Series presents the latest research and proceedings of conferences worldwide organized by the Association for Women in Mathematics (AWM). All works are peer-reviewed to meet the highest standards of scientific literature, while presenting topics at the cutting edge of pure and applied mathematics. Since its inception in 1971, The Association for Women in Mathematics has been a non-profit organization designed to help encourage women and girls to study and pursue active careers in mathematics and the mathematical sciences and to promote equal opportunity and equal treatment of women and girls in the mathematical sciences. Currently, the organization represents more than 3000 members and 200 institutions constituting a broad spectrum of the mathematical community, in the United States and around the world.

Anita T. Layton • Laura A. Miller
Editors

Women in Mathematical Biology

Research Collaboration Workshop, NIMBioS, Knoxville, June 2015

Springer

Editors
Anita T. Layton
Department of Mathematics
Duke University
Durham, North Carolina, USA

Laura A. Miller
Departments of Mathematics and Biology
University of North Carolina
Chapel Hill, North Carolina, USA

ISSN 2364-5733 ISSN 2364-5741 (electronic)
Association for Women in Mathematics Series
ISBN 978-3-319-60302-5 ISBN 978-3-319-60304-9 (eBook)
DOI 10.1007/978-3-319-60304-9

Library of Congress Control Number: 2017949389

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

On June 22, 2015, four research teams consisting of 24 women in mathematics and biology converged at the National Institute for Mathematics and Biological Synthesis (NIMBioS) in Knoxville, Tennessee, to participate in a Research Collaboration Workshop for Women in Mathematical Biology. The goal of the workshop was to kick-start multidisciplinary research projects and form lifetime connections to other women in STEM fields. This workshop was inspired by the highly successful 2013 Institute for Mathematics and its Applications (IMA) Special Workshop: WhAM! A Research Collaboration Workshop for Women in Applied Mathematics: Dynamical Systems with Applications to Biology and Medicine.

NIMBioS was an excellent venue to host this multidisciplinary workshop. Consistent with the institute's mission to trigger research in quantitative modeling and analysis in the life sciences, NIMBioS hosted four research teams consisting of women who are trained and working in both biology and mathematics departments. Each group consisted of a senior faculty and a junior faculty from the fields of mathematics, engineering, and the life sciences. These teams each mentored four junior women who ranged from graduate students to assistant professors.

This special volume contains research articles contributed by the four research teams. The topics include aerodynamics of spider ballooning; sleep, circadian rhythms, and pain; blood flow regulation in the kidney; and the effects of antimicrobial therapy on gut microbiota and Clostridium difficile. Each topic includes a review article and at least one research article. In addition, several papers are included in this volume on topics inspired by the workshop. This work includes contributions from junior researchers at the undergraduate, graduate, and postdoctoral levels on topics ranging from models of animal movement to the flow of blood cells in the embryonic heart.

We thank NIMBioS for its support, without which the workshop would not have become such a resounding success. And it is indeed our hope this workshop and this special volume will spark new ideas and new collaborations among female mathematicians.

Durham, North Carolina, USA Anita T. Layton
Chapel Hill, North Carolina, USA Laura A. Miller

Contents

The Modulation of Pain by Circadian and Sleep-Dependent Processes: A Review of the Experimental Evidence 1
Megan Hastings Hagenauer, Jennifer A. Crodelle, Sofia H. Piltz, Natalia Toporikova, Paige Ferguson, and Victoria Booth

Investigating Circadian Rhythmicity in Pain Sensitivity Using a Neural Circuit Model for Spinal Cord Processing of Pain 23
Jennifer A. Crodelle, Sofia H. Piltz, Victoria Booth, and Megan Hastings Hagenauer

A Two-Process Model for Circadian and Sleep-Dependent Modulation of Pain Sensitivity 49
Natalia Toporikova, Megan Hastings Hagenauer, Paige Ferguson, and Victoria Booth

Introduction to Mathematical Modeling of Blood Flow Control in the Kidney 63
Anita T. Layton and Aurélie Edwards

Modeling Autoregulation of the Afferent Arteriole of the Rat Kidney 75
Maria-Veronica Ciocanel, Tracy L. Stepien, Aurélie Edwards, and Anita T. Layton

Modeling Blood Flow and Oxygenation in a Diabetic Rat Kidney 101
Ioannis Sgouralis and Anita T. Layton

Tracking the Distribution of a Solute Bolus in the Rat Kidney 115
Anita T. Layton

Mathematical Modeling of the Effects of Nutrient Competition and Bile Acid Metabolism by the Gut Microbiota on Colonization Resistance Against *Clostridium difficile* 137
Arietta Fleming-Davies, Sara Jabbari, Suzanne L. Robertson, Tri Sri Noor Asih, Cristina Lanzas, Suzanne Lenhart, and Casey M. Theriot

Revisiting the Physics of Spider Ballooning 163
Kimberly S. Sheldon, Longhua Zhao, Angela Chuang, Iordanka N. Panayotova, Laura A. Miller, and Lydia Bourouiba

Flying Spiders: Simulating and Modeling the Dynamics of Ballooning 179
Longhua Zhao, Iordanka N. Panayotova, Angela Chuang, Kimberly S. Sheldon, Lydia Bourouiba, and Laura A. Miller

On the Dynamic Suction Pumping of Blood Cells in Tubular Hearts 211
Nicholas A. Battista, Andrea N. Lane, and Laura A. Miller

Undergraduate Research Highlight: Modeling Movement Behavior Among Interacting Species 233
Anne Talkington

Index 251

The Modulation of Pain by Circadian and Sleep-Dependent Processes: A Review of the Experimental Evidence

Megan Hastings Hagenauer, Jennifer A. Crodelle, Sofia H. Piltz, Natalia Toporikova, Paige Ferguson, and Victoria Booth

Abstract This proceedings paper is the first in a series of three papers developing mathematical models for the complex relationship between pain and the sleep–wake cycle. Here, we briefly review what is known about the relationship between pain and the sleep–wake cycle in humans and laboratory rodents in an effort to identify constraints for the models. While it is well accepted that sleep behavior is regulated by a daily (circadian) timekeeping system and homeostatic sleep drive, the joint modulation of these two primary biological processes on pain sensitivity has not been considered. Under experimental conditions, pain sensitivity varies across the 24 h day, with highest sensitivity occurring during the evening in humans. Pain sensitivity is also modulated by sleep behavior, with pain sensitivity increasing in response to the build-up of homeostatic sleep pressure following

M.H. Hagenauer
Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: hagenaue@umich.edu

J.A. Crodelle
Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
e-mail: kilej@rpi.edu

S.H. Piltz
Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kongens Lyngby, Denmark
e-mail: shpi@dtu.dk

N. Toporikova
Biology Department, Washington and Lee University, Lexington, VA 24450, USA
e-mail: toporikovan@wlu.edu

P. Ferguson
Department of Biological Sciences, University of Alabama, Tuscaloosa, AL 35487, USA
e-mail: pfferguson@ua.edu

V. Booth (✉)
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA

Department of Anesthesiology, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: vbooth@umich.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_1

sleep deprivation or sleep disruption. To explore the interaction between these two biological processes using modeling, we first compare the magnitude of their effects across a variety of experimental pain studies in humans. To do this comparison, we normalize the results from experimental pain studies relative to the range of physiologically meaningful stimulation levels. Following this normalization, we find that the estimated impact of the daily rhythm and of sleep deprivation on experimental pain measurements is surprisingly consistent across different pain modalities. We also review evidence documenting the impact of circadian rhythms and sleep deprivation on the neural circuitry in the spinal cord underlying pain sensation. The characterization of sleep-dependent and circadian influences on pain sensitivity in this review paper is used to develop and constrain the mathematical models introduced in the two companion articles.

MSC codes: 92B25, 92C20

1 Introduction: A Vicious Cycle

The experience of pain has a complex relationship with the sleep–wake cycle. Pain serves two important purposes: to motivate individuals to escape and avoid physical insult and to aid in healing by promoting the protection and immobilization of injured body parts. This first purpose necessitates rapid response and arousal, two processes that are suppressed by sleep, whereas the second purpose is closely tied to the concept of rest. Thus pain makes us tired (promotes the homeostatic drive to sleep), and increased sensitivity to pain during the night is coordinated with our daily circadian rhythm to promote immobilization and healing during the rest period [8]. However, the presence of pain is arousing and can inhibit our ability to initiate and maintain sleep, especially the deeper recuperative stages of sleep [35]. When sleep is disrupted or limited, the perception of pain further intensifies, healing is delayed, and pathological processes promoting the development of chronic pain can proceed unchecked [17]. Within clinical settings, this progression of events can create a vicious cycle of inadequate pain management [35], which is further complicated by similarly strong interdependencies between the sleep–wake cycle and the effectiveness of most forms of analgesia [8, 17, 35].

The development and analysis of mathematical models of this vicious cycle can lead to better understanding of the interactions between sleep and pain, which could improve pain management. In this article, we review the experimental and clinical evidence documenting the modulation of pain by sleep and circadian processes in humans and animals and introduce a novel analysis of this data that is used to justify and constrain the mathematical models introduced in the companion articles.

2 What Is Pain?

"Pain is an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" according to the International Association for the Study of Pain [40]. Pain can be caused by different types of actual or potential tissue damage, including adverse temperature conditions (heat, cold), intense mechanical stimulation or pressure, electric shock, constricted vasculature, or chemical irritation, as well as processes generated within the body, such as inflammation and pathological nerve damage (neuropathy). Pain can be derived experimentally or from natural conditions, and can occur on a variety of time scales. Experimental studies of "acute" pain sensitivity typically induce brief ("phasic"), localized, superficial pain to peripheral tissues. Such brief stimulation actually consists of two sensations: a fast, sharp pain and a slower, dull pain. Occasionally, experimental studies will induce longer duration ("tonic") acute pain that can last for hours [36]. Within clinical settings, chronic pain conditions can last for months or years.

As there are different types of pain that can be felt, there are different ways in which the body receives and processes pain signals. Sensory neurons (afferent neurons) in the peripheral nervous system sense stimuli and send that information to the spinal cord for processing. These neurons and their nerve fibers are specialized for detecting innocuous or noxious stimuli. Non-painful touch sensations are transmitted by Aβ afferent fibers while there are two major classes of nociceptive (pain-receptive) afferent fibers: Aδ and C. Medium diameter Aδ fibers mediate localized, sharp, fast pain sensations, while small diameter C fibers mediate the more diffuse and duller slow pain sensations [16]. The "fast pain" Aδ fibers are wrapped in a fatty sheath called myelin that allows for rapid transmission of signals, at speed of 4–30 m/s. This is also true for the Aβ fibers. In contrast, the "slow pain" C fibers are not myelinated and, due to their small diameter, transmit signals at speed of less than 2 m/s [36].

Different types of nerve fibers report to different areas in the spinal cord. In general, sensory neurons have their cell bodies in the dorsal root ganglia, a cluster of nerve cell bodies located in the spinal cord. Primary afferent fibers morphologically differ from other nerve fibers in that their axons and dendrites, usually responsible for sending and receiving signals, respectively, have equivalent biochemical make-up and thus these neurons can send and receive signals through both their axons and dendrites [3]. Signals in these afferent fibers are transmitted to the dorsal horn of the spinal cord, an area that is responsible for receiving information from the sensory neurons, processing it, and sending signals up to the brain. The dorsal horn contains many populations of neurons, including excitatory and inhibitory interneurons. One such population of neurons in the dorsal horn, called the Wide Dynamic Range (WDR) neurons, receive direct inputs from the touch and nociceptive afferent fibers as well as inputs from interneuron populations, and constitute the primary output from the dorsal horn to the brain. As such, pain intensity is correlated with the firing rate and the duration of firing of the WDR neurons.

Since pain is both an unpleasant sensory and emotional experience, pain-related input from the spinal cord engages multiple neural circuits in the brain, including the brainstem, thalamus, and cortex. These circuits involve a wide range of neurotransmitter systems, including the well-studied opioid system. Many of these higher-level cognitive and emotional responses to pain exert their influence over pain perception via descending projections to the dorsal horn of the spinal cord. This "top-down" feedback on sensory processing can act to either inhibit or facilitate pain sensation, essentially providing a "gate" for the transmission of nociceptive information to the brain [39]. Thus, there is a tradition of modeling pain processing by focusing exclusively on spinal cord circuitry.

3 The Relationship Between the Sleep Cycle and Pain Sensitivity in Humans

The daily timing of sleep is widely accepted as an interaction between two independent processes: a homeostatic drive to sleep, which builds up over the course of wakefulness in a saturating manner and dissipates during sleep, and a circadian timing system, which rhythmically influences the levels of sleep drive required to initiate and maintain sleep [13]. When exploring the literature documenting the relationship between sleep and pain, we found that the influences of both circadian rhythms and homeostatic sleep drive were rarely measured within the same experiment, despite ample evidence that both processes modulate pain sensitivity [8, 17, 35]. Instead, experimental studies tended to fall into two broad categories. In one variety of experiment, pain perception was measured across the day (24 h) in subjects maintaining their normal sleep schedule. Therefore, the data in these experiments should represent a combination of the influences of time-of-day and a normal modest 16 h build-up of homeostatic sleep drive during waking and 8 h dissipation of homeostatic sleep drive during sleep. In the other variety of experiment, subjects were sleep deprived for 1–3 nights or had their sleep restricted to less than a typical 8 h, and pain perception was recorded at various times. In these experiments, there should be a large build-up of homeostatic sleep drive, the effects of which may be more or less obvious at different times of day due to circadian modulation. We review these two forms of data below and introduce a novel analysis of the data that allows a comparison of results from these two categories of experiments and from studies using different pain modalities. For the sake of simplicity, we focus primarily on data derived from studies using pain modalities of experimentally induced brief (acute/phasic), superficial pain to peripheral tissues.

3.1 There Is a Daily Rhythm in Experimental Pain Sensitivity in Humans

Pain sensitivity follows a daily cycle in many clinical conditions [8], but it is currently unclear how much of that rhythmicity is derived from daily fluctuation in the underlying causes driving the pain (for example, nocturnal release of oxytocin induces contractions during labor) versus rhythmicity in the neural processing of pain. Within the experimental pain literature, rhythmic influences on pain sensation occur regardless of whether pain responses are measured subjectively or objectively [6, 12, 14, 56], suggesting that the rhythmic modulation of pain responses occurs at a basic physiological level. This rhythmic modulation of pain sensitivity increases with pain intensity [14, 21, 32], so that the more intense the pain is overall, the greater the change in the person's sensitivity to the pain across the day. Rhythmic influences on pain sensitivity are detectable in experiments involving a variety of different kinds of painful stimuli, including cold, heat, current, pressure, and ischemia (Tables 1 and 2). These stimuli are found to be most painful during hours of the day when experimental subjects are likely to be tired—late afternoon, evening, and night (Table 1).

To better characterize this rhythm, we constructed a prototypical "daily pain sensitivity" function by drawing data from four high-quality experiments that measured pain sensitivity at multiple time points around the 24 h day using diverse testing procedures:

1. The threshold for nociceptive pain reflex in response to electrical current (n = 5, [6]), an objective measure of pain sensitivity. In this study, measurements were taken from the same subjects every 4 h within a consecutive 24 h laboratory study (beginning at 13:00). The study states that subjects lived in "elementary conditions of social synchronization (08:00–23:00)" and remained in bed during night measurements.
2. The threshold for tooth pain in response to cold (n = 79, [48]), and the threshold for tooth pain in response to electrical stimulation (n = 56, [48]). In this large study, measurements were taken from the same subjects every 3 h across a 24 h day. From the methods, it is unclear whether these measurements were completed consecutively, but in a follow-up study in the same paper using a smaller sample size they replicate their results using measurements taken at 24+ h intervals. During the tests, the subjects maintained their normal living cycles.
3. The threshold for forearm pain in response to heat (n = 39, [50]). In this large study, measurements were taken from the same subjects at 4 time points across a 24 h day (8:00, 13:00, 18:00, 23:00). In women, this procedure was repeated at 3 different points across their menstrual cycle (days 7, 15, and 23). During the experiment, subjects maintained their normal daily routine (sleeping hours 24:00–7:00).

Table 1 Studies measuring daily rhythms in experimental human pain sensitivity

Citation	Detectable Rhythmicity	Site of Stimulus	Type of Stimulus	Around the Clock Sampling?	Peak "Most Painful"	Nadir "Least Painful"	Pain Parameter	Objective Pain Parameter?	Suprathreshold Stimulus?
Pollmann, 1984	Yes	Teeth	Cold	Yes: 8 time points	Night (24:00)	Afternoon (15:00-18:00)	PT	N	N
Pollmann, 1984	Yes	Teeth	Current	Yes: 8 time points	Night (24:00)	Afternoon (15:00-18:00)	PT	N	N
Procacci et al., 1974	Yes	Forearm	Heat	Yes: 4 time points (8, 13, 18, 23)	Evening (19:00-23:00)	Morning (6:30-9:30)	PT	N	N
Chapman and Jones, 1944	Yes	Forehead	Heat	No: 2 day time points (9 vs. 17)	Evening/Afternoon (17:00)	Morning (9:00)	PT, PRT	Y (Wincing)	Y
Aviram et al., 2015	Yes	Hand	Cold	No: 3 day time points ("morning", "afternoon", "evening")	Evening (17:00–19:00)	Morning (9:00-11:00)	PT, PTT, PS	N	Y
Aviram et al., 2015	Yes	Hand	Heat	No: 3 day time points ("morning", "afternoon", "evening")	Evening (17:00–19:00)	Morning (9:00-11:00)	PT, PS	N	Y
Gobel and Cordes, 1990	Yes	Pericranial Musculature	Constriction (Ischemia)	Yes: 6 time points	Night (2:00)	Afternoon (14:00)	PS	N	Y
Bourdalle-Badie et al., 1990	Yes	Sural nerve	Current	Yes: 6 time points	Night/Early Morning (1:00-5:00)	Morning/Afternoon (9:00-17:00)	PRT, PS	Y (NFR)	Y
Koch and Raschka, 2004	Yes	Forearm	Tourniquet (Ischemia)	Yes: 4 time points (6, 12, 18, 24)	Night (24:00)	Morning (6:00)	PS	N	Y
Davis et al., 1978	Yes	Forearm	Current	No: 2 time points (between 8 and 16)	Afternoon (14:00-17:00)	Morning (8:00-12:00)	PT, PS, EP	Y (EP)	Y
Grabfield and Martin, 1912	Yes	Finger	Current	No: hourly day time points (8:30-20:30)	Morning (10:30) or Evening (20:30)	Afternoon (16:30)	DT?	N	N
Rogers and Vilkin, 1978	Yes	Forearm	Current	No: 2 time points ("early morning" and "early evening")	"Early Evening"	"Early Morning"	DT, PT	N	N
Sandrini et al., 1986	Yes	Leg	Current	Yes: 4 time points (6, 12, 18, 24)	Morning (6:00)	Night (20-24)	PRT	Y (NFR)	Unknown
Strian et al., 1989	Maybe	Hand, Foot	Cold	Yes: 7 time points	NA	NA	DT, PT	N	N
Martin et al., 1914	Maybe	Hand	Current	No: night time points (20:30-8:30)	NA	NA	DT?	N	N
Bachmann et al., 2011	No	Foot	Heat	Almost: 7 time points (8 -22)	NA	NA	PS, EP	Y (EP)	Y
Aviram et al., 2015	No	Fingers	Mechanical (Pressure)	No: 3 day time points ("morning", "afternoon", "evening")	NA	NA	PT, PTT, PS	N	Y
Morawetz et al., 1984	No	Forearm	Current	Yes: 13 timepoints	NA	NA	DT, PT, IT, PTT	N	Y

Note that these studies focus on brief (acute, phasic) superficial pain in peripheral tissues. Clock hours are in military time (0:00–24:00)
DT Detection threshold
PT Pain threshold
PRT Pair response threshold
PS Pain sensitivity (intensity ratings)
PTT Pain tolerance threshold
IT Intervention threshold
NFR Nociceptive flexion reflex
EP Somatosensory evoked potential (EEG)

Table 2 Studies measuring daily rhythms in experimental human pain sensitivity that have been reviewed in previous publications [15, 3] but that we were either personally unable to review or found to be unreliable

Less Reliable Sources:	Detectable Rhythmicity	Site of Stimulus	Type of Stimulus	Notes:
Kobal et al., 1995	Yes	Nasal mucosa	Chemosomatosensory (CO2)	This is an abstract
Macht et al., 1916	No	Hand, Tongue, Lip, Nose	Current	Only one subject included in the time of day analysis?
Schumacher et al., 1940	No	Forehead	Heat	The 4 subjects included in the time of day analysis *were the author*s
Hardy et al., 1940	No	Forehead	Heat	Daily rhythms not systematically studied - just commented on.
Hummel et al., 1995	*NA*	Nasal mucosa	Chemosomatosensory (CO2)	Study only discusses relative change in morning and evening?
Stacher et al., 1982	No	Earlobe, Forearm	Current	Did not systematically study time of day, just commented on.
Stacher et al., 1982	No	Earlobe, Forearm	Heat	Did not systematically study time of day, just commented on.
Unable to Review the Citation:				
Strempel, 1977	Yes	Hand	Cold	Unable to obtain paper
Jores et al., 1937	Yes	Teeth	Current	(German)
Frees, 1937	Yes	Teeth	Current	(German)
Godt and Thiele, 1968	Yes	Teeth	Current	(German)
Procacci, 1972	Yes	Forearm	Current	(German)
Bragin and Durinyan, 1983	Yes	Forearm	Current	(Russian?)
Starek, 1957	No	Finger, Arm	Heat	(German)
Pollmann, 1974	No	Teeth	Cold	(German)
Pollmann, 1980	No	Teeth	Cold	(German)
Bochnik, 1958	No	Arm	Current	(German)
Ritter, 1958	No	Forearm	Current	(German)
Baginski, 1961	No	Forearm	Stitch	(German)
Schroder, 1905	No	Teeth	Current	(German)

For each study, we only had access to the summary data presented in the figures. Using these data, we standardized the pain measurements by converting them to percent of mean (or the mesor of the depicted rhythm). For ease of use, we inverted measures of pain threshold to pain sensitivity so that low pain threshold corresponded to high pain sensitivity. Thus, in our function, high measurement values are associated with greater pain. Time was standardized in relation to either scheduled or estimated morning wake time in order to organize the data in a manner more akin to the "zeitgeber time" used in sleep and circadian literature.

Following these transformations, the data collectively formed a tight curve that resembled a sinusoid. To produce a smoothed version of the curve for later use as our model input, we used the loess function in R 3.2.1 (loess{stats}, R Core Team 2014), which is a form of local polynomial regression that resembles a "vertical sliding window that moves across the horizontal scale axis of the scatterplot" [28]. The benefit of using loess() is that it does not assume a functional form for the relationship between X and Y and therefore, to some degree, "allows the data to speak for themselves" [28]. A traditional equation with coefficients is not produced. There is a parameter (alpha, sometimes called span) that controls the degree of smoothing via the width of the sliding window. The larger the alpha value, the smoother the curve. If alpha is too small, overfitting is possible. We used the default (alpha = 0.75). There is also a parameter (lambda) that specifies the degree of the polynomial. We used lambda = 2, meaning that quadratic equations were used, which can better capture "peaks" and "valleys" (local minima/maxima) in the data [28]. Using this unbiased approach, we still found that the output curve strongly resembled a simple sinusoid (R2loess = 0.64, Fig. 1).

We hypothesize that this best-fit curve represents an average daily rhythm in pain sensitivity for humans and that the rhythm is affected by both homeostatic sleep drive and a circadian rhythm in pain sensitivity. The curve has a sharp peak in pain sensitivity occurring close to sleep onset (18 h following wake, or approximately 1am) and then decreases during the night. This is consistent with an effect of homeostatic sleep drive on pain sensitivity, and fits previous demonstrations that mental fatigue can decrease pain threshold (i.e., increase pain sensitivity) by 8–10% [12]. The curve also has a distinct trough in pain sensitivity in the afternoon (following 9 h of wake, or approximately 4 pm). This pattern does not fit what would be expected due to an effect of homeostatic sleep drive and instead suggests the influence of a circadian rhythm.

3.2 Homeostatic Sleep Drive Increases Pain Sensitivity in Humans

Within the clinical literature, there are at least 14 studies demonstrating that the experience and intensity of pain correlate with sleep duration or quality [17]. However, the causal nature of this relationship is best evaluated within controlled

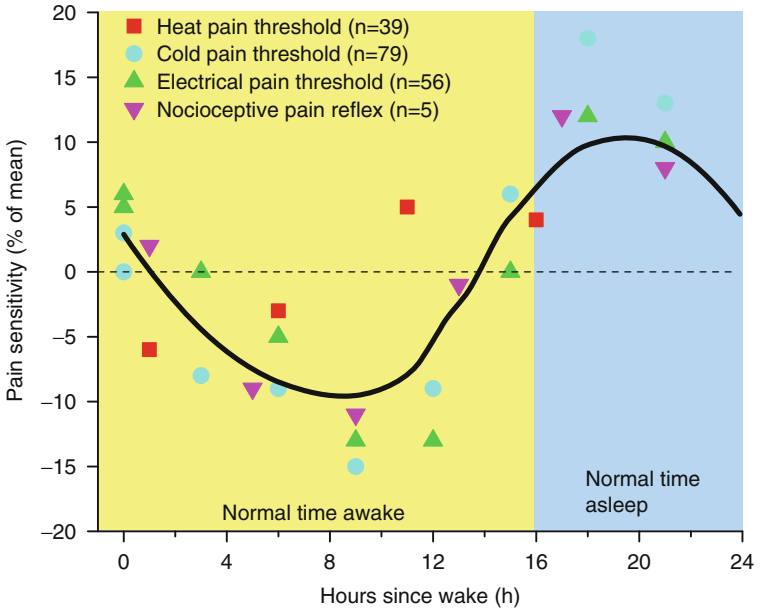


Fig. 1 Prototypical human "daily pain sensitivity" curve constructed from summarized data from four high-quality experimental studies of pain responses [6, 48, 50]. Time was standardized in relation to either scheduled or estimated morning wake time (time of wake = 0). Each point represents the mean value for that time point as derived from the published figures in each study, converted to a percent of the study's overall mean (or the mesor of the depicted rhythm). For ease of use, we inverted measures of pain threshold, so that low pain thresholds are presented in the graph as high "pain sensitivity." The smoothed curve was produced using an unbiased loess{stat} regression in R

experiments, and these experimental results have been wide-ranging. Within human experiments, sleep deprivation or restriction produced no effect on experimental pain [15, 61], small 2–10% increases in pain [33, 44, 57], or much larger 18–118% increases in pain [53, 59, 69]. The diversity of these effects may be due to the variety of sleep protocols used (as suggested by [59]) or the cognitive and emotional context accompanying each experiment (e.g., [61]). Even within a particular protocol, the intensity or quality of experimental pain may determine the impact of sleep deprivation, with one study observing increases in experimental pain that ranged from 6 to 118% depending on the method used to inflict and measure pain [59] (Table 3).

However, we noted that much of the variability in pain sensitivity across studies could be accounted for by the method of normalization used to compare data. For example, when using percentage change as our standardized unit, we can artificially see a larger effect of sleep deprivation on cold pain threshold if the original units are in degrees Celsius instead of in degrees Fahrenheit (Table 4).

Table 3 The estimated impact of sleep deprivation on experimental pain can vary greatly across experimental pain measures within the same study [59]

Schuh-Hofer et al. [59]	Control	Total sleep dep	Difference	% Increase in pain
Cold pain threshold (°C)	14.7	20.4	5.7	38.8
Heat pain threshold (°C)	44.2	41.4	2.8	6.3
Pressure pain threshold (kPa)	446.6	376.6	70	18.6
Mechanical pain threshold (mN)	67.8	43.5	24.3	55.9
Mechanical pain sensitivity (NRS 0–100)	5.5	12	6.5	118.2

In this study, the responses of 14 healthy subjects to five different measurements of evoked pain were assessed after a night of undisturbed sleep (control) and after a night of total sleep deprivation (Total Sleep Dep). The change in response is measured as a percentage of the control response

Table 4 An example of why it is difficult to compare magnitude of effect across different units for measuring pain [59]

Schuh-Hofer et al. [59]	Control	Total sleep dep	% Increase in pain
Cold pain threshold (°C)	14.7	20.4	38.8
Cold pain threshold (°F)	58.5	68.7	17.6

Changes in the threshold temperature for evoked pain by cold stimulus to the hand was measured after a night of undisturbed sleep (control) and after a night of total sleep deprivation (Total Sleep Dep). When measured as a percentage of the control response, the same change in response is computed as a larger percentage change when degrees Celsius are used compared to degrees Fahrenheit

Similarly, across studies, the effect of sleep deprivation in percentage change units was consistently larger on cold pain threshold than on heat pain threshold, simply because the temperature values for cold pain threshold under the control condition were lower, making the denominator in the percentage change equation (i.e., (change from control threshold temperature) / (control threshold temperature)) smaller. Likewise, percentage change increases in subjective rating scales were almost always exaggerated, since ratings from control subjects were often extremely low, making the denominator in the percentage change equation diminutive.

The typical rationale for using percentage change units for comparing data of different units is the idea that the biological impact of changes in the unit depends on its initial values. For example, if a disease condition increases the number of mRNA transcripts for a particular gene from 100 to 110, this is likely to matter more biologically than an increase from 1000 to 1010. It is not clear that this logic holds true for the units used in pain research (as indicated by the particularly irrational examples above). What is likely to matter more biologically is the percentage of the range of stimulation possible before tissue is genuinely damaged. For example, in a heat threshold experiment, you would expect that pain sensation might reflect a range of temperatures between physiological levels (37 °C) and a level of heat that rapidly causes damage (60 °C) [77]. In that case, a drop in pain threshold of 3° would cover 13% of the full range of pain sensation possible ((3°)/(60–37 °C) = 0.13). This is a much more interpretable value than simply saying that a drop in pain threshold from 49 ° in controls to 46° following sleep deprivation is a 6% change from the original pain threshold ((49–46°)/49° = 0.061).

Table 5 Within a study, the estimated impact of sleep deprivation on experimental pain is quite consistent if the data are normalized as a percentage of the full estimated range for painful sensation in that modality [59]

Schuh-Hofer et al. [59]	Observed difference	Minimum pain (no stimulation)	Maximum pain (threshold for tissue damage)	Full range of stimulation (max–min)	% Range of stimulation
Cold pain threshold (°C)	5.7	37	-51	88	6.5
Heat pain threshold (°C)	2.8	37	60	23	12.2
Pressure pain threshold (kPa)	70	0	686	686	10.2
Mechanical pain threshold (mN)	24.3	0	512	512	4.7
Mechanical pain sensitivity (NRS 0–100)	6.5	0	100	100	6.5

The observed difference in response values between the control and total sleep deprivation conditions, as listed in Table 3, was computed as a percentage of the full range of stimulation response values

Using this logic, we found that the data documenting the impact of sleep deprivation on experimental pain was much more consistent than it initially appeared. For the different experimental pain measures used in the study of [59], we determined minimum and maximum response values corresponding to the absence of stimulation and the value when tissue damage would occur, respectively. The full range of response values was computed as the difference between the maximum and minimum response values. We then computed the observed difference in response values between the control and total sleep deprivation conditions, as listed in Table 3, as a percentage of the full range of stimulation. Computed in this way, all experimental pain measures indicate that sleep deprivation increases pain by ∼5–12% of the full range of painful stimulation (Table 5).

To apply this logic to multiple studies on the effects of sleep deprivation and circadian modulation of pain, we estimated minimum and maximum stimulation levels necessary to produce a full range of pain responses to a number of different experimental pain modalities, as well as typical pain thresholds (Table 6). From these measurements, we computed the range of physiologically meaningful stimulation as the difference between the maximum and minimum values, and the range of painful stimulation as the difference between the maximum and pain threshold values. We then normalized the results across studies by converting changes in pain response values to percentage changes within the full range of physiologically meaningful stimulation or percentage changes within the range of painful stimulation. Following this normalization, the magnitudes of the effects of sleep deprivation and daily rhythms were less variable across studies. This implied that normalizing data based on percentage changes within the range of painful stimulation was superior to using a simple percentage of the mean. (However, please note that the data necessary to perform this improved normalization were not available for all studies—for example, several studies used to construct Fig. 1.)

Table 6 Determining the range of painful stimulation possible within human experimental pain studies before the occurrence of tissue damage, as well as the typical threshold for pain sensation

Type of pain measurement	Minimum pain (no stimulation)	Maximum pain (threshold for tissue damage)	Full range of meaningful stimulation (max–min)	Typical pain threshold	Pain threshold as % of range of painful stimulation	Range of painful stimulation (max-threshold)
Cold (°C)[a]	37	-51	88	12	28%	59
Heat (°C)[b]	37	60	23	46	37%	15
Heat (mcal/cm^2)[c]	0	2420	2420	871	36%	1540
Electric stimulation of sural nerve (mA)[d]	0	40	40	12	30%	28
Pressure (kPa)[e]	0	686	686	447	65%	240
Mechanical (mN)[f]	0	512	512	68	13%	444
Tolerable pressure duration (s)[g]	0	360	360	15	13%	345
Mechanical pain sensitivity (NRS 0–100)[h]	0	100	100	1	1%	99
Laser heat VAS pain ratings (0–100)[i]	0	100	100	1	1%	99
Electrical stimulation pain ratings (0–10)[j]	0	10	10	1	10%	9
Tourniquet pain intensity ratings (0–10)[k]	0	10	10	1	10%	9

Sources for computations:
[a]http://-www.ehs.neu.edu/laboratory_safety/fact_sheets/-cryogenic_liquids
[b]Reference [77]
[c]Reference [65]
[d]Maximum for Instrument, the typical mA eliciting nociceptive pain reflex by someone who is under general anesthesia for surgery [72]
[e]Pressure of about 100 lb/in^2 (7 kg/cm^2) is required to penetrate the epidermis (1 kg/cm^2 = 98.07 kPa) [9]
[f]Maximum for Instrument [59]
[g]Maximum for Test, rated "Very strong pain" by all participants [21]
[h]Reference [59]
[i]Reference [69]
[j]Reference [6]
[k]Reference [32]

Using this improved normalization method, we also found that the magnitude of the effects of sleep deprivation and daily rhythms were roughly equivalent (Table 7). Specifically, we found that, on average, evoked pain responses, measured relative to the range of painful stimulation, varied by approximately 14% due to daily rhythms and by approximately 13% in response to sleep deprivation.

Table 7 Experimental results illustrating the effects of daily rhythm and sleep deprivation on pain thresholds in humans for multiple pain modalities

Type of pain measurement	In original units		As % of mean		As % of full range (max–min)		As % of painful range (max-threshold)	
	Daily rhythm	Sleep deprivation	Daily rhythm	Sleep deprivation	Daily rhythm	Sleep deprivation	Daily rhythm	Sleep deprivation
Cold (°C) [33, 59]		3.7		32%		4%		6%
Heat (°C) [33, 57, 59]		2.8		6%		12%		19%
Heat (mcal/cm^2) [50]	122		14%		5%		8%	
Electric stimulation of sural nerve (mA) [6]	2.4		20%		6%		9%	
Pressure (kPa) [59]		70.0		16%		10%		29%
Mechanical (mN) [59]		24.3		36%		5%		5%
Tolerable pressure duration (s) [21]	41		36%		11%		12%	
Mechanical pain sensitivity (NRS 0–100) [59]		6.5		118%		7%		7%
Laser heat VAS pain ratings (0–100) [69]		9.5		23%		10%		10%
Electrical stimulation pain ratings (0–10) [6]	2.3		48%		23%		26%	
Tourniquet pain intensity ratings (0–10) [32]	1.45		18%		15%		16%	
	Average		27%	38%	12%	8%	14%	13%
	STDEV		14%	40%	7%	3%	7%	10%
	Signal/Noise		1.92	0.95	1.65	2.45	1.94	1.33

These results are normalized with respect to the range of physiologically meaningful or painful stimulation computed in Table 6. Across studies, the estimated impact of the daily rhythm and sleep deprivation on experimental pain is quite consistent if the data are normalized as a percentage of either the estimated range for physiologically meaningful sensation or painful sensation in that modality

3.3 A Cross-Species Comparison: Circadian Rhythms and Homeostatic Sleep Drive Influence Pain Sensitivity in Laboratory Rodents

The vast majority of what is known regarding the influence of circadian rhythms and sleep–wake cycles on pain processing circuitry comes from studies on laboratory rodents. In order to properly compare these data with that of humans, it is important to understand that the circadian and sleep systems of laboratory rodents differ from humans in several fundamental ways. To begin with, laboratory mice and rats are nocturnal, which means that most of their wakefulness occurs at night and most of their sleep occurs during the day. They are also polyphasic sleepers, which means that they sleep in short, multi-minute bouts, interrupted by waking, and rarely exhibit consolidated wakefulness that extends beyond several hours. Despite their unconsolidated wake and sleep, they still generally exhibit a progressive build-up of homeostatic sleep drive across the nighttime active period, and dissipation during the daytime rest period [70].

Similar to humans, there is clear evidence that pain sensation in laboratory rodents is modulated by both time-of-day [11, 18, 30, 34, 43, 51, 55, 68, 78] and sleep deprivation [27, 42, 71, 73–75]. Unlike humans, we can easily place laboratory rodents into constant environmental conditions and thus be able to demonstrate with certainty that the influence of time-of-day on pain sensation is due to an endogenous circadian clock instead of simple passive responses to a rhythmic environment [51]. However, the timing of the daily peak in pain sensitivity varies in different strains of inbred rodents by as much as 12 h [11], making it sometimes difficult to draw generalized conclusions about the influence of circadian rhythms on pain sensitivity. Another notable difference between humans and rodents is that the duration of sleep deprivation necessary to observe an effect on pain responses is much smaller, since rodents typically do not exhibit consolidated wakefulness on the scale of multiple hours.

4 Circadian Rhythms and Homeostatic Sleep Drive Modulate Pain Neural Circuitry

The neural location for the circadian modulation of pain begins at the most fundamental level of the pain circuitry: sensory afferent input into the spinal cord. Within the dorsal root ganglia, which are the neural structures that contain the cell bodies for the sensory afferent neurons, there is clear evidence for endogenous circadian rhythmicity. The dorsal root ganglia rhythmically express a full complement of clock genes, which are the genes responsible for generating daily rhythmicity throughout the body [78]. The dorsal root ganglia also demonstrate rhythmic expression of genes necessary for synaptic transmission, including voltage-gated calcium channel subunits [34] and NMDA glutamate receptor subunits [78]. However, since the

dorsal root ganglia contain the cell bodies for a wide variety of afferent neurons, it could be argued that measuring rhythmicity in the dorsal root ganglia as a whole does not necessarily indicate that the nociceptors responsible for pain transmission are rhythmic. Two pieces of evidence suggest otherwise. First, an estimated 82% of afferents are nociceptors [16], thus it is likely that the majority of mRNA collected from the dorsal root ganglia in these experiments represents mRNA from pain transmitting cells. Second, researchers have discovered rhythmic expression of the mRNA and protein for Substance P, a neurotransmitter important for pain-signaling from C fibers [78]. Therefore, it is likely that the nociceptive afferent neurons themselves are rhythmic. That said, the cell bodies for the non-nociceptive fibers in the dorsal root ganglia probably also contain endogenous rhythmicity. In human studies the influence of time-of-day on non-noxious mechanical sensitivity, which is conveyed by Aβ fibers, differs from that of painful stimuli, which is conveyed by Aδ and C fibers, with the rhythm in mechanical sensitivity peaking in the late afternoon (15:00–18:00) and the rhythm in pain sensitivity peaking in the middle of the night (between midnight and 03:00 [48]).

The top-down inhibition of pain processing in the dorsal horn also exhibits a daily rhythm. In humans, placebos best alleviate pain in the early afternoon [48]. In laboratory rodents, there is a daily rhythm in the strength of stress-induced analgesia and endogenous opioid-release, and this rhythm persists under constant environmental conditions [51, 76]. Opioid receptors in the brainstem, which are important for analgesia, exhibit a strong daily rhythm [68]. However, it is possible that these daily rhythms in the top-down inhibition of pain do not represent direct influences of the circadian clock, but instead are a response to the rhythmic build-up and dissipation of homeostatic sleep drive across the day. In support of this theory, there is strong evidence demonstrating that sleep deprivation influences the highest levels of pain processing. Sleep deprivation is already well known to disproportionately affect the energy and resource-needy cortex. Therefore, it is unsurprising that sleep deprivation in humans eliminates distraction-based analgesia [69] and decreases central pain modulation [10, 24]. Even top-down pain inhibition originating from lower levels of the central nervous system is crippled by sleep deprivation, including diffuse noxious inhibitory controls in humans [24, 61] and stress-induced hyperalgesia in rodents [71]. Pharmacological manipulations that mimic top-down pain inhibition, such as morphine, are ineffective following severe sleep deprivation [42, 71].

Sleep deprivation can also alter more fundamental levels of pain processing in the spinal cord, including neurotransmission via glutamate (mGLUR5, NMDA), GABA, and NOS, as well as the passive spread of electric potential via astrocytic gap junctions and the production of reactive oxygen species [73–75]. Despite these effects, under conditions in which the top-down inhibition of pain is minimal, there seems to be less evidence that homeostatic sleep drive influences pain processing. For example, there is some evidence that sleep deprivation may not affect the processing of fast pain (Aδ input). Cortical responses to fast pain actually diminish following sleep restriction [69], and, for faster reflexive behaviors (such as tail withdrawal latency), sleep deprivation effects are sometimes not found [75]. Like-

wise, homeostatic sleep drive does not seem to contribute much to the typical daily rhythm in acute pain sensitivity in rodents. For example, experiments performed in mice with a critical mutation in the essential clock gene Per2 show a complete elimination of daily rhythms in acute pain under typical housing conditions [45, 78], despite maintaining elevated nocturnal activity in response to the laboratory light–dark cycle [62], and thus presumably also retaining a daily rhythm in the build-up and dissipation of homeostatic sleep drive.

In the case of more severe or chronic pain, the influences of homeostatic sleep drive on top-down inhibition may be more relevant. For example, Per2 mutant mice continued to exhibit daily rhythms in inflammatory pain in a manner that matched a predicted build-up and dissipation of homeostatic sleep drive in response to nocturnal behavioral patterns [78]. Both inflammatory and neuropathic conditions are also characterized by a 8–12 h shift in the phasing of daily rhythms in pain sensitivity [20, 34, 68, 78], which may represent an increased influence of homeostatic drive on the top-down inhibition of pain when pain is extended over a longer time scale.

5 Discussion

In summary, there is a substantial body of work documenting the effects of both daily rhythms and sleep deprivation on acute pain sensitivity under experimental conditions in humans and rodents. These results appear divergent at first glance, but upon closer inspection seem to generally agree that peak acute/phasic pain sensitivity in humans occurs during the evening. Our data analysis reveals that the influence of both daily rhythms in pain sensitivity and 24 h of sleep deprivation typically alter pain sensitivity under experimental conditions by 13–14% of the full range of painful stimulation. Other studies suggest that the influence of daily rhythms and sleep deprivation may increase with pain intensity. Where these effects originate physiologically is a more recent source of discussion, but it is likely that they represent the intersecting influence of homeostatic sleep drive and the circadian timekeeping system on the central nervous system. There is clear evidence for circadian effects at the level of the spinal cord and there are equally clear effects for sleep-dependent modulation of the top-down inhibition of pain, although it is possible that both processes influence all levels of the central nervous system. Finally, the effect of circadian rhythms and homeostatic sleep pressure on pain sensitivity may differ depending on the type of pain measured, with data clearly indicating that slower C fiber input and tonic pain sensitivity are influenced by both endogenous circadian rhythms and homeostatic sleep drive, whereas fast Aδ input and faster reflexive pain measures may be less susceptible.

In the companion articles, we introduce two mathematical models to investigate the joint modulation of the circadian rhythm and homeostatic sleep drive on pain. These models address pain at two different levels: at the organismal level as the experience of pain sensitivity and at the neural level as the firing rates of pain processing circuits in the spinal cord.

The organismal-level model addresses a clear gap in our current knowledge: the lack of experimental data measuring the dissociated influence of circadian rhythms and homeostatic sleep drive on pain sensitivity in humans, as would be obtained from a forced desynchrony or constant routine protocol. To address this gap, we adapt the formalism of a classic mathematical model for the regulation of sleep behavior by the circadian rhythm and homeostatic sleep drive, called the Two Process model [13], to simulate the interaction of these two processes on pain sensitivity. The data analysis presented here is used to define a generic "daily pain sensitivity" function (Fig. 1), which we decompose into two independent circadian and homeostatic components (Process C and Process S) using a range of potential relative magnitudes constrained to produce results resembling the normalized data in Table 7. Then we use this model to simulate the resultant changes in the daily pain sensitivity rhythm in response to a variety of altered sleep schedules: sleep deprivation, sleep restriction, and shift work.

The neural-level model is based on the circuitry in the dorsal horn of the spinal cord consisting of synaptically coupled populations of excitatory and inhibitory interneurons that process input signals in the primary afferent fibers and influence the output signal of the WDR neurons. The temporal profile of inputs on the different types of afferent fibers and excitability properties of the included neuronal populations are constrained by experimental results. We validate the model by replicating experimentally observed phenomena of A fiber inhibition of pain and wind-up. We then use the model to investigate mechanisms for the observed phase shift in circadian rhythmicity of pain that occurs with neuropathic pain conditions.

In conclusion, while experimental evidence indicates both circadian and sleep-dependent effects on daily pain rhythms, dissecting their interactions that contribute to changes in pain rhythms under varying normal or pathological conditions is difficult experimentally. The mathematical models developed in this series of papers provide frameworks to incorporate the known experimental results of these effects and to investigate their potential interactions under different conditions. By addressing both the behavioral and cellular levels, these models are useful tools to identify how the primary biological processes of sleep, circadian rhythmicity, and pain interact.

Acknowledgements This work was conducted as a part of A Research Collaboration Workshop for Women in Mathematical Biology at the National Institute for Mathematical and Biological Synthesis, sponsored by the National Science Foundation through NSF Award DBI-1300426, with additional support from The University of Tennessee, Knoxville. This work was additionally partially supported by the following sources: NSF Award DMS-1412119 (VB), NSF RTG grant DMS-1344962 (JC), and the Pritzker Neuropsychiatric Disorders Research Consortium (MH). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Aviram, J., Schochat, T., Pud, D.: Pain perception in healthy young men is modified by time of day and is modality dependent. Pain Med. **16**, 1137–1144 (2015)
2. Bachmann, C.G., Nitsche, M.A., Pfingsten, M., Gersdorff, N., Harder, C., Baier, P.C., Antel, A., Treede, R.D., Paulus, W., Happe, S.: Diurnal time course of heat pain perception in healthy humans. Neurosci. Lett. **489**, 122–125 (2011)
3. Basbaum, A.I., Bautista, D.M., Scherrer, G., Julius, D.: Cellular and molecular mechanisms of pain. Cell **139**(2), 267–284 (2009)
4. Beginski, R.: Untersuchungen über die Schmerzreizschwel-lender Haut unter der Wirkung analgetischer Pharmaka. Med. Dissertation, Universität Kiel (1961)
5. Bochnik, H.J.: Mehrgleisig-simultane Untersuchungen spontaner Tagesschwankungen sensibler, motorischer und vegetativer Funktionen. Nervenarzt **29**, 307 (1958)
6. Bourdalle-Badie, C., Andre, M., Pourquier, P., Robert, S., Cambar, J., Erny, P.: Circadian rhythm of pain in man: study of measure of nociceptive flexion reflex. In: Reinberg, A., Smolensky, M., Labrecque, G. (eds.) Annual Review of Chronopharmacology. Biological Rhythms and Medications, vol. 7, pp. 249–252. Pergamon Press, New York (1990)
7. Bragin, E.O., Durinyan, R.A.: Study of the time course of changes in the daily rhythm of pain sensitivity in rats and in man. Patolgicescaja Ficiologija Eksperimentalnaja Terapija (Moscau) **5**, 22 (1983)
8. Bruguerolle, B., Labrecque, G.: Rhythmic pattern in pain and their chronotherapy. Adv. Drug Deliv. Rev. **59**(9–10), 883–895 (2007)
9. Bullocks, J.M., Hsu, P.W., Izaddoost, S.A., Hollier, L.H., Stal, S.: Plastic Surgery Emergencies: Principles and Techniques. Thieme, Stuttgart (2008)
10. Campbell, C.M., Bounds, S.C., Simango, M.B., Witmer, K.R., Campbell, J.N., Edwards, R.R., Haythornthwaite, J.A., Smith, M.T.: Self-reported sleep duration associated with distraction analgesia, hyperemia, and secondary hyperalgesia in the heat-capsaicin nociceptive model. Eur. J. Pain **15**(6), 561–567 (2011)
11. Castellano, C., Puglisi-Allegra, P., Renzi, S., Oliverio, A.: Genetic differences in daily rhythms of pain sensitivity in mice. Pharmacol. Biochem. Behav. **23**, 91–92 (1985)
12. Chapman, W.P., Jones, C.M.: Variations on cutaneous and visceral pain sensitivity in normal subjects. J. Clin. Invest. **23**(1), 81–91 (1944)
13. Daan, S., Beersma, D.G., Borbely, A.A.: Timing of human sleep: recovery process gated by a circadian pacemaker. Am. J. Physiol. Regul. Integr. Comp. Physiol. **246**(2), R161–R183 (1984)
14. Davis, G.C., Buchsbaum, M.S., Bunney, W.E.: Naloxone decreases diurnal variation in pain sensitivity and somatosensory evoked potentials. Life Sci. **23**(14), 1449–1459 (1978)
15. Drewes, A.M., Rossel, P., Arendt-Nielsen, L., Nielsen, K.D., Hansen, L.M., Birket-Smith, L., Stengaard-Pedersen, K.: Sleepiness does not modulate experimental joint pain in healthy volunteers. Scand. J. Rheumatol. **26**, 399–400 (1997)
16. Dubin, A.E., Patapoutian, A.: Nociceptors: the sensors of the pain pathway. J. Clin. Invest. **120**(11), 3760–3772 (2010)
17. Finan, P.H., Goodin, B.R., Smith, M.T.: The association of sleep and pain: an update and a path forward. J. Pain **14**(12), 1539–1552 (2013)
18. Frederickson, R.C., Burgis, V., Edwards, J.D.: Hyperalgesia induced by naloxone follows diurnal rhythm in responsivity to painful stimuli. Science **198**(4318), 756–758 (1977)
19. Frees, H.J.: Rhythmik in der Schmerzempfindlichkeit der Zähne. Med. Dissertation, Universität Rostock (1937)
20. Gilron, I., Ghasemlou, N.: Chronobiology of chronic pain: focus on diurnal rhythmicity of neuropathic pain. Curr. Opin. Support. Palliat. Care **8**(4), 429–436 (2014)
21. Göbel, H., Cordes, P.: Circadian variation of pain sensitivity in pericranial musculature. Headache **30**(7), 418–422 (1990)
22. Godt, H., Thiele, E.: Über die Abkängigkeit der Zahnschmer-zempfindlichkeit vom individuellen Tagesablauf. Stoma **21**, 184–190 (1968)

23. Grabfield, G.P., Martin, E.G.: Variations in the sensory threshold for faradic stimulation in normal subjects. The diurnal rhythm. Am. J. Physiol. **31**, 300–308 (1912)
24. Haack, M., Lee, E., Cohen, D.A., Mullington, J.M.: Activation of the prostaglandin system in response to sleep loss in healthy humans: potential mediator of increased spontaneous pain. Pain **145**(1–2), 136–141 (2009)
25. Hardy, D.J., Wolff, H.G., Goodell, H.: Studies on pain. A new method for measuring pain threshold. J. Clin. Invest. **19**, 349 (1940)
26. Hummel, T., Kraetsch, H.G., Lotsch, J., Hepper, M., Liefhold, J., Kobal, G.: Analgesic effects of dihydrocodeine and tramadol when administered either in the morning or evening. Chronobiol. Int. **12**(1): 62–72 (1995)
27. Ibironke, G.F., Ajonijebu, C.O.: Sleep deprivation-induced hyperalgesia in rodents: some neurochemical mechanisms. Neurophysiology **46**(5), 411–414 (2014)
28. Jacoby, W.G.: Loess:: a nonparametric, graphical tool for depicting relationships between variables. Elect. Stud. **19**(4), 577–613 (2000)
29. Jores, A., Frees, J.: Die Tagesschwankungen der Schmerz empfindung. DMW **25**, 962–963 (1937)
30. Kavaliers, M., Hirst, M.: Daily rhythms of analgesia in mice: effects of age and photoperiod. Brain Res. **279**(1–2), 387–393 (1983)
31. Kobal, G., Hummel, T., Kraetsch, H.G., Lötsch, J.: Circadian analgesic effect of dihydromorphione and tramadol. In: 5th International Conference on Biological Rhythms and Medications, Amelia Island (FL), Abstract XIII-5 (1992)
32. Koch, H.J., Raschka, C.: Diurnal variation of pain perception in young volunteers using the tourniquet pain model. Chronobiol. Int. **21**(1), 171–173 (2004)
33. Kundermann, B., Spernal, J., Huber, M.T., Krieg, J.-C., Lautenbacher, S.: Sleep deprivation affects thermal pain thresholds but not somatosensory thresholds in healthy volunteers. Psychosom. Med. **66**(6), 932–937 (2004)
34. Kusunore, N., Koyanagi, S., Hamamura, K., Matsunaga, N., Yoshida, M., Uchida, T., Tsuda, M., Inoue, K., Ohdo, S.: Molecular basis for the dosing time-dependency of anti-allodynic effects of gabapentin in a mouse model of neuropathic pain. Mol. Pain **6**(83), 1–8 (2010)
35. Lautenbacher, S., Kundermann, B., Krieg, J.-C.: Sleep deprivation and pain perception. Sleep Med. Rev. **10**(5), 357–369 (2006)
36. Le Bars, D., Gozariu, M., Cadden, S.W.: Animal models of nociception. Pharmacol. Rev. **53**(4), 597–652 (2001)
37. Macht, D.J., Herman, N.B., Levy, C.S.: A quantitative study of the analgesia produced by opium alkaloids, individually and in combination with each other, in normal man. J. Pharm. Exp. Ther. **8**(1), 1–37 (1916)
38. Martin, E.G., Bigelow, G.H., Grabfield, G.B.: Variations in the sensory threshold for faradic stimulation in normal human subjects. II. The nocturnal rhythm. Am. J. Physiol. **33**, 415–422 (1914)
39. Millan, M.J.: Descending control of pain. Prog. Neurobiol. **66**(6), 355–474 (2002)
40. Moayedi, M., Davis, K.D.: Theories of pain: from specificity to gate control. J. Neurophysiol. **109**(1), 5–12 (2013)
41. Morawetz, R.F., Parth, P., Pöppel, E.: Influence of the pain measurement technique on the diurnal variation of pain perception. In: Bromm B (ed) Pain Measurement in Man, pp 409–416. Elsevier, Amsterdam (1984)
42. Nascimento, D.C., Andersen, M.L., Hipólide, D.C., Nobrega, J.N., Tufik, S.: Pain hypersensitivity induced by paradoxical sleep deprivation is not due to altered binding to brain-opioid receptors. Behav. Brain Res. **178**(2), 216–220 (2007)
43. Oliverio, A., Castellano, C., Puglisi-Allegra, S.: Opiate analgesia: evidence for circadian rhythms in mice. Brain Res. **249**(2), 265–270 (1982)
44. Onen, S.H., Alloui, A., Gross, A., Eschallier, A., Dubray, C.: The effects of total sleep deprivation, selective sleep interruption and sleep recovery on pain tolerance thresholds in healthy subjects. J. Sleep Res. **10**(1), 35–42 (2001)

45. Oster, H., Yasui, A., van der Horst, G.T., Albrecht, U.: Disruption of mCry2 restores circadian rhythmicity in mPer2 mutant mice. Genes Dev. **16**(20), 2633–2638 (2002)
46. Pöllmann, L.: Über den Tagesrhythmus der Schmerzemp-findlichkeit der Zähne. Wehrmedizinische Monatschrift **18**, 142–144 (1974)
47. Pöllmann, L.: Der Zahnschmerz. Chronobiologie, Beurteilung und Behandlung. München Wien, Hanser (1980)
48. Pöllmann, L.: Duality of pain demonstrated by the circadian variation in tooth sensitivity. In: Erhard, H., Kabat, H.F. (eds.) Chronobiology 1982–1983, chap. 39, pp. 225–228. Karger, Basel (1984)
49. Procacci, P.: Änderungen der Schmerzschwelle für Haut-stiche während des Tages- und Monatsrhythmus. In: Janzen R, Keidel WD, Herz A, Streichele C (eds.) Schmerz. Grund-lagen, Pharmakologie, Therapie, pp. 47–50. Stuttgart, Thieme (1972)
50. Procacci, P., Della Corte, M., Zoppi, M., Maresca, M.: Rhythmic changes of the cutaneous pain threshold in man: a general review. Chronobiologia **1**(1), 77–96 (1974)
51. PuglisiAllegra, S., Castellano, C., Oliverio, A.: Circadian variations in stress-induced analgesia. Brain Res. **252**(2), 373–376 (1982)
52. Ritter, O.: Untersuchungen über die Bestimmung der Schmerzempfindlichkeit beim Menschen mit Hilfe einer faradischen Reizmethode. Med. Dissertation, Universität Kiel (1958)
53. Roehrs, T., Hyde, M., Blaisdell, B., Greenwald, M., Roth, T.: Sleep loss and REM sleep loss are hyperalgesic. Sleep **29**(2), 145–151 (2006)
54. Rogers, E.J., Viikin, B.: Diurnal variations in sensory and pain thresholds correlated with mood states. J. Clin. Psych. **39**, 431–432 (1978)
55. Roseboom, P.H., Namboodiri, M.A., Zimonjic, D.B., Popescu, N.C., Rodriguez, I.R., Gastel, J.A., Klein, D.C.: Natural melatonin 'knockdown' in C57BL/6J mice: rare mechanism truncates serotonin N-acetyltransferase. Brain Res. Mol. Brain Res. **63**(1), 189–197 (1998)
56. Sandrini, G., Alfonsi, E., Bono, G., Facchinetti, F., Montalbetti, L., Nappi, G.: Circadian variations of human flexion reflex. Pain **25**(3), 403–410 (1986)
57. Schestatsky, P., Dall-Agnol, L., Gheller, L., Stefani, L.C., Sanches, P.R.S., de Souza, I.C., Torres, I.L., Caumo, W.: Pain-autonomic interaction after work-induced sleep restriction. Eur. J. Neurol. **20**(4), 638–646 (2013)
58. Schröder: Der Inductionsstrom als Diagnosticum in der zahnärztlichen Praxis. Korrespondenzblatt für Zahnärzte **34**, 68–78 (1905)
59. Schuh-Hofer, S., Wodarski, R., Pfau, D.B., Caspani, O., Magerl, W., Kennedy, J.D., Treede, R.-D.: One night of total sleep deprivation promotes a state of generalized hyperalgesia: a surrogate pain model to study the relationship of insomnia and pain. Pain **154**(9), 1613–1621 (2013)
60. Schuhmacher, G.A., Goodell, H., Hardy, J.D., Wolff, H.G.: Uniformity of pain threshold in man. Science **92**, 110–112 (1940)
61. Smith, M.T., Edwards, R.R., McCann, U.D., Haythornthwaite, J.A.: The effects of sleep deprivation on pain inhibition and spontaneous pain in women. Sleep **30**(4), 494–505 (2007)
62. Spoelstra, K., Albrecht, U., van der Horst, G.T.J., Brauer, V., Daan, S.: Phase responses to light pulses in mice lacking functional per or cry genes. J. Biol. Rhythms **19**(6), 518–529 (2004)
63. Stacher, G., Bauer, P., Schneider, C., Winklehner, S., Schmierer, G.: Effects of combination of oral naproxen sodium and codeine on experimentally induced pain. J. Clin. Pharm. **21**(6), 485 (1982)
64. Starek, H.: Untersuchungen der Schmerzschwelle des Menschen mit dem Algesimeter für strahlende Wärme. Med. Dissertation, Universität Zürich (1957)
65. Stoll, A.M., Greene, L.C.: Relationship between pain and tissue damage due to thermal radiation. J. Appl. Physiol. **14**(3), 373–382 (1959)
66. Strempel, H.: Circadian cycles of epicritic and protopathic pain threshold. J. Interdis. Cycle Res. **8**, 276–280 (1977)
67. Strian, F., Lautenbacher, S., Galfe, G., Hölzl, R.: Diurnal variations in pain perception and thermal sensitivity. Pain **36**, 125–131 (1989)

68. Takada, T., Yamashita, A., Date, A., Yanase, M., Suhara, Y., Hamada, A., Sakai, H., Ikegami, D., Iseki, M., Inada, E., Narita, M.: Changes in the circadian rhythm of mRNA expression for μ-opioid receptors in the periaqueductal gray under a neuropathic pain-like state. Synapse **67**(5), 216–223 (2013)
69. Tiede, W., Magerl, W., Baumgartner, U., Durrer, B., Ehlert, U., Treede, R.D.: Sleep restriction attenuates amplitudes and attentional modulation of pain-related evoked potentials, but augments pain ratings in healthy volunteers. Pain **148**(1), 36–42 (2010)
70. Tobler, I., Borbély, A.A.: Sleep eeg in the rat as a function of prior waking. Electroencephalogr. Clin. Neurophysiol. **64**(1), 74–76 (1986)
71. Ukponmwan, O.E., Rupreht, J., Dzoljic, M.R.: REM sleep deprivation decreases the antinociceptive property of enkephalinase-inhibition, morphine and cold-water-swim. Gen. Pharmacol. **15**(3), 255–258 (1984)
72. von Dincklage, F., Hackbarth, M., Mager, R., Rehberg, B., Baars, J.H.: Monitoring of the responsiveness to noxious stimuli during anaesthesia with propofol and remifentanil by using RIII reflex threshold and bispectral index. Br. J. Anaesth. **104**(2), 201–208 (2010)
73. Wei, H., Gong, N., Huang, J.L., Fan, H., Ma, A.N., Li, X.Y., Wang, Y.X., Pertovaara, A.: Spinal D-amino acid oxidase contributes to mechanical pain hypersensitivity induced by sleep deprivation in the rat. Pharmacol. Biochem. Behav. **111**, 30–36 (2013)
74. Wei, H., Hao, B., Huang, J.-L., Ma, A.-N., Li, X.-Y., Wang, Y.-X., Pertovaara, A.: Intrathecal administration of a gap junction decoupler, an inhibitor of Na+ K+ 2Cl cotransporter 1, or a GABAA receptor agonist attenuates mechanical pain hypersensitivity induced by REM sleep deprivation in the rat. Pharmacol. Biochem. Behav. **97**(2), 377–383 (2010)
75. Wei, H., Zhao, W., Wang, Y.X., Pertovaara, A.: Pain-related behavior following REM sleep deprivation in the rat: influence of peripheral nerve injury, spinal glutamatergic receptors and nitric oxide. Brain Res. **1148**, 105–112 (2007)
76. Wesche, D.L., Frederickson, R.C.A.: Diurnal differences in opioid peptide levels correlated with nociceptive sensitivity. Life Sci. **24**(20), 1861–1867 (1979)
77. Burn exposure chart. www.antiscald.com
78. Zhang, J., Li, H., Teng, H., Zhang, T., Luo, Y., Zhao, M., Li, Y.Q., Sun, Z.S.: Regulation of peripheral clock to oscillation of substance P contributes to circadian inflammatory pain. Anesthesiology **117**(1), 149–160 (2012)

Investigating Circadian Rhythmicity in Pain Sensitivity Using a Neural Circuit Model for Spinal Cord Processing of Pain

Jennifer A. Crodelle, Sofia H. Piltz, Victoria Booth, and Megan Hastings Hagenauer

Abstract Primary processing of painful stimulation occurs in the dorsal horn of the spinal cord. In this article, we introduce mathematical models of the neural circuitry in the dorsal horn responsible for processing nerve fiber inputs from noxious stimulation of peripheral tissues and generating the resultant pain signal. The differential equation models describe the average firing rates of excitatory and inhibitory interneuron populations, as well as the wide dynamic range (WDR) neurons whose output correlates with the pain signal. The temporal profile of inputs on the different afferent nerve fibers that signal noxious and innocuous stimulation and the excitability properties of the included neuronal populations are constrained by experimental results. We consider models for the spinal cord circuit in isolation and when top-down inputs from higher brain areas that modulate pain processing are included. We validate the models by replicating experimentally observed phenomena of A fiber inhibition of pain and wind-up. We then use the models to investigate mechanisms for the observed phase shift in circadian

J.A. Crodelle
Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
e-mail: kilej@rpi.edu

S.H. Piltz
Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kongens Lyngby, Denmark
e-mail: piltz@umich.edu

V. Booth (✉)
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA

Department of Anesthesiology, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: vbooth@umich.edu

M.H. Hagenauer
Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: hagenaue@umich.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_2

rhythmicity of pain that occurs with neuropathic pain conditions. Our results suggest that changes in neuropathic pain rhythmicity can occur through dysregulation of inhibition within the dorsal horn circuit.

MSC codes: 37C99, 37N25, 92B25, 92C20

1 The Neural Processing of Pain

The ability for an organism to detect pain is essential for its survival. It is intuitive that the processing of pain must engage a wide variety of neural circuits ranging from the spinal cord, up through the brainstem, thalamus, and cortex. Though this is true, many of the higher-level cognitive and emotional influences re-converge at the level of the spinal cord, to gate the input of nociceptive information entering the dorsal horn. The dorsal horn serves as a processing center for incoming pain signals, while the midbrain and cortex, as a whole referred to as descending or top-down inhibition [21], serve as a modulator of the pain circuit in the dorsal horn. As a result, there is a tradition of modeling pain processing by focusing exclusively on spinal cord circuitry.

The neural circuit in the dorsal horn receives information about stimulation of peripheral tissues from several types of primary afferent nerve fibers. Nerve fibers called nociceptors detect painful stimuli and are only activated when a stimulus exceeds a specific threshold. These afferents have their cell bodies in the dorsal root ganglia, a cluster of nerve cell bodies located in the back of the spinal cord, and their axons, or afferent fibers, reach to the dorsal horn [3].

There are two major classes of nociceptive fibers: medium diameter, myelinated, fast conducting Aδ fibers that mediate localized, fast pain, and small diameter, unmyelinated, slow conducting C fibers that mediate diffused, slow pain. In addition to the two nociceptive fibers, there are large diameter, myelinated, rapidly conducting Aβ fibers that respond to innocuous, mechanical stimulation [24]. The dorsal horn circuit is composed of many populations of neurons, including excitatory and inhibitory interneurons, and the wide dynamic range (WDR) neurons, or projection neurons. These WDR neurons respond to input from all fibers and constitute the majority of the output from the dorsal horn circuit up to the brain.

In this article, we introduce a mathematical model of the pain processing neural circuit in the dorsal horn. We are particularly interested in using the model to investigate mechanisms for circadian and sleep-dependent modulation of pain sensitivity. As reviewed in [9], pain sensitivity exhibits a daily rhythm with a trough in the late afternoon and a peak sometime after midnight for humans. There are several hypotheses for the source of this circadian rhythm, including the sensory afferent fibers, and the top-down inhibition. Since the dorsal root ganglia rhythmically express clock genes responsible for generating circadian rhythmicity of other physiological processes [35], and the dorsal root ganglia contain the cell

bodies for a wide variety of afferent neurons, we assume for our model that circadian modulation occurs at the level of primary afferent input to the spinal cord. In the case of more severe or chronic pain, the influences of homeostatic sleep drive on top-down inhibition may be more relevant [35]. We aim to use our mathematical model of the pain circuit in the dorsal horn to form hypotheses on where the modulation might occur and how this placement can affect the firing behavior of the projection neurons.

1.1 Previous Models of Pain Processing

There is a long history to understanding how the body perceives pain, including many conflicting theories. Today's main theory of pain, the gate control theory of pain, was developed in 1965 by Ronald Melzack and Charles Patrick Wall [19]. These researchers revolutionized the understanding of the pain pathway by scrutinizing previous conceptual models of pain processing and developing a model that accounts for the experimental evidence seen thus far. The gate control theory of pain posits that the neural circuitry in the dorsal horn exhibits a gating mechanism that is modulated by activity in the Aβ and C afferent fibers [22]. The nociceptive C fibers facilitate activity in the dorsal horn circuit, whereas the Aβ fibers inhibit activity. When the amount of painful stimuli (activity in C fibers) outweighs the inhibition from the Aβ fibers, the "gate opens" and activates the WDR neurons and, thus, the experience of pain. Experimentalists have used this theory to frame their investigations on the types of fibers that project to the spinal cord, as well as the role of different neuron types in the dorsal horn.

Although the gate control theory of pain [19] is a simplification and not a complete representation of the physiological underpinnings of pain [22], it has been a productive starting point for several mathematical and computational models of pain. These models in turn have given insight into the underlying mechanism of pain. The gate control theory was shown to explain several observed phenomena in pain and suggested a possible mechanistic explanation for rhythmic pain (i.e., a sudden change in the input from fast or slow afferent fibers) [6]. Later in [5], the authors considered an excitatory and inhibitory connection from the mid-brain to the inhibitory interneurons and projection neurons, respectively, to be included in the model developed in [6]. This generalization made it possible to take the effect of N-methyl-D-aspartate (NMDA) receptors into account, and therefore, allowed for the resulting model to successfully reproduce the "wind-up" mechanism [5] —that is, an increased level of activity in a neuron that is being repeatedly stimulated [20].

Similarly to [5, 6], more recent models of pain have considered a modeling framework at the level of a single neuron. These biophysically detailed models of pain have been constructed by connecting compartmental models of individual neurons in the dorsal horn according to the circuit architecture proposed by the gate control theory [19]. In these models, the action potential firing of an individual neuron is described by a Hodgkin–Huxley model of membrane current [13] with

appropriate membrane dynamics and synaptic strengths based on experimental data. This approach allows for a detailed representation of the geometry and biophysics of each neuron connected to the other neurons via a network whose biophysical behavior and characteristics are then calculated numerically [12]. Such a network model has been previously constructed for the interaction between a deep dorsal horn neuron and Aδ fibers [17], for a wide dynamic range projection neuron [1], and for the dorsal horn circuit between a projection, inhibitory, and excitatory neuron [36]. All these models were validated by showing that they are able to reproduce observed phenomena such as wind-up in the presence of nonzero calcium conductances and NMDA [1, 17, 36]. In addition to wind-up, the model in [36] exhibits also pain inhibition via a response to a stimulus in the A fibers, as has been observed experimentally [34]. On the other end of the modeling spectrum, Arle et al. [2] have constructed a very large-scale, physiologically accurate network model of spinal cord neural circuitry that includes numerous known cell types, their laminar distribution, and their modes of connectivity. In addition to simulating pain signaling, the network accounts for the primary motor reflex responses. They applied the model to investigate the mechanisms of pain relief through dorsal column stimulation (DCS), a procedure used to treat neuropathic pain. Their results identify limitations of gate control theory and propose alternate circuitry that more accurately accounts for the effects on nociceptive and neuropathic pain of DCS.

In this work, we take a similar approach to the previous models of pain in terms of the network architecture in the dorsal horn proposed by the gate control theory [20]. However, instead of considering a detailed biophysical model of a single neuron as in [1, 17, 36] or a large-scale network of individual neurons as in [2], we construct equations to describe the population activity of projection, inhibitory, and excitatory neurons in the dorsal horn. As a result, we work with average firing rates of each of the three neuron populations according to the formalism developed in [33]. Therefore, our modeling approach is similar to [6], but we give our model predictions in terms of average firing rates of neuron populations instead of potentials of individual cells.

Thus far, we have not encountered an average firing rate model for pain in the literature. Our choice of modeling framework and dynamic variables is motivated by our long-term aim to integrate a model for pain into an existing model for the sleep–wake cycle constructed in terms of the average firing rate of sleep- and wake-promoting neuron populations. Such a combined sleep–wake–pain model would allow us to test existing hypotheses and ask several biologically motivated questions from the model, for example, on the coupling between sleep, circadian rhythms, and pain sensitivity [7], including the case of chronic pain which is not assessed by the existing biophysical models of pain in [17, 36].

2 Mathematical Model

In this section, we construct our model for the dorsal horn circuit. We choose a firing rate model in which we describe the firing rate of the projection, inhibitory interneuron, and excitatory interneuron populations in the dorsal horn circuit. The following sections define equations of time evolution for the firing rate of populations, as well as the response functions for each population, arrival times for the afferent fibers, and connectivity between populations.

2.1 Equations of Time Evolution

In our model for pain processing, we focus on the dorsal horn and construct equations for the average firing rate of three interconnected neuron populations in the dorsal horn circuit. We assume that the input to our model is a stimulation of the afferent fibers that has been pre-processed in the dorsal root ganglion. Based on this model input, and on the connections between the neuron populations in the dorsal horn, our model predicts the activity of the projection neurons that then proceeds to the mid-brain (see Fig. 1).

In the dorsal horn circuit, the population of the wide dynamic range (i.e., neurons that respond to both nociceptive and non-nociceptive stimuli) projection neurons (W) is connected to the population of inhibitory interneurons (I) and excitatory interneurons (E) (see Fig. 1). According to the formalism of the average firing rate models, we follow [33] and assume that the rate of change of the average firing rate in Hz (i.e., average number of spikes per unit time) of the projection, inhibitory, and excitatory neuron populations, f_W, f_I, and f_E, respectively, is determined by a nonlinear response function (that we define in Sect. 2.2). These response functions

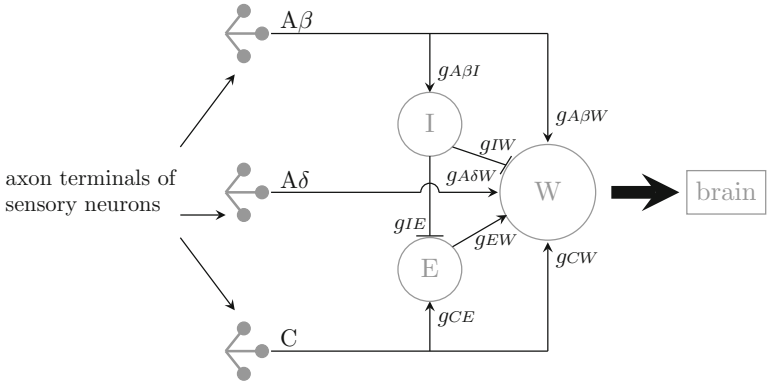


Fig. 1 Diagram of our biophysical model for the dorsal horn circuit. For variable names and default parameter values, see Table 1

determine the average firing rate of a neuron population based on the external inputs (i.e., stimulations of the afferent fibers pre-processed in the dorsal root ganglion) and the firing rates of the presynaptic neuron populations (see Fig. 1). In the absence of input from other neuron populations and afferent fibers, we assume the average firing rate of the neuron population decays exponentially. These assumptions yield the following equations for the average firing rate of each population:

$$\begin{aligned} \frac{df_W}{dt} &= \frac{W_\infty(g_{ABW}f_{AB} + g_{ADW}f_{AD} + g_{CW}f_C + g_{EW}f_E - g_{IW}f_I) - f_W}{\tau_W}, \\ \frac{df_E}{dt} &= \frac{E_\infty(g_{CE}f_C - g_{IE}f_I) - f_E}{\tau_E}, \\ \frac{df_I}{dt} &= \frac{I_\infty(g_{ABI}f_{AB}) - f_I}{\tau_I}, \end{aligned} \tag{1}$$

where τ_W, τ_E, and τ_I are the intrinsic timescales of the projection, excitatory, and inhibitory neuron populations, respectively. Weight g_{NM} denotes the strength of the effect a change in an external input or presynaptic neuron population N has on neuron population M. We indicate inhibitory synaptic input with a negative sign and excitatory synaptic input with a positive sign. We define the step functions of the external inputs, f_{AB}, f_{AD}, and f_C, and the monotonically increasing firing rate response functions W_∞, E_∞, and I_∞, in the following Sects. 2.1.1 and 2.2, respectively. For a summary of all model variables and parameters, see Table 1.

Table 1 Summary of our model [in Eq. (1)] variables, parameters, and default parameter values

Name	Description	Default value
g_{ABW}	Weight of the synaptic connection from Aβ fibers to W	0.6
g_{ABI}	Weight of the synaptic connection from Aβ fibers to I	0.6
g_{ADW}	Weight of the synaptic connection from Aδ fibers to W	0.3
g_{CE}	Weight of the synaptic connection from C fibers to E	0.6
g_{CW}	Weight of the synaptic connection from C fibers to W	0.4
g_{EW}	Weight of the synaptic connection from E to W	0.4
g_{IE}	Weight of the synaptic connection from I to E	0.05
g_{IW}	Weight of the synaptic connection from I to W	0.15
$\max_{AB}$	Amplitude of the Aβ fiber model input	2
$\max_{AD}$	Amplitude of the Aδ fiber model input	0.5
$\max_C$	Amplitude of the C fiber model input	1.5
$\max_E$	Maximum firing rate of E	60 Hz
$\max_I$	Maximum firing rate of I	80 Hz
$\max_W$	Maximum firing rate of W	50 Hz
τ_E	Intrinsic timescale of E	20 ms
τ_I	Intrinsic timescale of I	20 ms
τ_W	Intrinsic timescale of W	1 ms

2.1.1 Model Inputs from the Dorsal Root Ganglion

The three different types of afferent fibers not only have different sizes of diameter, but they also differ at the level of myelination that provides insulation. As a result, impulses are transmitted at different conductance speeds in the three afferent fibers. To determine the pattern of nerve input from a painful stimulus to the spinal cord (see Fig. 1a in [25]), we simulate the arrival of 1000 nerve impulses. The majority (82%) of these fibers consist of slow C fibers (with an average conduction velocity of 1.25 m/s and a standard deviation of 0.75 m/s), 9% as Aδ fiber fibers (with an average conduction velocity of 0.12 m/s and a standard deviation of 0.083 m/s), and 9% as Aβ fibers (with an average conduction velocity of 0.024 m/s and standard deviation of 0.013 m/s). We assume that the time of initiation of each of the nerve pulses in each of these fibers in the periphery in response to painful stimulation is roughly equivalent and that they need to travel 1 meter to reach the spinal cord (e.g., the length of a leg). We choose these proportions and conductance speeds based on the literature [16, 24]. Our simulated data from fibers with different conductance speeds reproduces the observed pattern [25] of a fast response to the Aβ and Aδ fibers (i.e., first pain) followed by a slow response to the C fibers (i.e., second pain) (see Fig. 2).

To generate a simplified model input similar to the simulated input in Fig. 2, we use Heaviside step functions to represent how a stimulation (of the afferent Aβ, Aδ, and C fibers) and its pre-processing in the dorsal root ganglion is received by the dorsal horn circuit. Thus, the external inputs f_{AB}, f_{AD}, and f_C to the model in Eq. (1) are given by

$$\begin{aligned} f_{AB}(t) &= \max_{\mathrm{AB}} H(t - t_{ON_{AB}}) H(t_{OFF_{AB}} - t)\,, \\ f_{AD}(t) &= \max_{\mathrm{AD}} H(t - t_{ON_{AD}}) H(t_{OFF_{AD}} - t)\,, \\ f_C(t) &= \max_{\mathrm{C}} H(t - t_{ON_C}) H(t_{OFF_C} - t)\,, \end{aligned} \tag{2}$$

where $\max_{AB}$, $\max_{AD}$, and $\max_C$ are the amplitudes of the signals from Aβ, Aδ, and C fibers, respectively; $t_{ON_{AB}}$, $t_{ON_{AD}}$, and t_{ON_C} are the time points when an input from Aβ, Aδ, and C fibers, respectively, is received by the dorsal horn circuit; $t_{OFF_{AB}}$, $t_{OFF_{AD}}$, and t_{OFF_C} are the time points when an input from Aβ, Aδ fiber, and C fibers, respectively, has decayed; and $H(x)$ is a Heaviside step function

$$H(x) = \begin{cases} 0, & \text{if } x < 0\,, \\ \frac{1}{2}, & \text{if } x = 0\,, \\ 1, & \text{if } x > 0\,. \end{cases} \tag{3}$$

In Fig. 3, we show an example input that mimics the combined signal from the three afferent fibers and that we use as an input to our model in Eq. (1).

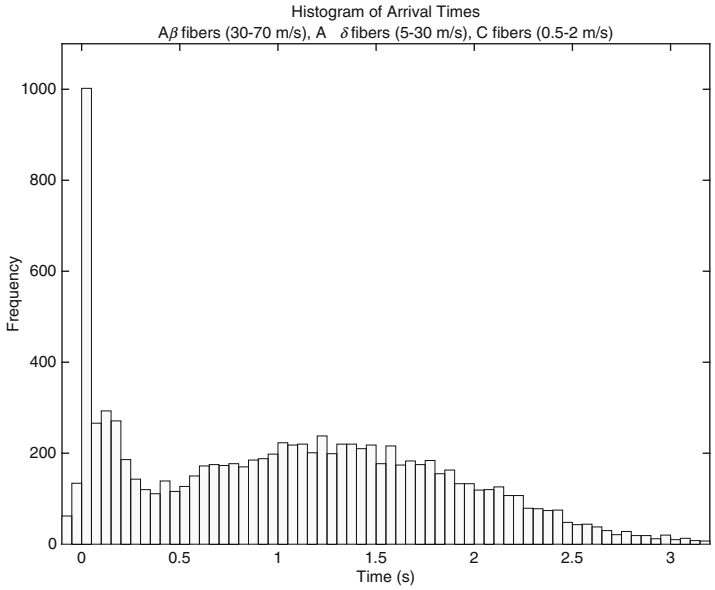


Fig. 2 Simulated post-stimulus histogram from three afferent fibers with different conductance speeds reproduces a pattern observed in the action potential of projection neurons [25] where fast pain response is composed of a response to stimuli in the Aβ and Aδ fibers followed by slow pain response to stimuli in the C fibers

2.2 Firing Rate Response Functions

In our modeling framework, we assume a sigmoidal shape for the monotonically increasing firing rate response functions W_∞, E_∞, and I_∞, and use hyperbolic tangent functions to represent them

$$\begin{aligned} W_\infty(c) &= \max_W \frac{1}{2}\left(1+\tanh\left(\frac{1}{\alpha_W}(c-\beta_W)\right)\right), \\ E_\infty(c) &= \max_E \frac{1}{2}\left(1+\tanh\left(\frac{1}{\alpha_W}(c-\beta_W)\right)\right), \\ I_\infty(c) &= \max_I \frac{1}{2}\left(1+\tanh\left(\frac{1}{\alpha_W}(c-\beta_W)\right)\right), \end{aligned} \tag{4}$$

where $\max_W$, $\max_E$, and $\max_I$ are the maximum firing rates of the projection, excitatory, and inhibitory neuron population, respectively. In Eq. (4), the shape of the tanh-functions is determined by the input c at which the average firing rate of the projection, excitatory, and inhibitory neuron population reaches half of its maximum

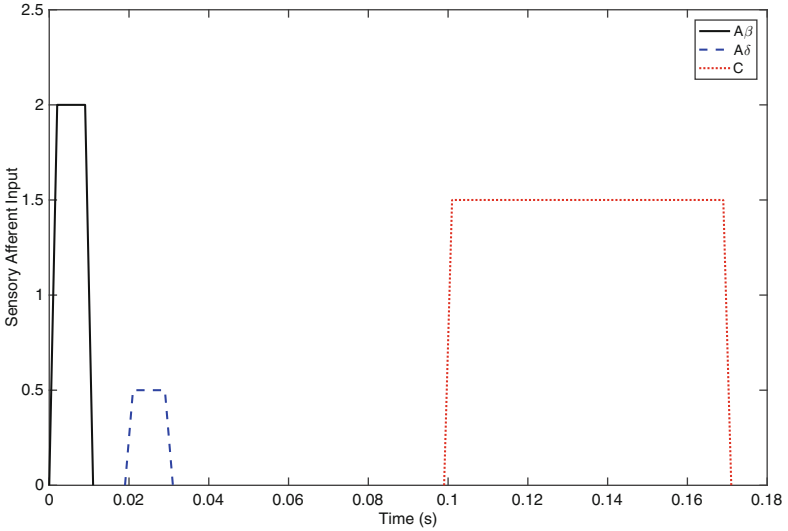


Fig. 3 Simulated model input to the dorsal horn circuit from the afferent fibers after preprocessing in the dorsal root ganglion

value, $c = \beta_W$, $c = \beta_E$, and $c = \beta_I$, respectively. The slope of the transition from non-firing to firing in the projection, excitatory, and inhibitory neuron population is given by $1/\alpha_W$, $1/\alpha_E$, and $1/\alpha_I$, respectively. See Table 1 for default parameter values. We choose the parameter values for the tanh-functions in such a way that the input–output curve of the projection, excitatory, and inhibitory neuron population agrees qualitatively with experimental observations (see Fig. 4). Hence, we assume the inhibitory interneuron population has a nonzero resting firing rate, as has been reported in [3, 21], and a higher maximum firing rate than that of the projection and excitatory interneuron populations, as has been assumed in a biophysically detailed model of the dorsal horn circuit [36]. In our model assumptions for the response functions, we mimic the model predictions of [36] that agree with data from experimental observations in [18, 26]. Thus, we assume that for a small input, the excitatory interneuron population has a small average firing rate that, however, reaches a higher maximum than that of the projection neuron population for a large input (see Fig. 4).

3 Model Validation

In this section, we set out to show that our model reproduces various experimental observations such as pain inhibition, wind-up, and neuropathic phase changes.

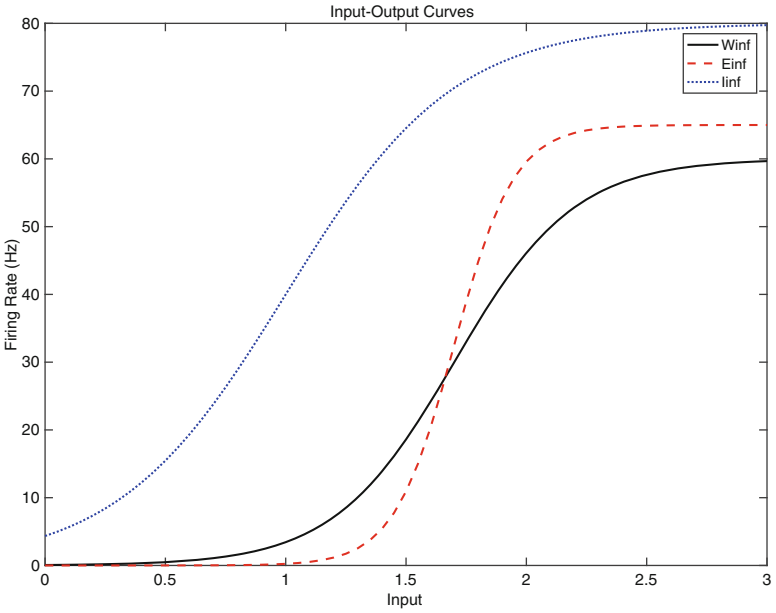


Fig. 4 Response functions of the projection (*black*), excitatory (*red*), and inhibitory (*blue*) neuron populations for different constant inputs (on the x-axis). For parameter values, see Table 1

First, we show that our model reproduces the average firing rate pattern of the populations of neurons in the dorsal horn when the three afferent fibers differ in their conductance speeds. That is, as a response to the input from the afferent fibers as shown in Fig. 2, the average firing rates of the projection and interneuron populations [which are connected to each other as shown in Fig. 1 and whose dynamics are modeled as in Eq. (1)] are qualitatively similar to the simulated histogram in Fig. 2 and also seen experimentally [e.g., see Fig. 1a in [25] (see Fig. 5)]. In addition, the model captures the expected tonic firing rate in the inhibitory neuron population [3, 21], as well as captures the low firing rate of the excitatory neurons [18, 26] (see Fig. 4). We use the model output shown in Fig. 5 as our point of comparison when choosing "default" values for the weights g_{NN} (see Table 1) representing the strength of the connections between the neuron populations as shown in Fig. 1.

3.1 Pain Inhibition

It has been experimentally observed that stimulation of A fiber afferents can lead to inhibition in some wide range projection neurons that typically follows from a stimulation of the C fiber afferents [36]. This is related to the idea that when you

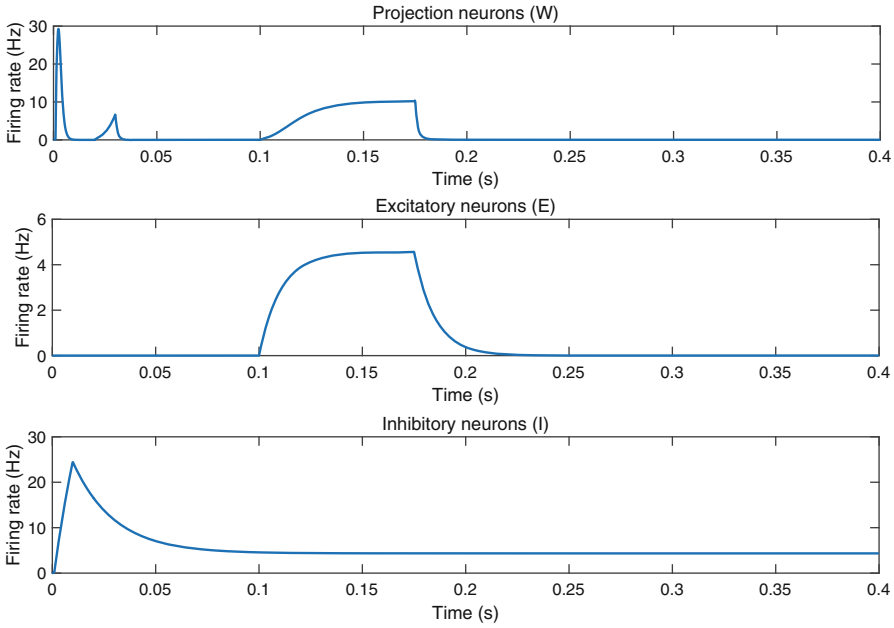


Fig. 5 Firing rates for each population in response to the input from afferent fibers as described in Fig. 3

stub your toe, you immediately apply pressure on the toe and feel some lessening of pain. To capture this phenomenon in our model, we stimulate all three fibers (stubbing of the toe) and then deliver a pulse to the Aβ fiber a short time thereafter (pressure applied to toe), shown in Fig. 6 by the red arrows. The arrival time of the second pulse to the Aβ fiber is increased by 10ms for each simulation, and the response in the projection neurons is shown in blue. As can be seen in Fig. 6, the timing of the second pulse gets closer to the arrival of the C fiber stimulation, and there is a brief period of excitation followed by a longer period of inhibition, as seen in experiments [34]. Thus, our model successfully captures this delayed inhibition phenomenon.

3.2 Wind-Up

We aim to further validate our dorsal horn circuit model (1) by showing that it reproduces wind-up —that is, increased (and frequency-dependent) excitability of the neurons in the spinal cord because of repetitive stimulation of afferent C fibers [20]. Wind-up serves as an important tool for studying the role the spinal cord plays in sensing of pain, and it has been often used as an example phenomenon to validate

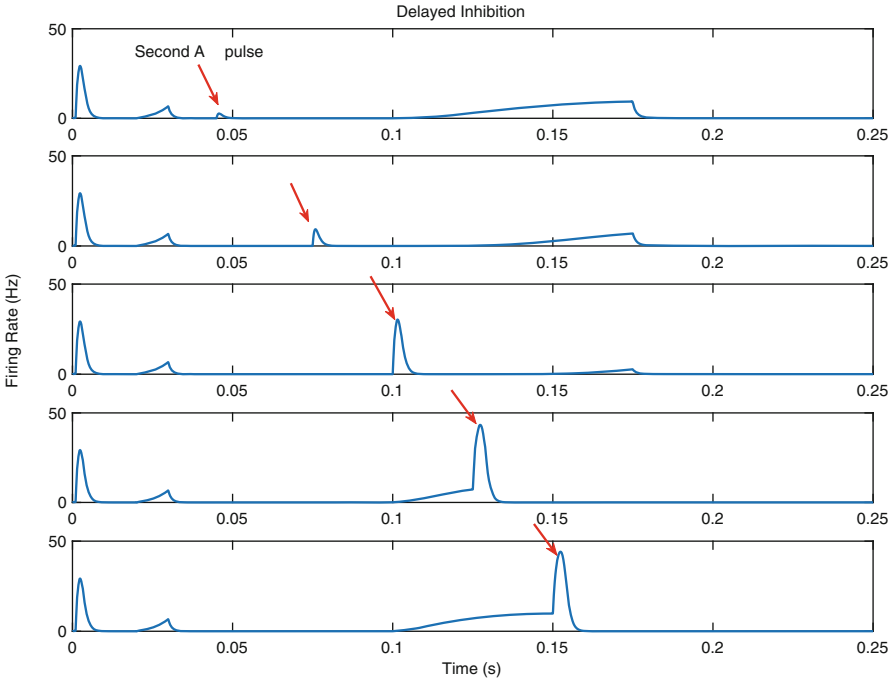


Fig. 6 Pain inhibition phenomenon captured in the model. Response of the projection neuron population to the initial fiber pulse stimulation (at $t = 0$) and the second pulse stimulation only to Aβ fibers (*red arrow*) for increasing in time between fiber stimulations

single neuron models of the dorsal horn (see [1, 17, 36], for example). However, both the physiological meaning and the generation of wind-up remain unclear (see [11] for a review).

There are several possible molecular mechanisms proposed for the generation of wind-up (see Fig. 6 in [11]). Earlier work on single neuron models suggests that wind-up is generated by a combination of long-lasting responses to NMDA and calcium currents providing for cumulative depolarization [1]. Indeed, calcium conductances and NMDA receptors of the projection/deep dorsal horn neurons are included in all previous models of the dorsal horn [1, 17, 36]. In contrast to the model in [1], wind-up can also be reproduced in the absence of synapses that express gamma-Aminobutyric (GABA) from C fibers to the projection neuron [36]. The study done in [17] emphasizes the effect (direct or via influencing the dependence of the deep dorsal horn neurons on their intrinsic calcium currents) NMDA and inhibitory conductances have on the extent of wind-up in the deep dorsal horn neurons [17].

Experimental data on superficial and deep dorsal horn suggest that wind-up is exhibited more by the deep than by the superficial dorsal horn neurons [28]. However, wind-up in the potential of the C fibers is observed in the superficial but not in the deep dorsal horn [27]. Similarly to [1], we investigate whether wind-

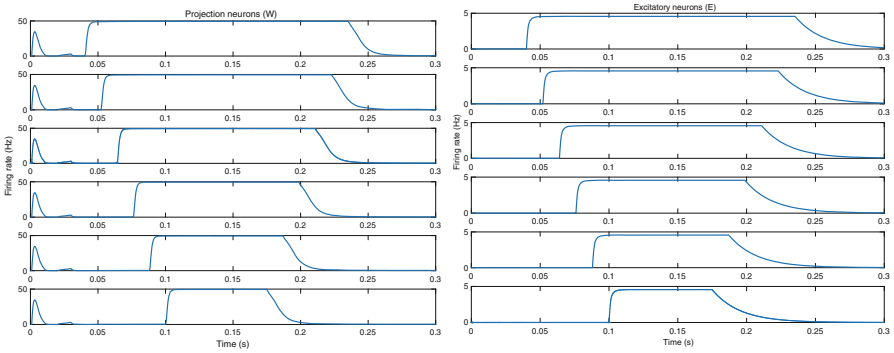


Fig. 7 (*Left*) Projection and (*right*) excitatory interneuron activity predicted by the model in (1) and otherwise as shown in Table 1] when C fibers are stimulated repeatedly. We assume that repeated stimulation of C fibers is experienced in the neurons upstream from the deep dorsal horn and, thus, seen as an increase in the duration of the C fiber model input (see Fig. 3) to the dorsal horn circuit (1). In other words, the model input from C fibers arrives to the dorsal horn circuit at time $t_{ONC} = 0.01s + n0.012s$, where $n = 0, 1, 2, 3, 4, 5$ and n increasing from bottom to top panel

up of the wide dynamic range projection neurons in the dorsal horn circuit can be explained by an increase in the C fiber response before the C-input reaches the dorsal horn circuit. Thus, we assume wind-up occurs "upstream" from the dorsal horn circuit described by our model in (1) and represent it as an increase in the duration, and as a decrease in the arrival time, of the C fiber model input to the dorsal horn circuit.

Increase in C fiber synaptic efficacy has been proposed as a possible generation mechanism for wind-up in the literature [27] and suggested as one of the molecular mechanisms underlying wind-up (see Fig. 6 in [11]). Similarly to [1], our model predicts an increase in the activity of the projection neurons for an increase in the width of the step input from C fibers (see Fig. 7, left). Furthermore, as in [1], our model also predicts that wind-up in the excitatory interneurons (as a response to the change in the C fiber model input) is similar to that seen in the projection neurons (see Fig. 7, right). However, such behavior of the excitatory interneurons is not well supported by experiments where wind-up is mostly observed in the projection neurons [28]. Because wind-up in the excitatory interneurons had not been reported by 2010, C fiber presynaptic facilitation was discarded as a possible mechanistic explanation for wind-up in the modeling work done by [1]. Nonetheless, the authors note that there is a possibility for underestimating the extent of wind-up in interneurons because they are smaller in size than projection neurons and, therefore, more difficult to sample for electrophysiology experiments than projection neurons [1].

We note here that the proposed mechanism we simulate in Fig. 7 involves changing the profile of the model input (in Fig. 3) which leads to an obvious change in the model output. We discuss implementations of dynamic wind-up mechanisms in our conclusions in Sect. 5.

3.3 Neuropathy

Following model creation, we next set out to determine whether changes in the balance of excitation and inhibition within spinal cord pain circuitry could explain changes in pain processing under pathological conditions. We were particularly interested in the case of inflammation and neuropathic pain, in which non-noxious mechanical stimuli become painful following peripheral nerve injury. Both nerve injury and inflammation can cause a deregulation of chloride ion transporters in the dorsal horn. Maintaining a low intracellular chloride concentration is important for the functioning of inhibitory neurotransmission. Under typical conditions, the neurotransmitter GABA produces an inhibitory postsynaptic response by binding to the $GABA_A$ receptor, which allows negatively charged chloride ions to flow into the postsynaptic neuron, thus producing hyperpolarization. Under neuropathic or inflammatory conditions, intracellular chloride concentrations may stay semi-permanently elevated, allowing chloride ions to flow out of the cell in response to $GABA_A$ receptor activity, producing excitatory rather than inhibitory effects. Several authors have hypothesized that this deregulation of spinal pain inhibition could explain the development of pain sensation in response to non-noxious stimuli under neuropathic conditions [10, 32].

Neuropathic conditions are characterized by an 8–12 h shift in the phasing of daily rhythms in pain sensitivity [8, 15, 29, 35]. As an application of the model, we investigate whether a large phase shift could be produced by a combination of deregulated neural inhibition and differentially phased rhythmic afferent input from $A\beta$ and C fibers [23], see [9].

Several inflammatory pain conditions, such as osteoarthritis and rheumatoid arthritis, have been shown to exhibit circadian rhythm in pain, with the peak of pain intensity being felt during the night; see Fig. 1 in [8]. Neuropathic pain occurs from various conditions involving the brain, spinal cord, and nerves. It is distinguished from inflammatory conditions, like arthritis, in that it often appears in body parts that are otherwise normal under inspection and imaging, and is also characterized by pain being evoked by a light touch. Experiments on pain in neuropathic patients suggest that neuropathic pain has a circadian rhythm as well, having its peak in the afternoon; see Fig. 2 in [8]. An afternoon peak in pain sensitivity is opposite of the daily rhythm in pain sensitivity under normal conditions [9]. We use our model to further investigate this phenomenon and propose that a possible mechanism for this shift in rhythm is due to the interaction between the $A\beta$ and C fibers.

It has been seen experimentally that the A fiber activity can have an inhibitory influence on C fibers, and that under neuropathic conditions, this inhibition can turn to excitation [10, 31]. Using both of these experimentally observed results, as well as the idea that the circadian rhythm comes into the dorsal horn at the level of the fiber inputs, we show that we can get a change in phase of the firing rate of the projection neurons with a change from inhibition (normal conditions) to excitation (neuropathic conditions) in the influence from the A fibers to the C fibers. In order to test our hypothesis that under neuropathy, response to acute phasic pain peaks in the

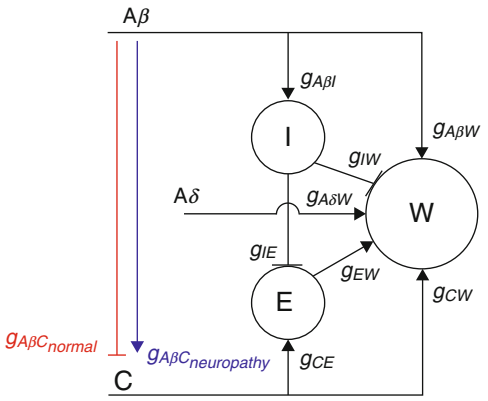


Fig. 8 Diagram of our biophysical model for the dorsal horn circuit under (*blue*) neuropathic and (*red*) normal conditions

late afternoon when, under normal conditions, pain sensation reaches its minimum value, we introduce two principal modifications to our model in Eq. (1). First, we impose a circadian rhythm on the maximum amplitude of the model inputs from Aβ and C fibers (i.e., on parameters $\max_{AB}$ and $\max_{C}$, respectively). Second, we assume an amplitude modulation of the C fibers by the Aβ fibers (see Fig. 8).

Our motivation for the second assumption comes from experimental data suggesting that A fibers can decrease the activity of C nociceptors [31]. To represent such an inhibitory effect of A fibers, we model the amplitude modulation between the Aβ and C fibers with a weight $g_{A\beta C}$, which under normal conditions is inhibitory and $g_{A\beta C} < 0$, whereas under neuropathy, the inhibitory interneuron population through Aβ fibers has an excitatory effect on the C fibers and $g_{A\beta C} > 0$. By simulating our modified model, we investigate whether the activity of the projection neurons follows the circadian rhythm in C fibers under normal conditions and that of the Aβ fibers under neuropathic conditions.

Earlier work suggests circadian rhythmicity in both the touch and pain sensitivity (see Figs. 1 and 2 in [23]). Namely, the pain sensitivity is at its lowest in the early afternoon and at its highest in night, while the highest sensitivity for tactile discrimination is reached in the late afternoon and the lowest in the late morning [23]. These experimental observations motivate us to introduce a circadian rhythm to the model input from $A\beta$ fibers that is in antiphase with the circadian rhythm of the C fiber model inputs, while keeping the arrival times from the three afferent fibers at their default values (see Fig. 3 and Table 1). Thus, in our modified model for neuropathy, $\max_{AB} = 1.5(1+\sin\left(\frac{\pi}{4}t-\pi\right))$ and $\max_{C} = 1.3(1+\sin\left(\frac{\pi}{4}t\right))+1.6$ (see blue and red curves, respectively, in Fig. 9). In addition, because of the synaptic connection from the inhibitory interneuron populations to C fibers, we compute the effective maximum height of the C fiber model input as

$$\max_{C_{eff}} = \max_{C} + g_{A\beta C}\max_{AB}\,, \tag{5}$$

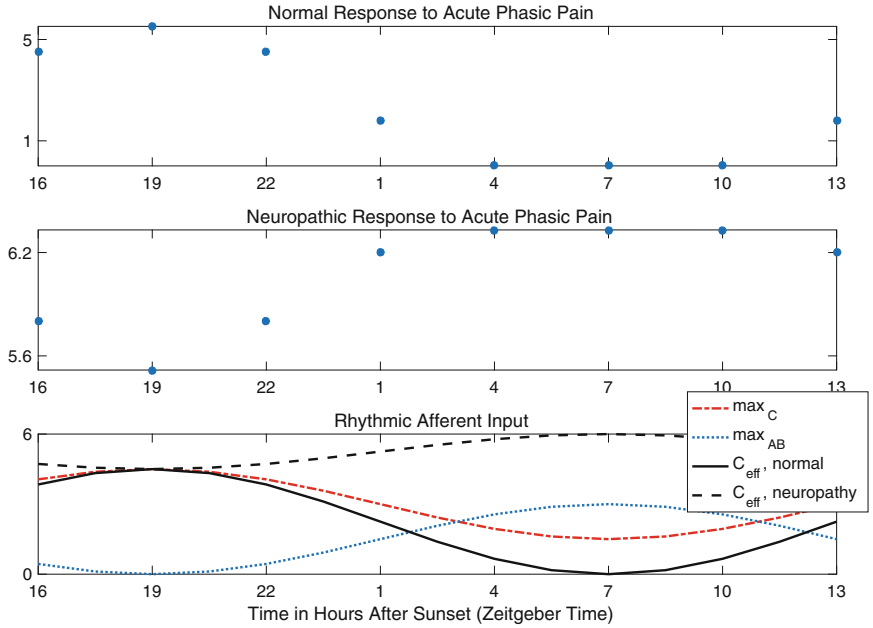


Fig. 9 (*Circles*) Integrated projection neuron activity predicted by the model in (1) under (*top*) normal and (*middle*) neuropathic conditions, and (*bottom*) the different circadian rhythms of the height of the afferent fiber model inputs. Under normal conditions, the interneuron population I decreases the height of the model input from C fibers (see *black solid curve in the bottom panel*), whereas under neuropathic conditions the connection between I and C fiber model input is excitatory. As a result, the effective height of the C fiber model input (*black dashed curve*) is higher than its baseline value (*red curve*). We calculate the activity of the projection neuron population as the area under the C-response. Thus, for each zeitgeber time point (with the corresponding maximum heights of the model inputs from Aβ and C fibers shown in the *bottom panel*), we simulate the neuropathy model for 0.4s and determine the integral under the projection neuron response for $t = [0.007s, 0.4s]$

where $g_{A\beta C} = -0.5$ under normal conditions, and $g_{A\beta C} = 1.5$ under neuropathic conditions. In order to assess the extent of experienced pain, we compute the integral of the firing rate of the projection neurons to a stimulus in C fibers (i.e., painful stimuli) using the trapezoidal method from $t > 0.07s$ onward (i.e., the C-response of W; see model output curve in Fig. 5). Indeed, our model simulations suggest that inhibition turned excitation at the level of the fibers is a possible mechanistic explanation for the flip in phase of pain sensitivity seen under neuropathic conditions. Our model shows that under normal conditions, the pain sensitivity rhythm follows the circadian rhythm of the C fibers (see top and bottom panels in Fig. 9) but mimics the rhythm in the Aβ fibers under neuropathic conditions (see middle and bottom panels in Fig. 9).

We note here that under neuropathic conditions, the firing rate of the projection neurons is at, or near, its maximum value throughout the day. While patients with

neuropathic pain experience an increased level of pain throughout the day than those without, it is unrealistic for the pain to be at, or near, its maximum all day. In the next section, we propose an amended model in which we include top-down inhibition from the mid-brain where this is not the case.

4 Model with Descending Control from the Mid-Brain

4.1 Introduction

In its current form, our biophysical model of the dorsal horn pain circuit includes response functions for the three neuron populations that mimic empirical observations. Importantly, our model reproduces the phenomenon of pain inhibition in which a brief mechanical stimulus applied after a painful stimulus can decrease the activity of the projection neurons and, thereby, decrease the sensation of pain. Our model also captures the phase shift in pain intensity for neuropathic pain; however, the amplitude of the pain intensity under neuropathic conditions is very small, and the firing rate of the projections neurons is basically at its maximum throughout the day. In an attempt to both make the model more realistic and explain the neuropathic phase flip, we introduce an amended model in which we consider communication from the dorsal horn to the mid-brain (see Fig. 10). Influence from the mid-brain to the dorsal horn plays an important role in modulating inhibition within the pain circuit of the dorsal horn [5, 21, 35]. There are several descending pathways from the brain down to the spinal cord that could affect the afferent fibers, the inhibitory interneurons, and the projection neurons; see Fig. 2 in [21]. We choose to model the inhibitory descending pathway, as done in [5]. The motivation for this added

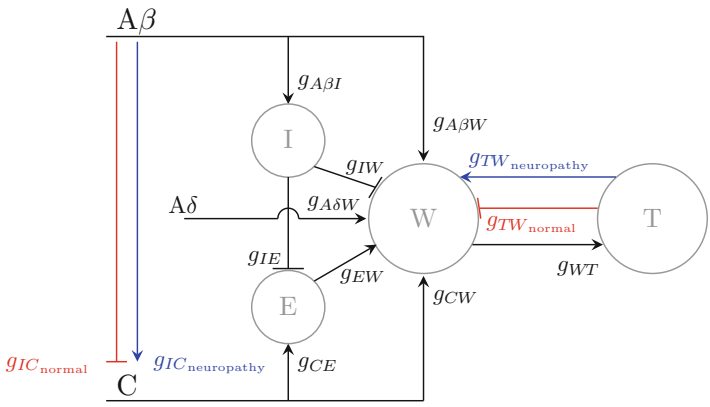


Fig. 10 Diagram of our biophysical model for the dorsal horn circuit including connections to and from the mid-brain under (*blue*) neuropathic and (*red*) normal conditions

mechanism is to enable the projection neurons to exhibit a more realistic phase flip in pain sensitivity throughout the day under neuropathic conditions. We also aim to include a mechanism for the effect of the homeostatic sleep drive on this top-down inhibition. As shown in [14] and [30], the build-up of the homeostatic sleep drive is reflected in the daily rhythm of the pro-inflammatory cytokines, whose increased levels are associated with an increased firing rate of the WDR projection neurons. We model this by assuming that the connection from the mid-brain to the dorsal horn circuit is a function of the time spent awake, or the build-up of the homeostatic sleep drive. We verify this amended model by showing that it can reproduce the same phenomena as the earlier model, as well as show that this amended model can better capture the observed change in phase of pain sensitivity rhythm for neuropathic patients.

4.2 Amendments to Model

In our modified biophysical model for pain, we add a connection between the projection neurons and the mid-brain as shown in Fig. 10 and a dimension to the mathematical model in Eq. (1). Thus, the dynamics for the average firing rate of neuron populations I and E remain as they are in Eq. (1), and the equations for the projection neurons W and neuron population in the mid-brain (T) become as follows:

$$\frac{df_W}{dt} = \frac{W_\infty(g_{ABW}f_{AB} + g_{ADW}f_{AD} + g_{CW}f_C + g_{EW}f_E - g_{IW}f_I - g_{TW}(H)f_T) - f_W}{\tau_W},$$

$$\frac{df_T}{dt} = \frac{T_\infty(g_{WT}f_W) - f_T}{\tau_T}, \quad (6)$$

where τ_T is the intrinsic timescale of the population T and weight g_{WT} denotes the strength of the effect a change in the W population has on the neuron population T. To investigate the coupling between neuropathic pain and sleep deprivation, we allow g_{WT}, the strength of the effect of a change in the T population on the W population, to depend on the homeostatic sleep drive H. Hence, we write the weight of the connection from T to W as $g_{TW}(H)$. As in the case of the other neuron populations in Sect. 2.2, we assume a monotonically increasing firing rate response function (with respect to input c) for the mid-brain population T:

$$T_\infty(c) = \max_T \frac{1}{2}\left(1 + \tanh\left(\frac{1}{\alpha_T}(c - \beta_T)\right)\right), \quad (7)$$

where $\max_T$ is the maximum firing rate of the mid-brain population, $c = \beta_T$ is the input at which the average firing rate of the mid-brain population reaches half of its maximum value, and $1/\alpha_T$ determines the slope of the transition from non-firing to firing in mid-brain population.

4.3 Model Validation

In regard to parameter values, we introduce a lag in the connection from the mid-brain to the projection neurons, and choose $\tau_T = 0.05$ ms, which is larger than the intrinsic timescales of the other populations (see Table 1). We do this as a result of the assumption that the signal must travel much further to interact with the mid-brain than it would to other populations within the dorsal horn circuit. In addition, we assume that the maximum amplitude and the slope of the response function of T are smaller than those of the other neuron populations of the model and we pick $(\max_T, \alpha_T, \beta_T) = (30, 0.75, 1.4)$ (see Fig. 11). As in the beginning of Sect. 3, we choose values for the weights $g_{WT} = 0.1$ and $g_{TW} = 0.05$ using the model output of the average firing rates of the four neuron populations to the model input (shown in Fig. 3) as our point of comparison (see Fig. 12).

The modified model including connections to and from the mid-brain can capture the delayed inhibition response in the projection neurons from delayed stimulation of the Aβ fibers (see Fig. 13).

As concerns neuropathy, we use a similar approach as in Sect. 3.3 and consider both (a) an inhibitory effect of the A fibers to the C fibers represented by amplitude modulation of the C fibers by the Aβ fibers as given in Eq. (5) and (b) circadian rhythm in the C and Aβ fibers. In addition, we use the amended model to investigate the hypothesis that under neuropathy, time spent awake causes increased excitatory input from the mid-brain to the dorsal horn circuit [14]. Thus, we assume that the strength of the connection from the mid-brain population T to the projection neuron population W given by the weight g_{TW} (see Fig. 10) increases during wake and decreases during sleep. Moreover, we assume that under normal conditions, the T population inhibits the activity of the W population, while under neuropathy, the connection from T to W is excitatory (see red and blue lines in Fig. 10). Thus, under normal conditions, the weight g_{TW} has a daily rhythm shown in red, and

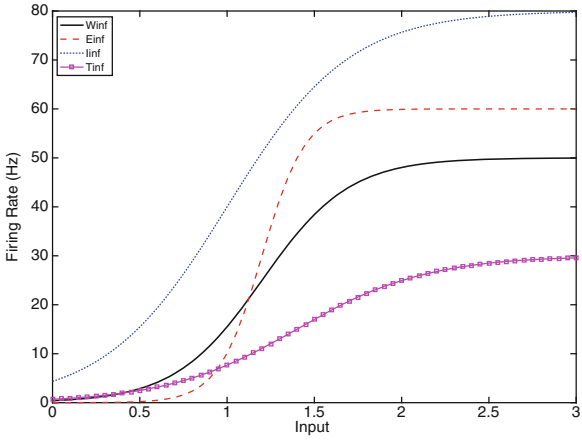


Fig. 11 Response functions of the projection (*black*), excitatory (*red*), inhibitory (*blue*), and mid-brain (*magenta*) neuron populations for different constant inputs (on the x-axis)

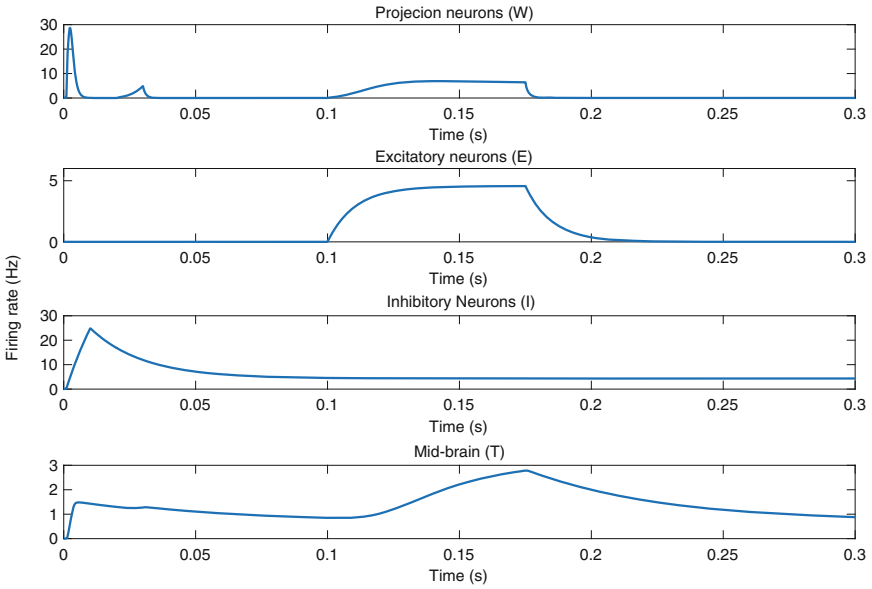


Fig. 12 Firing rates for each population (connected as described in Fig. 10) in response to the input from afferent fibers as described in Fig. 3 including the amendments to the dorsal horn circuit model given in Eq. (6)

under neuropathic conditions, it has the rhythm shown in blue in Fig. 14, where negative values result in excitatory input from T to W because of our choice of using a negative sign in front of g_{TW} in Eq. (6).

With these abovementioned modifications to the model of the dorsal horn circuit, our amended model reproduces a more pronounced flip in the phase of the W population (see Fig. 15) than in the case of only amplitude modulation of the C fibers through Aβ fibers (see Fig. 9). We note as well that the W population does not saturate to its maximum firing rate in Fig. 15 (right), as it did in the earlier model in Fig. 9.

5 Conclusions and Future Work

We have constructed a biophysical model of the pain processing circuit in the dorsal horn that represents the interactions between inhibitory and excitatory interneurons, and WDR projection neurons. Our model considers the average firing rate of each of these three neuron populations and therefore includes less biophysical detail than previous circuit models consisting of single spiking neurons. However, our choice of modeling framework is motivated by our ongoing work to incorporate this model for the pain processing circuit with sleep–wake regulatory network models (see [4] for

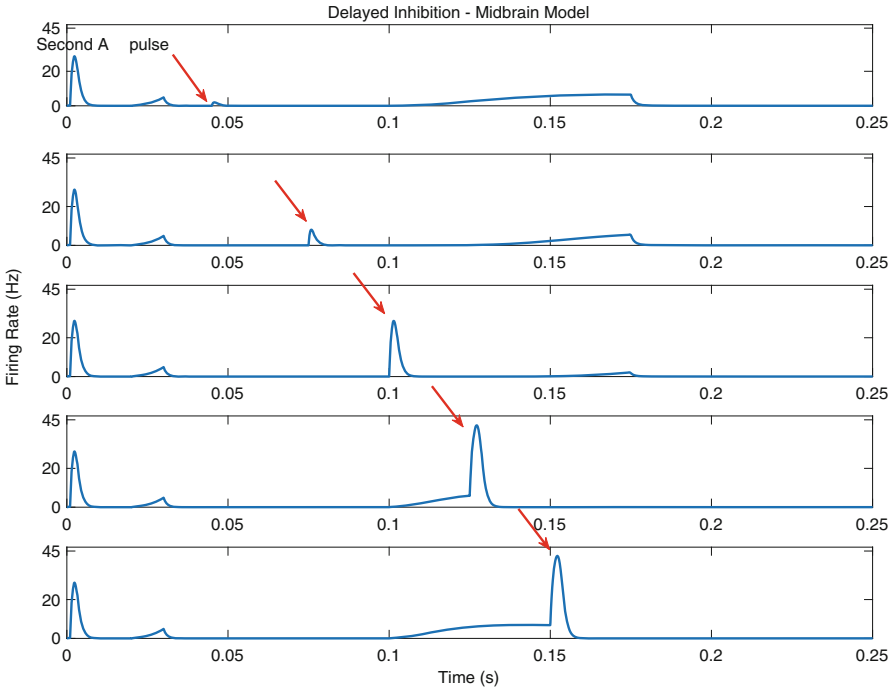


Fig. 13 Pain inhibition phenomenon captured in the modified model including top-down inhibition. Response of the projection neuron population to the initial fiber pulse stimulation (at $t = 0$) and the second pulse stimulation only to Aβ fibers (*red arrows*) for increasing in time between fiber stimulations

a review). Such an extended sleep–wake–pain model would allow us to test several existing hypotheses on the effects of sleep-dependent and circadian modulation of pain sensitivity. In addition, we have chosen to use a simplified modeling approach, because it allows us to examine whether suggested mechanisms (i.e., rhythmicity in afferent fibers and their interaction changing from inhibitory to excitatory under neuropathic conditions) are capable of explaining observed rhythmicity in pain before we incorporate more physiological details into our model.

Concerning the phenomenon of wind-up, we simulate it as an increase in the synaptic efficacy of the C fibers before their input reaches the deep dorsal horn circuit. This is an assumption that is supported by experimental evidence of wind-up in the potential of the C fibers observed in the superficial but not in the deep dorsal horn [27]. However, experimental data also suggest that wind-up is more pronounced in the deep than in the superficial dorsal horn neurons [28]. Therefore, our model assumption of wind-up occurring only upstream from the deep dorsal horn is not widely supported by the data. Moreover, at the current stage, our model incorporates no information on possible mechanistic explanations of wind-up. This

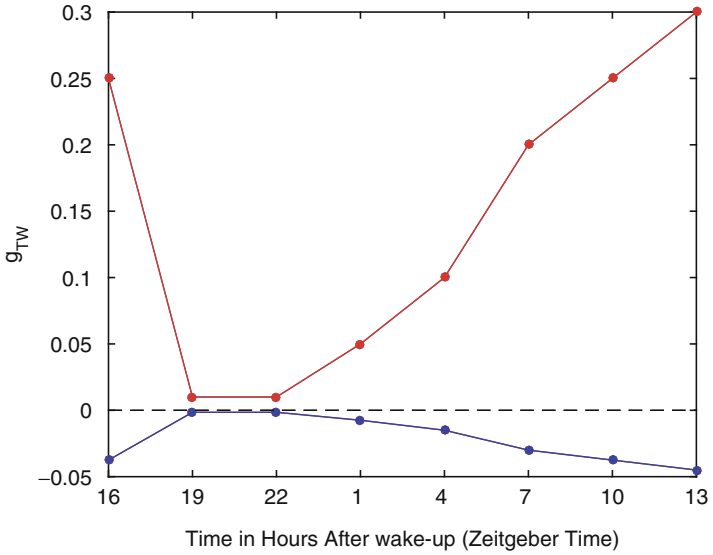


Fig. 14 The strength of the connection from the mid-brain to the projection neuron population as a function of hours after wake-up under (*blue*) neuropathic and (*red*) normal conditions. We note that negative values of g_{TW} result in excitatory input from T to W population; see Eq. (6)

is a limitation of the model as it restricts our ability to test existing hypotheses (see Fig. 6 in [11]) and increase the current understanding of the generation of wind-up.

As concerns neuropathy, by taking into account both amplitude modulation of the C fibers by the A fibers and normal inhibitory effect switching to excitatory under neuropathic conditions, our model reproduces a change in the daily rhythm seen in the activity of the WDR projection neurons and predicts a higher baseline of pain under neuropathy than under normal conditions, both of which agree with experimental evidence. In the case where a connection to, and from, the mid-brain is included in the dorsal horn circuitry, the flip of the rhythm in the projection neurons is more pronounced and does not evoke a response in the WDR neurons that is

Fig. 15 (continued) Integrated (*circles, top 2 panels*) and maximum (*asterisks, bottom 2 panels*) projection neuron firing rates predicted by the amended equations of time evolution given in (6) under normal (*top and 4th panels*) and neuropathic (*2nd and bottom panels*) conditions showing predicted circadian modulation over 24 h. Circadian rhythmicity of responses is generated by the different circadian rhythms in the amplitudes of the afferent fiber model inputs (*middle panel*). Under normal conditions, the interneuron population I decreases the amplitude of the model input from C fibers (*red curve*) leading to a reduced effective amplitude of C fiber input (*black solid curve*). Under neuropathic conditions, the connection between I and C fiber model input is excitatory resulting in a higher effective amplitude of the C fiber model input (*black dashed curve*). We calculate the activity of the projection neuron population as the area under the C-response in the same way as in Fig. 9

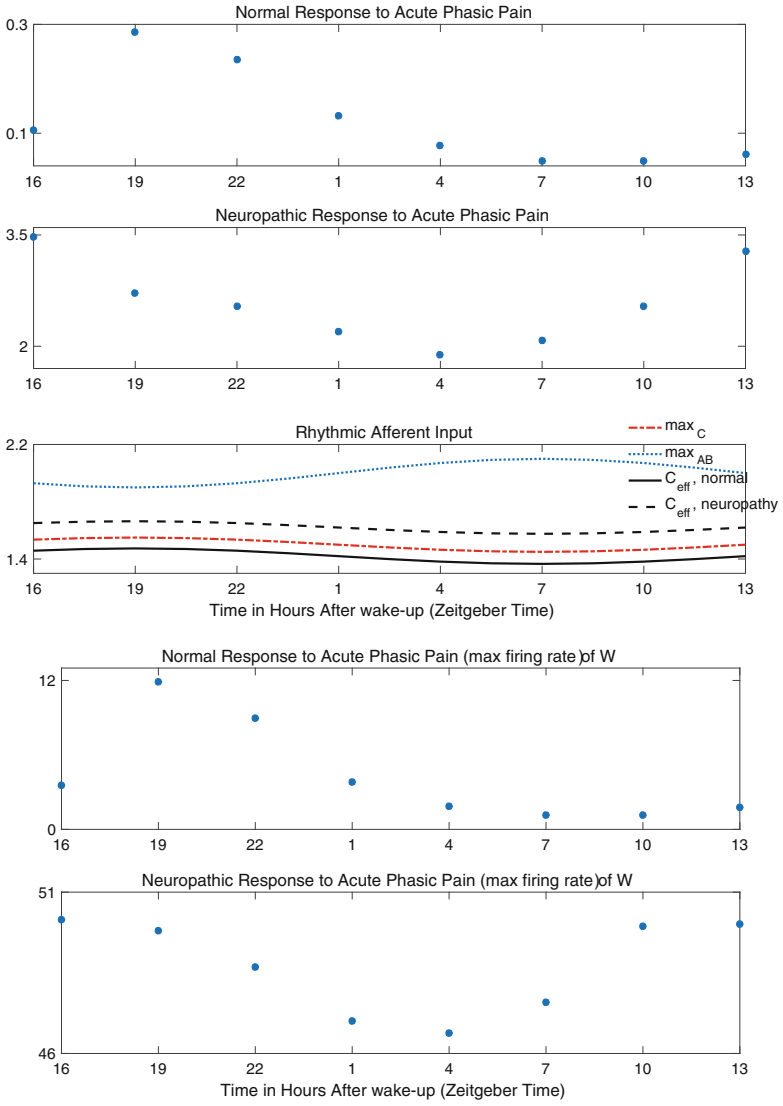


Fig. 15 (continued)

at its maximum as in the case where there is no connection to the mid-brain. In our ongoing work, we investigate the robustness of these abovementioned model predictions for neuropathy, in particular as concerns the range of parameter values that represent the strength of the connection between the WDR projection neurons and the mid-brain.

We have incorporated neuropathy in an attempt to validate that our model can replicate known circadian pain effects. It is important to note that a more biologically realistic model has been developed including large networks of individual neurons [2]. Similarities between our model and the one proposed by Arle et al. are the lack of connection from the Aβ fibers to the inhibitory interneurons, but a major difference is that their model has two distinct circuits for nociceptive and neuropathic pain. We instead use the same circuit but propose different mechanisms within the circuit that contribute to neuropathic pain (e.g., inhibition switching to excitation under neuropathic conditions). We justify our use of a simplified model by emphasizing that our motivation is in understanding the effect of circadian and sleep-dependent processes on pain sensitivity and note that our model does capture circadian effects in neuropathic pain patients.

In our ongoing work, we are constructing equations for the time evolution of the average activity of each of the three afferent fiber populations. Such a generalization would not only increase our ability to include possible mechanisms of wind-up but also help in connecting models of pain and sleep together. That is, in the future generalized model, a change in the homeostatic sleep drive (that is an output of the sleep–wake model) could be directly fed into the pain circuit model by influencing the sensitivity of the afferent fibers to external stimuli. This will allow us to more thoroughly investigate several hypotheses on the coupling between sleep deprivation and pain sensitivity.

Acknowledgements This work was conducted as a part of A Research Collaboration Workshop for Women in Mathematical Biology at the National Institute for Mathematical and Biological Synthesis, sponsored by the National Science Foundation through NSF Award DBI-1300426, with additional support from the University of Tennessee, Knoxville. This work was additionally partially supported by the following sources: NSF Award DMS-1412119 (VB) and the Pritzker Neuropsychiatric Disorders Research Consortium (MH). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Aguiar, P., Sousa, M., Lima, D.: NMDA channels together with L-type calcium currents and calcium-activated nonspecific cationic currents are sufficient to generate windup in WDR neurons. J. Neurophysiol. **104**(2), 1155–1166 (2010)
2. Arle, J.E., Carlson, K.W., Mei, L., Iftimia, N., Shils, J.L.: Mechanism of dorsal column stimulation to treat neuropathic but not nociceptive pain: analysis with a computational model. Neuromodulation Technol. Neural Interface **17**(7), 642–655 (2014)
3. Basbaum, A.I., Bautista, D.M. Scherrer, G., Julius, D.: Cellular and molecular mechanisms of pain. Cell **139**(2), 267–284 (2009)

4. Booth, V., Diniz Behn, C.G.: Physiologically-based modeling of sleep-wake regulatory networks. Math. Biosci. **250**, 54–68 (2014)
5. Britton, N.F., Chaplain, M.A., Skevington, S.M.: The role of N-methyl-D-aspartate (NMDA) receptors in wind-up: a mathematical model. IMA J. Math. Appl. Med. Biol. **13**(3), 193–205 (1996)
6. Britton, N.F., Skevington, S.M.: A mathematical model of the gate control theory of pain. J. Theor. Biol. **137**(1), 91–105 (1989)
7. Foo, H., Mason, P.: Brainstem modulation of pain during sleep and waking. Sleep Med. Rev. **7**(2), 145–154 (2003)
8. Gilron, I., Ghasemlou, N.: Chronobiology of chronic pain: focus on diurnal rhythmicity of neuropathic pain. Curr. Opin. Support. Palliat. Care **8**(4), 429–436 (2014)
9. Hagenauer, M.H., Kile, J., Piltz, S., Toporikova, N., Ferguson, P., Booth, V.: The modulation of pain by circadian and sleep-dependent processes: a review of the experimental evidence. In: Proceedings from A Research Collaboration Workshop for Women in Mathematical Biology. Springer, Berlin (2016)
10. Hasbargen, T., Ahmed, M.M., Miranpuri, G., Li, L., Kahle, K.T., Resnick, D., Sun, D.: Role of NKCC1 and KCC2 in the development of chronic neuropathic pain following spinal cord injury. Ann. N. Y. Acad. Sci. **1198**, 168–172 (2010)
11. Herrero, J.F., Laird, J.M., Lopez-Garcia, J.A.: Wind-up of spinal cord neurones and pain sensation: much ado about something? Prog. Neurobiol. **61**(2), 169–203 (2000)
12. Hines, M.L., Carnevale, N.T.: NEURON: a tool for neuroscientists. Neuroscientist **7**(2), 123–135 (2001)
13. Hodgkin, A.L., Huxley, A.F.: A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. (Lond.) **117**(4), 500–544 (1952)
14. Huang, C.-T., Chiang, R.P.-Y., Chen, C.-L., Tsai, Y.-J.: Sleep deprivation aggravates median nerve injury-induced neuropathic pain and enhances microglial activation by suppressing melatonin secretion. Sleep **37**(9), 1513–1523 (2014)
15. Kusunore, N., Koyanagi, S., Hamamura, K., Matsunaga, N., Yoshida, M., Uchida, T., Tsuda, M., Inoue, K., Ohdo, S.: Molecular basis for the dosing time-dependency of anti-allodynic effects of gabapentin in a mouse model of neuropathic pain. Mol. Pain **6**(83), 1–8 (2010)
16. Le Bars, D., Gozariu, M., Cadden, S.W.: Animal models of nociception. Pharm. Rev. **53**(4), 597–652 (2001)
17. Le Franc, Y., Le Masson, G.: Multiple firing patterns in deep dorsal horn neurons of the spinal cord: computational analysis of mechanisms and functional implications. J. Neurophysiol. **104**(4), 1978–1996 (2010)
18. Melnick, I.V., Santos, S.F., Szokol, K., Szucs, P., Safronov, B.V.: Ionic basis of tonic firing in spinal substantia gelatinosa neurons of rat. J. Neurophysiol. **91**(2), 646–655 (2004)
19. Melzack, R., Wall, P.D.: Pain mechanisms: a new theory. Science **150**(3699), 971–979 (1965)
20. Mendell, L.M. Wall, D.P.: Responses of single dorsal cells to peripheral cutaneous unmyelinated fibers. Nature **206**, 97–99 (1965)
21. Millan, M.J.: Descending control of pain. Prog. Neurobiol. **66**(6), 355–474 (2002)
22. Moayedi, M., Davis, K.D.: Theories of pain: from specificity to gate control. J. Neurophysiol. **109**(1), 5–12 (2013)
23. Pöllmann, L.: Duality of pain demonstrated by the circadian variation in tooth sensitivity. In: Erhard, H., Kabat, H.F. (eds.) Chronobiology 1982–1983, chap. 39, pp. 225–228. Karger, Basel (1984)
24. Purves, D., Augustine, G.J., Fitzpatrick, D., et al.: Neuroscience. Sinauer Associates, Sunderland (2001)
25. Reeve, A.J., Walker, K., Urban, L., Fox, A.: Excitatory effects of galanin in the spinal cord of intact, anaesthetized rats. Neurosci. Lett. **295**(1–2), 25–28 (2000)
26. Ruscheweyh, R., Sandkühler, J.: Lamina-specific membrane and discharge properties of rat spinal dorsal horn neurones in vitro. J. Physiol. (Lond.) **541**(Pt 1), 231–244 (2002)
27. Schouenborg, J.: Functional and topographical properties of field potentials evoked in rat dorsal horn by cutaneous C-fibre stimulation. J. Physiol. (Lond.) **356**, 169–192 (1984)

28. Schouenborg, J., Sjölund, B.H.: Activity evoked by A- and C-afferent fibers in rat dorsal horn neurons and its relation to a flexion reflex. J. Neurophysiol. **50**(5), 1108–1121 (1983)
29. Takada, T., Yamashita, A., Date, A., Yanase, M., Suhara, Y., Hamada, A., Sakai, H., Ikegami, D., Iseki, M., Inada, E., Narita, M.: Changes in the circadian rhythm of mRNA expression for μ-opioid receptors in the periaqueductal gray under a neuropathic pain-like state. Synapse **67**(5), 216–223 (2013)
30. Vgontzas, A.N., Bixler, E.O. Lin, H.-M., Prolo, P., Trakada, G., Chrousos, G.P.: Il-6 and its circadian secretion in humans. Neuroimmunomodulation **12**(3), 131–140 (2005)
31. Wan-Ru, D., Yi-Kuan, X.: Modulation of c-nociceptive activities by inputs from myelinated fibers. Adv. Exp. Med. Biol. **904**, 33–40 (2016)
32. Wei, H., Hao, B., Huang, J.-L., Ma, A.-N., Li, X.-Y., Wang, Y.-X., Pertovaara, A.: Intrathecal administration of a gap junction decoupler, an inhibitor of Na+ K+ 2Cl cotransporter 1, or a GABAA receptor agonist attenuates mechanical pain hypersensitivity induced by REM sleep deprivation in the rat. Pharmacol. Biochem. Behav. **97**(2), 377–383 (2010)
33. Wilson, H.R., Cowan, J.D.: Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. **12**(1), 1–24 (1972)
34. Woolf, C.J., Wall, P.D.: Chronic peripheral nerve section diminishes the primary afferent A-fibre mediated inhibition of rat dorsal horn neurones. Brain Res. **242**(1), 77–85 (1982)
35. Zhang, J., Li, H., Teng, H., Zhang, T., Luo, Y., Zhao, M., Li, Y.Q., Sun, Z.S.: Regulation of peripheral clock to oscillation of substance P contributes to circadian inflammatory pain. Anesthesiology **117**(1), 149–160 (2012)
36. Zhang, T.C., Janik, J.J. Grill, W.M.: Modeling effects of spinal cord stimulation on wide-dynamic range dorsal horn neurons: influence of stimulation frequency and GABAergic inhibition. J. Neurophysiol. **112**(3), 552–567 (2014)

A Two-Process Model for Circadian and Sleep-Dependent Modulation of Pain Sensitivity

Natalia Toporikova, Megan Hastings Hagenauer, Paige Ferguson, and Victoria Booth

Abstract Pain sensitivity is strongly modulated by time of day and by prior sleep behavior. These two factors, governed by the circadian rhythm and homeostatic sleep drive, respectively, likewise dictate the timing and duration of sleep. The fields of sleep and circadian research have identified much of the physiology underlying the circadian rhythm and homeostatic sleep drive with mathematical modeling playing an important role in understanding how these two processes interact to affect sleep behavior. We hypothesize that the daily rhythm of pain sensitivity and its sleep-dependent modulation reflect an interaction of the circadian rhythm and homeostatic sleep drive. To investigate this hypothesis, we adapt the formalism of a classic mathematical model for the regulation of sleep behavior by the circadian rhythm and homeostatic sleep drive, called the Two-Process model, to simulate the interaction of these two processes on pain sensitivity. To construct the model, we utilize data from experimental reports on the daily rhythmicity of pain sensitivity in humans to define a "daily pain sensitivity" function. We decompose this function into two processes: a sleep-dependent process $S(t)$ that follows the homeostatic sleep drive and a circadian process $C(t)$ that is dictated by the circadian rhythm. By simulating different sleep schedules with the original Two-Process model, we compute changes in the sleep-dependent process $S(t)$ that modulates pain sensitivity. By combining $S(t)$ with the circadian process $C(t)$, our model predicts resultant

N. Toporikova
Biology Department, Washington and Lee University, Lexington, VA 24450, USA
e-mail: toporikovan@wlu.edu

M.H. Hagenauer
Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: hagenaue@umich.edu

P. Ferguson
Department of Biological Sciences, University of Alabama, Tuscaloosa, AL 35487, USA
e-mail: pfferguson@ua.edu

V. Booth (✉)
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA

Department of Anesthesiology, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: vbooth@umich.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_3

changes in the daily pain sensitivity rhythm. We illustrate model predictions for changes in pain sensitivity due to sleep deprivation, sleep restriction and shift work schedules. We believe that this model may be a useful tool for pain management by providing predictions of the variations in pain sensitivity due to changing sleep schedules.

MSC codes: 39A23, 39A60, 92B25, 92C30

1 Introduction

Pain sensitivity is strongly modulated by time of day and by prior sleep behavior. As reviewed in the preceding chapter [7], in humans highest sensitivity to painful stimuli occurs during the night and lowest sensitivity occurs in the late afternoon. Additionally, sleep deprivation increases pain sensitivity. Experimental studies of pain sensitivity have only considered either time-of-day (circadian) effects or sleep-dependent effects. However, it is well known that the effects of prior sleep behavior and the circadian rhythmicity of sleep propensity interact to govern the timing and duration of sleep. For example, sleep deprivation causes an increase in the drive for sleep and can promote the occurrence of sleep during daytime hours. However, the circadian rhythm acts to promote wakefulness during the day. The interaction of these two processes, namely the homeostatic sleep drive and the circadian rhythm, results in limited durations of daytime sleep episodes despite elevated sleep drive levels which would prolong sleep if it occurred during the evening.

We hypothesize that the 24 h and sleep-dependent modulation of pain sensitivity may likewise reflect a combined interaction of these two processes: circadian rhythm and homeostatic sleep drive. As an example, consider a study on the effects of sleep deprivation on pain sensitivity. If pain measurements are conducted in the late afternoon, effects of sleep deprivation may be underestimated, while they may be overestimated if measurements are taken during the night because of circadian modulation of pain sensitivity. To investigate this hypothesis, we adapt the formalism of a classic and influential mathematical model for the regulation of sleep behavior by the circadian rhythm and homeostatic sleep drive, called the Two-Process model [3], to model how the interaction of these two processes may affect pain sensitivity.

2 Background: Two-Process Model for Circadian Modulation of Sleep Timing

The original Two-Process model for sleep regulation was constructed to account for the interaction of the homeostatic sleep drive and the circadian rhythm in the

timing and duration of sleep [3]. While the physiology of the homeostatic sleep drive has not been completely determined [6, 9, 12], a biological marker for it has been identified as the power of low frequency, delta range ($0.5 - 4$ Hz) oscillations in EEG recordings during human sleep in the non-rapid eye movement (nREM) or slow wave stage of sleep. Specifically, at the initiation of a sleep episode, the power of delta oscillations is high and decays roughly exponentially as sleep continues through the night. These observations motivated modeling the homeostatic sleep drive, Process S $\mathscr{S}(t)$ in the model, as an exponential function that decreases during sleep and increases during wake. The time constant for the decay during sleep was fit to the decay in delta oscillation power observed in sleep EEG recordings. The time constant for the increase of the homeostatic sleep drive during wake was constrained to match the fit to data during sleep. Specifically, Process S increases exponentially during wake with time constant $\tau_w = 18.2$ h and decreases exponentially during sleep with time constant $\tau_s = 4.2$ h. In wake, $\mathscr{S}(t)$ is governed by:

$$\mathscr{S}_w(t) = 1 + (\mathscr{S}_{wo} - 1)e^{\frac{-(t_{wo}-t)}{\tau_w}} \tag{1}$$

where time t is in hours, $\mathscr{S}_{wo}$ is set to the $\mathscr{S}$ value at the previous wake onset (wo), and t_{wo} is the time of the previous wake onset. In sleep, $\mathscr{S}(t)$ is governed by:

$$\mathscr{S}_s(t) = \mathscr{S}_{so}e^{\frac{(t_{so}-t)}{\tau_s}} \tag{2}$$

where $\mathscr{S}_{so}$ is set to the $\mathscr{S}$ value at the previous sleep onset (so) and t_{so} is the time of the previous sleep onset.

While more is known about the physiology of the circadian rhythm that modulates sleep timing [5, 10], the Two-Process model focuses on time-of-day effects on sleep propensity to determine the equations for Process C $\mathscr{C}(t)$. Experiments measuring typical sleep latencies and durations across the day suggested that sleep timing follows a skewed sinusoidal function such that sleep is minimal near midday and is strongly promoted in the early morning hours [3]. This circadian rhythmicity in sleep propensity is combined with Process S by Process C dictating the threshold values at which Process S transitions from sleep to wake and vice versa. As such, Process C consists of two sinusoidally varying functions $\mathscr{C}_w(t)$ and $\mathscr{C}_s(t)$ such that $\mathscr{C}_w(t)$ dictates the $\mathscr{S}(t)$ values when the transition from wake to sleep should occur and $\mathscr{C}_s(t)$ dictates the $\mathscr{S}(t)$ values at which sleep to wake transitions occur:

$$\mathscr{C}_w(t) = 0.67 + c(t), \tag{3}$$

$$\mathscr{C}_s(t) = 0.17 + c(t), \tag{4}$$

$$\begin{aligned} c(t) = {} & 0.12[0.97\sin(2\pi(t-t_0)/24) + 0.22\sin(4\pi(t-t_0)/24) + 0.07\sin(6\pi(t-t_0)/24) \\ & +0.03\sin(8\pi(t-t_0)/24) + 0.001\sin(8\pi(t-t_0)/24)] \end{aligned} \tag{5}$$

where t_0 sets the circadian phase at the initial time.

With these parameter values, the model generates a 24 h cycle of sleep–wake behavior with 16 h in wake and 8 h in sleep (Fig. 1). We initialize t, c, and $\mathscr{S}$ so

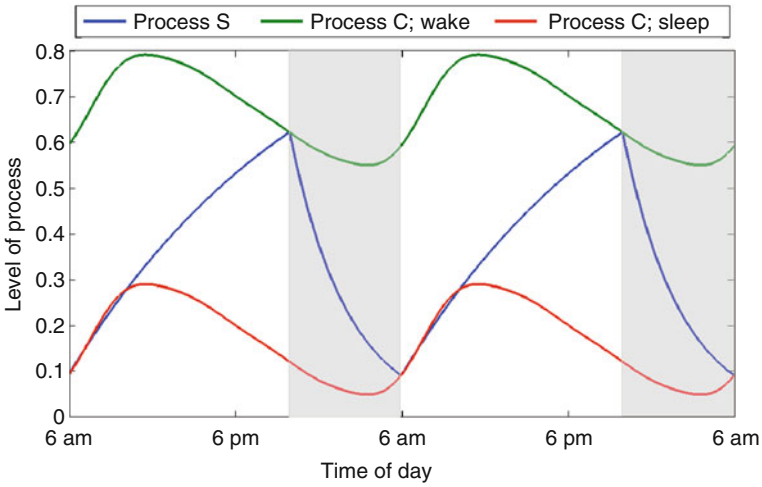


Fig. 1 Two-Process model predicting the timing and duration of sleep and wake episodes as governed by the homeostatic sleep drive (Process S, $\mathscr{S}(t)$, *blue curve*) whose levels for transitions between wake and sleep are dictated by the circadian rhythm (Process C, *green and red curves*). Wake occurs as Process S is increasing with sleep initiated when Process S intersects $\mathscr{C}_w(t)$ (*green curve*). Sleep (shaded regions) continues as Process S decreases until it intersects $\mathscr{C}_s(t)$ (*red curve*)

that the model starts at $t = 6$ for 6 am at the beginning of a wake episode with the following initial conditions: $t = 6$, $t_0 = 7.5$, $\mathscr{S} = \mathscr{S}_{wo} = \mathscr{C}_s(6)$ and $t_{wo} = 6$. With these initial conditions, the model simulates a sleep schedule with wake onset at 6 am and sleep onset at 10pm. The wake state occurs during the interval that $\mathscr{S}(t)$ is increasing and the sleep state occurs when it is decreasing.

3 Two-Process Model for Pain Sensitivity

While there is strong experimental evidence that increased homeostatic sleep drive that occurs due to sleep deprivation increases pain sensitivity (as reviewed in [7]), the correlation between pain sensitivity and the homeostatic sleep drive under normal sleep conditions has not been determined. To construct our model, we assume that pain sensitivity exhibits increases and decreases throughout the 24 h day that follow the increases and decreases of the homeostatic sleep drive. In particular, we assume pain sensitivity increases with time spent awake even under normal daily schedules of sleep–wake behavior. We implement the simplest form of this assumption by assuming that sleep-dependent modulation of pain sensitivity follows the exponential increase during wake and decrease during sleep of the homeostatic sleep drive predicted by the original Two-Process Model. A consequence of this primary assumption is the further assumption that the experimentally observed daily

rhythm of pain sensitivity is the combined result of circadian and sleep-dependent modulation. For our model, we develop an estimate of the sleep-dependent component of this modulation by determining an appropriate scaling factor of Process S from the original Two-Process Model. We then estimate the circadian modulation of pain sensitivity by subtracting the estimated sleep-dependent component from the experimentally observed daily rhythm of pain sensitivity. We assume that this circadian modulatory component is not affected by sleep behavior. To implement the model, we use the Two-Process model to compute variations of the sleep-dependent modulatory component under different patterns of sleep behavior and combine it with the circadian component to predict their combined effect on pain sensitivity with changes in sleep behavior.

As discussed in the companion article [7], multiple studies report a consistent daily rhythm of pain sensitivity that peaks during the night and is at a minimum during the late afternoon. In order to quantify this rhythm, in [7] we constructed a prototypical "daily pain sensitivity function" by normalizing data from four studies that tested experimentally induced pain responses across 24 h. For each dataset, we transformed time to relation to morning wake time and transformed units to percent of the mean of the reported daily variation. Following these transformations, the data collectively formed a tight, sinusoidal curve with a trough $\sim$ 9 h after usual wake onset and a peak $\sim$ 2 h after usual sleep onset (see Fig. 1 in [7]). This constructed curve provides the qualitative shape of the daily rhythm in pain sensitivity, but does not accurately reflect the amplitude of the rhythm due to the units transformation. In order to compare the amplitude of effects across different studies on the daily fluctuation of pain sensitivity using different pain modalities, and additionally to compare effects due to sleep deprivation, in [7] we identified new normalizations for units of change in pain sensitivity. Namely, for the particular pain modality used, we converted the observed changes in pain threshold to a percentage of the range of physiologically meaningful stimulation values or to a percentage of the range of painful stimulation values. Using these units for changes in pain sensitivity, we found that the average amplitude (max - min) of the daily rhythm measured in multiple pain modalities was 12–14% (see Table 6 in [7]). Thus, for our model, we define the experimentally observed daily pain rhythm, $P_{obs}(t)$, as the "daily pain sensitivity function" scaled so that its oscillation amplitude (peak - trough) matches this average amplitude. The appropriate scaling parameters for the "daily pain sensitivity function" were in the range $[0.6, 0.7]$. In the model, this curve is assumed to be the combined result of circadian and sleep-dependent modulation of pain sensitivity and will be decomposed into sleep-dependent and circadian components, $S(t)$ and $C(t)$, respectively.

To explicitly define the model, let $S(t)$ represent the time-varying, sleep-dependent modulation of pain sensitivity that follows the homeostatic sleep drive as predicted by the Two-Process model:

$$S(t) = \sigma(\mathscr{S}(t) - \mathscr{S}_{0min}), \tag{6}$$

where $\mathscr{S}(t)$ is Process S (Eqs. (1) and (2)), $\mathscr{S}_{0min}$ is the minimum value of $\mathscr{S}$ under normal sleep–wake behavior, and σ is an appropriate scaling factor. In this way, during normal sleep–wake behavior predicted by the Two-Process Model, $S(t)$ varies between 0 at times of wake onset to a maximum value at times of sleep onset. During other sleep patterns, such as sleep deprivation or restriction when the sleep homeostatic drive may be elevated, increases in $S(t)$ are thus measured relative to its minimum possible value. We define $C(t)$ as the circadian modulation of pain sensitivity. Then the observed daily rhythm of pain sensitivity $P_{obs}(t)$ is defined as

$$P_{obs}(t) = C(t) + S(t) = C(t) + \sigma(\mathscr{S}(t) - \mathscr{S}_{0min}). \tag{7}$$

As the physiological processes regulating circadian rhythms are believed to be independent of the processes regulating sleep homeostasis, we assume that $C(t)$ and $S(t)$ are likewise regulated independently. Thus, under conditions of sleep deprivation or restriction that do not change the circadian rhythm, we assume that the circadian modulation of pain sensitivity $C(t)$ does not vary, while the sleep-dependent component $S(t)$ would vary with variation in Process S. For example, consider the scenario of 8 h sleep deprivation due to extending wake 8 h beyond the normal sleep onset time of 10pm in the Two-Process Model. Let $S_{SD8}(t) = \sigma(\mathscr{S}_{SD8}(t) - \mathscr{S}_{0min})$ be the modified sleep-dependent pain modulation where $\mathscr{S}_{SD8}(t)$ is Process S under this instance of 8 h sleep deprivation and $\mathscr{S}_{0min}$ is the minimum value of Process S under normal sleep–wake behavior. As in the original Two-Process Model, this sleep disruption is assumed not to significantly affect circadian rhythms, thus the predicted pain sensitivity in this scenario, $P_{SD8}(t)$, is given by

$$P_{SD8}(t) = C_0(t) + S_{SD8}(t) = (P_{obs}(t) - S_0(t)) + S_{SD8}(t) \tag{8}$$

where $S_0(t) = \sigma(\mathscr{S}_0(t) - \mathscr{S}_{0min})$ and $\mathscr{S}_0(t)$ is Process S under normal sleep–wake behavior.

To identify an appropriate value for the scaling parameter σ, we constrain the model to replicate the change in pain sensitivity observed after one night of sleep deprivation. Specifically, σ is chosen so that $S(t)$ reaches values between 12–14 after 8 h of sleep deprivation as modeled with the Two-Process Model to reflect the experimentally observed approximately 13% increase in evoked pain responses measured relative to the range of painful stimulation (see Table 6 in [7]). This yields values for σ in the interval $[18.1433, 21.1672]$.

We now compute $C_0(t)$ as given in Eq. (8) as

$$C_0(t) = P_{obs}(t) - S_0(t) = P_{obs}(t) - \sigma(\mathscr{S}_0(t) - \mathscr{S}_{0min}), \tag{9}$$

where $\mathscr{S}_{0min} = 0.0953$, the minimum value reached by $\mathscr{S}_0(t)$ under normal sleep behavior.

In summary, our model for the predicted rhythm of pain sensitivity due to circadian and sleep-dependent modulation under a specific sleep–wake behavior pattern α, $P_\alpha(t)$, is defined as

$$P_\alpha(t) = C_0(t) + S_\alpha(t) = C_0(t) + \sigma(\mathscr{S}_\alpha(t) - \mathscr{S}_{0min}), \tag{10}$$

where $\mathscr{S}_\alpha(t)$ is Process S as simulated by the Two-Process model under sleep–wake behavior pattern α and $\mathscr{S}_{0min}$ is its minimum under normal sleep–wake behavior. We note that the range of scalings for $P_{obs}(t)$ (namely, $[0.6, 0.7]$) and σ ($[18.1433, 21.1672]$), reflecting the experimentally reported average effects of daily rhythm and sleep deprivation on evoked pain responses, leads to a range of predicted values for $C_0(t)$, $S_\alpha(t)$ and $P_\alpha(t)$ which are indicated by the thickness of the curves in the middle and bottom panels of Figs. 2, 3, and 4.

4 Model Predictions

To predict how sensitivity to pain was affected by sleep schedules, we applied our model to three different simulations of disrupted sleep, namely sleep deprivation, sleep restriction, and a shift work schedule. To simulate the effect of these sleep schedule perturbations on the homeostatic sleep drive, we manually induced sleep or wake transitions in the Two-Process model and ignored the state transition thresholds dictated by Process C. At the end of the sleep perturbation protocol, we re-initiated the Process C threshold crossing rules for sleep initiation and termination.

4.1 Pain Sensitivity Under Sleep Deprivation

First, we tested the change in pain sensitivity due to a continuous sleep deprivation protocol (Fig. 2). This numerical experiment simulated a common protocol in which human subjects are kept awake for 12 h beyond their customary bed time. To obtain the behavior of Process S in this protocol, $\mathscr{S}_\alpha(t) = \mathscr{S}_{SD12}(t)$ in Eq. (10), we ran the Two-Process model for 5 days during which the first 24 h simulated the customary wake time at 6 am and sleep initiation at 10 pm. On day 2 of the simulation, we initiated the continuous sleep deprivation protocol, by ignoring the threshold crossing condition for Process C for 28 h. After 28 h, we re-initiated the rule for sleep initiation when Process S is above $\mathscr{C}_W$, which resulted in immediate sleep onset. For the remainder of the simulation, the Two-Process model followed its usual evolution (Fig. 2, top panel).

During the sleep deprivation protocol, Process S continued to increase exponentially beyond its usual values (top panel, blue curve), which drove an increase in sleep-dependent modulation of pain, $S_\alpha(t) = S_{SD12}(t)$ in Eq. (10) (middle panel, cyan curve). When $S_{SD12}(t)$ was added to the circadian modulation of pain function, $C_0(t)$ (middle panel, magenta curve), the resulting predicted pain sensitivity function, $P_\alpha(t) = P_{SD12}(t)$ in Eq. (10) (lower panel, red curve) showed a large increase in sensitivity, compared to normal levels (lower panel, blue curve), during the deprivation period which continued into the subsequent sleep episode. An interesting prediction of the model is a reduction in pain sensitivity due to the sleep

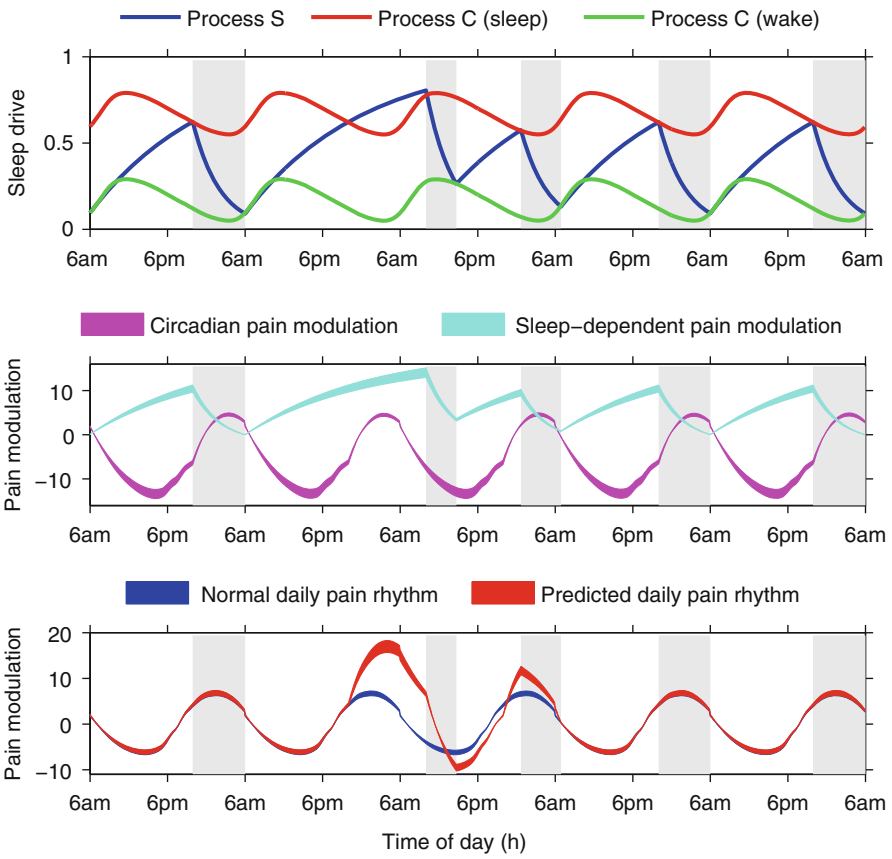


Fig. 2 Predicted pain sensitivity during a simulation of continuous sleep deprivation protocol for 28 h. Shaded regions represent sleep times. *Top*: Simulation of the Two-Process model with 12 h of sleep deprivation initiated at the customary sleep onset time on day 2. *Middle*: The sleep-dependent pain modulation (*cyan curve*) increased during the sleep deprivation protocol while the circadian pain modulation was unperturbed (*magenta curve*). *Bottom*: Combining the sleep-dependent and circadian pain modulation yielded the predicted daily rhythm in pain sensitivity (*red curve*) that showed increases in sensitivity during the deprivation protocol and decreases the following day, compared to the rhythm under the normal sleep schedule (*blue curve*). The thickness of the curves in the middle and bottom panels indicates the range of values obtained due to the range of scalings for $P_{obs}(t)$ and σ (see Sect. 3). Vertical axis units for top: level of homeostatic sleep drive normalized to be between 0 and 1; for middle and bottom: levels of sleep-dependent and circadian pain modulation are set so that the normal daily pain rhythm (*blue curve*) has amplitude 12–14 with mean 0

deprivation. Right after awakening from the recovery sleep episode, pain sensitivity was lower than under the normal sleep schedule. This occurs because of a phase shift in the perturbed sleep-dependent pain modulation, whose minimum coincides with the minimum of the circadian pain modulation. This reduction reversed as the circadian modulation neared its peak because the model remains in waking

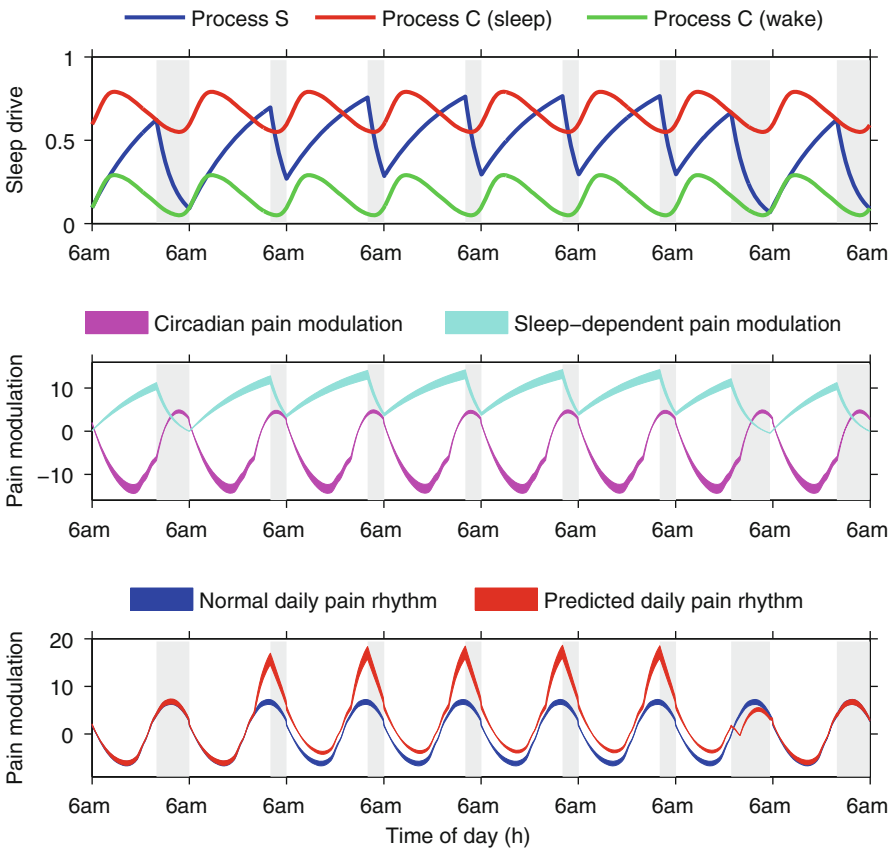


Fig. 3 Predicted pain sensitivity during a simulation of sleep restriction to 4 h a night for 5 consecutive nights. Shaded regions represent sleep times. *Top*: Simulation of the Two-Process model with sleep restricted to occur between 2 − 6 am for days 2 − 6. *Middle*: Sleep-dependent pain modulation (*cyan curve*) remained elevated for days 2−6 while the circadian pain modulation was unperturbed (*magenta curve*). *Bottom*: The predicted daily pain sensitivity (*red curve*) showed non-uniform increases across the days when sleep was restricted, compared to levels under the normal sleep schedule (*blue curve*). Vertical axis units are the same as in Fig. 2

beyond the normal sleep onset time, leading to an increase in sleep-dependent pain modulation.

4.2 Pain Sensitivity Under Sleep Restriction

Another common sleep disruption that occurs in people's daily lives and is induced in experimental settings is sleep restriction during which the time allowed for sleep is restricted over several consecutive days. We simulated a sleep restriction protocol

in which sleep is allowed for only 4 h per night, during 2–6 am, for 5 consecutive nights as might occur in a typical work week (Fig. 3). We simulated the Two-Process model for 8 days where on days 2–6 sleep onset and wake onset were manually induced for the sleep restriction protocol (top panel) to obtain $\mathscr{S}_\alpha(t) = \mathscr{S}_{SR4}(t)$ (top panel). Following Process S, the sleep-dependent pain modulation $S_\alpha(t) = S_{SR4}(t)$ was elevated during days 2–6 (middle panel, cyan curve) predicting continuous increased pain sensitivity over those days (bottom panel, red curve). The interaction of the sleep-dependent and circadian pain modulation led to non-uniform increases in pain sensitivity across the day. The smallest increase occurred during the evening hours after 6 pm, but pain sensitivity increased significantly during the hours between the customary sleep onset time of 10 pm and the allowed sleep onset time of 2 am. During the allowed sleep episode, pain sensitivity decreased but remained elevated at wake onset at an intermediate level. On day 7 when sleep was allowed to occur normally, pain sensitivity decreased during the sleep episode as the homeostatic sleep drive decayed to normal levels.

4.3 Pain Sensitivity Under Shift Work Schedules

Shift work and the resulting misalignment of sleep schedules with the circadian rhythm have been associated with a myriad of adverse health conditions, including increased rates of obesity, cardiac disease, and cancer [8, 18]. Additionally, shift work has been correlated with increased rates of reported musculoskeletal and lower back pain [1, 15, 17]. We simulated an 8 h shift work schedule between 11 pm and 7 am for 5 consecutive days as may occur for a typical shiftwork week (Fig. 4). Sleep onset was assumed at 8 am to allow for commuting time between work and home. We simulated the Two-Process model (top panel) for 8 days during which on day 2 sleep onset was suspended until 8 am on the morning of day 3. The duration of sleep was allowed to be dictated by the model. Subsequent shifts were simulated by again suspending sleep onset until 8 am on the morning of the following day. After each shift, sleep behavior was allowed to resume as predicted by the model. This simulation of the Two-Process model generated $\mathscr{S}_\alpha(t) = \mathscr{S}_{SW}(t)$ from which the sleep-dependent pain modulation $S_\alpha(t) = S_{SW}(t)$ was computed (middle panel, cyan curve). The predicted pain sensitivity (bottom panel, red curve) was significantly elevated during the majority of time during the work shift and during the subsequent sleep episode. A decrease in pain sensitivity occurred for a short duration upon awakening due to a decrease in sleep-dependent pain modulation when the circadian pain modulation was near its minimum.

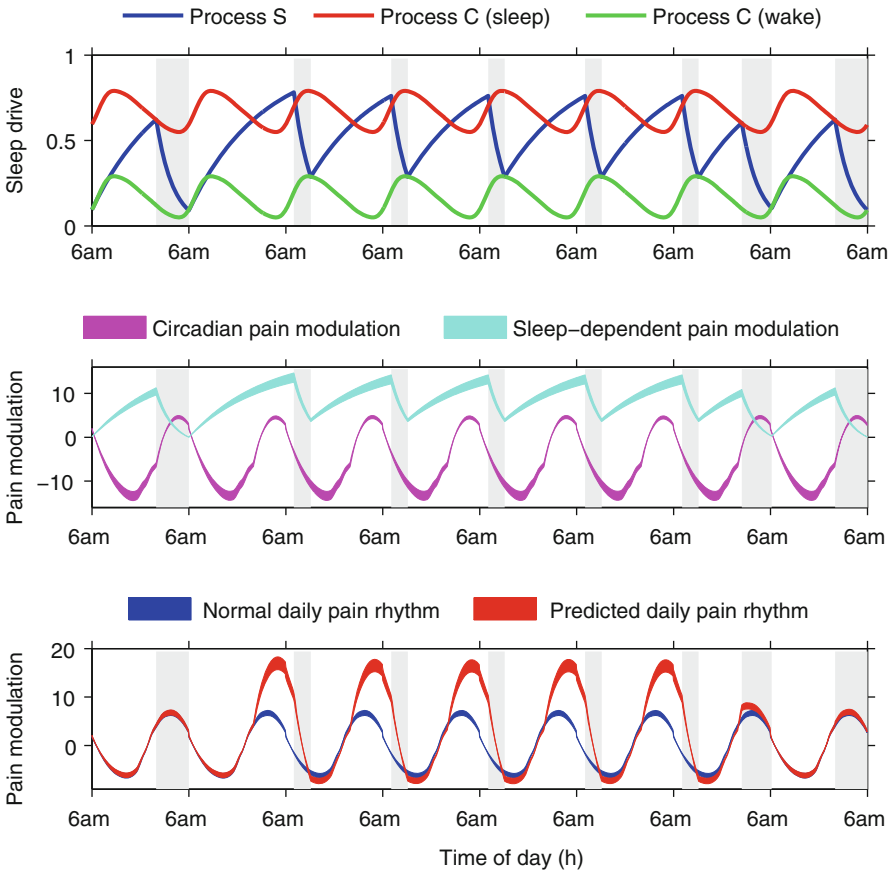


Fig. 4 Predicted pain sensitivity under simulated shift work schedule from 11 pm to 7 am for 5 consecutive days. Shaded regions represent sleep times. *Top*: Simulation of the Two-Process model with sleep onset suspended until 8 am on days 3–7. *Middle*: Sleep-dependent pain modulation (*cyan curve*) remained elevated for days 2–6 while the circadian pain modulation was unperturbed (*magenta curve*). *Bottom*: The predicted pain sensitivity (*red curve*) showed significant increases during the majority of the shift work hours and during sleep episodes, compared to levels under the normal sleep schedule (*blue curve*). Vertical axis units are the same as in Fig. 2

5 Discussion

We have developed a mathematical model to investigate the combined influences and interactions of the homeostatic sleep drive and the circadian rhythm on the variation of pain sensitivity across the day. While experimental results indicate that these two processes have strong effects on pain sensitivity, their combined influences, especially under conditions of disrupted or limited sleep behavior, have not been explored. The model adapts the formalism of a successful mathematical model for the regulation of sleep behavior by the circadian rhythm and homeostatic

sleep drive, called the Two-Process model [3]. Similar to the Two-Process model, our model for pain sensitivity $P(t)$ assumes that the sensitivity level is independently modulated by a circadian process $C_0(t)$ and by a sleep-dependent process $S(t)$. The effects of these two processes were derived from the experimentally observed daily rhythm of pain, $P_{obs}(t)$, computed from multiple experimental results, by applying the assumption that the sleep-dependent process $S(t)$ is driven by the homeostatic sleep drive as predicted by the Two-Process model. The resulting circadian modulation $C_0(t)$ imparts a roughly sinusoidal modulation to $P(t)$ that peaks during the early morning hours before customary awakening and has a trough in the late afternoon hours (middle panels of Figs. 2, 3, and 4, magenta curves). The sleep-dependent modulation $S(t)$ exponentially increases during wake and decreases during sleep (cyan curves).

The strength of the model comes from the ability to predict how the daily cycle of pain sensitivity will change due to changes in sleep behavior. The original Two-Process model has been validated against many diverse disrupted sleep schedules, such as sleep deprivation and sleep restriction. Our model incorporates these results of the Two-Process model to predict the consequent variations in pain sensitivity, as illustrated here for 12 h of sleep deprivation, 4 h sleep restriction, and shift work schedules.

Another advantage of the model is the ability to quantitatively compare the effects of different sleep schedules on pain sensitivity. For example, the simulated shift work schedule shown in Fig. 4 results in 4.1 h of sleep on the 5 shift work days, approximately 30 min more total sleep time as during the simulated restricted sleep schedule shown in Fig. 3. Despite similar amounts of sleep during these two schedules, there are large differences in the effect of these schedules on pain sensitivity. The shift work schedule resulted in higher maximum pain sensitivity, inducing an increase equal to 98.7% of the maximum amplitude of the daily observed rhythm in pain sensitivity $P_{obs}(t)$, while the sleep restriction schedule induced a 78.9% increase. The timing of this maximum also differed with it occurring 2 h prior to sleep onset in the shift work schedule and at sleep onset in the sleep restriction schedule. Thus, while the shift work schedule allows for slightly more sleep, it results in higher peak pain sensitivity that may be more apparent or debilitating since it occurs during active waking periods compared to the sleep restriction schedule. However, the sleep restriction schedule results in a larger portion of the waking period with increased pain sensitivity compared to the shift work schedule. Under shift work, during approximately 70% of the waking period there are negligible changes in pain sensitivity while under sleep restriction pain sensitivity is increased throughout the entire waking period. This type of quantitative comparison possible with the model would be especially useful in evaluating potential sleep interventions, such as nap schedules, in alleviating pain sensitivity increases.

Our model assumes that pain sensitivity is determined by the homeostatic sleep drive. This assumption is based on findings from sleep deprivation experiments, however additional experiments are needed to validate it. Sleep drive is believed to be due to some structure that tracks or some substance that accumulates a

"need to sleep" during prolonged wakefulness, and discharges this homeostatic need during sleep. Several neuromodulators/neurotransmitters have been proposed to serve a role of a homeostatic accumulator for the need to sleep, with adenosine as a leading candidate [2, 14]. During prolonged wakefulness, adenosine levels rise in some sleep-related areas of the brain [11]. Injection of adenosine causes sleep in cats and rats [13, 14]. Hence, at least one mechanism for homeostatic sleep drive might be an accumulation of adenosine that enhances the activity of sleep-promoting brain areas and reduces activity in wake-promoting brain areas. Therefore, a test of our modeling assumption that the homeostatic sleep drive increases pain sensitivity would be to investigate changes in pain sensitivity in animals injected with adenosine. If our assumption is correct, pain sensitivity will increase after the adenosine injection.

Patients with chronic pain often suffer from insomnia and report pain as the primary reason for their disrupted sleep. Our model assumes that the circadian rhythm and homeostatic sleep drive determine daily cycles of pain sensitivity. However, other physiological or cognitive factors surely affect responses to painful stimuli. For example, the perception of painful stimuli might be affected by its timing relative to wake onset. While cognitive functions are reduced for a short time after waking from sleep, a phenomenon known as sleep inertia [16], this cognitive state has interesting effects on the perception of pain. It has been shown that sleep inertia has no effect on pain perception when subjects are awoken abruptly from slow wave or non-rapid eye movement (non-REM) sleep but it reduces pain perception when awoken abruptly from REM sleep [4]. While the reasons for this difference are not known, these results suggest that patients who wake up in pain either perceive accurately the pain they are experiencing, or at worst underestimate the level of pain if woken from REM sleep.

The model is phenomenological in nature; specific physiological mechanisms underlying sleep-dependent and circadian modulation of pain are not explicitly modeled. Instead, the pain sensitivity variable of the model has units of percent change in sensitivity relative to the estimated range of painful stimulation for the particular pain modality, and we've shifted it so that the mean sensitivity is zero. This phenomenological form has the advantage that the model can be validated against data for different modalities and measures of pain sensitivity. In this way, we believe the model has the potential to be a useful tool for pain management by providing predictions of the variations in pain sensitivity due to changing sleep schedules. Future work will focus on identifying data sets for model validation. The advent of activity monitoring devices, such as fitbit (fitbit.com), that can continuously track sleep behavior, and wearables that can prompt users to easily record pain sensitivity at multiple time points per day would provide the type of continuous data on wake and sleep schedules and pain that would be ideal for further developing our model.

Acknowledgements This work was conducted as a part of A Research Collaboration Workshop for Women in Mathematical Biology at the National Institute for Mathematical and Biological Synthesis, sponsored by the National Science Foundation through NSF Award DBI-1300426,

with additional support from The University of Tennessee, Knoxville. This work was additionally partially supported by the following sources: NSF Award DMS-1412119 (VB) and the Pritzker Neuropsychiatric Disorders Research Consortium (MH). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Barro, D., Olinto, M.T.A., Macagnan, J.B.A., Henn, R.L., Pattussi, M.P., Faoro, M.W., Garcez, A.d.S., Paniz, V.M.V.: Job characteristics and musculoskeletal pain among shift workers of a poultry processing plant in southern brazil. J. Occup. Health **57**(5), 448–456 (2015)
2. Benington, J.H., Heller, H.C., Restoration of brain energy metabolism as the function of sleep. Prog. Neurobiol. **45**(4), 347–360 (1995)
3. Daan, S., Beersma, D.G., Borbely, A.A.: Timing of human sleep: recovery process gated by a circadian pacemaker. Am. J. Physiol. Regul. Integr. Comp. Physiol. **246**(2), R161–R183 (1984)
4. Daya, V.G., Bentley, A.J.: Perception of experimental pain is reduced after provoked waking from rapid eye movement sleep. J. Sleep Res. **19**(2), 317–322 (2010)
5. Fisher, S.P., Foster, R.G., Peirson, S.N.: The circadian control of sleep. In: Clocks, C., Kramer, A., Merrow, M. (eds.) Handbook of Experimental Pharmacology, vol. 217, pp. 157–183. Springer, Berlin/Heidelberg (2013)
6. Frank, M.G.: Astroglial regulation of sleep homeostasis. Curr. Opin. Neurobiol. **23**(5), 812–818 (2013)
7. Hagenauer, M.H., Kile, J., Piltz, S., Toporikova, N., Ferguson, P., Booth, V.: The modulation of pain by circadian and sleep-dependent processes: a review of the experimental evidence. In: Proceedings from a Research Collaboration Workshop for Women in Mathematical Biology. Springer, Berlin (2016)
8. Knutsson, A.: Health disorders of shift workers. Occup. Med. (Lond.) **53**(2), 103–108 (2003)
9. Landolt, H.-P.: Sleep homeostasis: a role for adenosine in humans? Biochem. Pharmacol. **75**(11), 2070–2079 (2008)
10. Mistlberger, R.E.: Circadian regulation of sleep in mammals: role of the suprachiasmatic nucleus. Brain Res. Brain Res. Rev. **49**(3), 429–454 (2005)
11. Porkka-Heiskanen, T., Strecker, R.E., McCarley, R.W.: Brain site-specificity of extracellular adenosine concentration changes during sleep deprivation and spontaneous sleep: an in vivo microdialysis study. Neuroscience **99**(3), 507–517 (2000)
12. Richter, C., Woods, I.G., Schier, A.F.: Neuropeptidergic control of sleep and wakefulness. Ann. Rev. Neurosci. **37**, 503–531 (2014)
13. Scammell, T.E., Gerashchenko, D.Y., Mochizuki, T., McCarthy, M.T., Estabrooke, I.V., Sears, C.A., Saper, C.B., Urade, Y., Hayaishi, O.: An adenosine a2a agonist increases sleep and induces fos in ventrolateral preoptic neurons. Neuroscience **107**(4), 653–663 (2001)
14. Strecker, R.E., Morairty, S., Thakkar, M.M., Porkka-Heiskanen, T., Basheer, R., Dauphin, L.J., Rainnie, D.G., Portas, C.M., Greene, R.W., McCarley, R.W.: Adenosinergic modulation of basal forebrain and preoptic/anterior hypothalamic neuronal activity in the control of behavioral state. Behav. Brain Res. **115**(2), 183–204 (2000)
15. Takahashi, M., Matsudaira, K., Shimazu, A.: Disabling low back pain associated with night shift duration: sleep problems as a potentiator. Am. J. Ind. Med. **58**(12), 1300–1310 (2015)
16. Tassi, P., Muzet, A.: Sleep inertia. Sleep Med. Rev. **4**(4), 341–353 (2000)
17. Ursin, H., Endresen, I.M., Ursin, G.: Psychological factors and self-reports of muscle pain. Eur. J. Appl. Physiol. Occup. Physiol. **57**(3), 282–290 (1988)
18. Wang, X.-S., Armstrong, M.E.G., Cairns, B.J., Key, T.J., Travis, R.C.: Shift work and chronic disease: the epidemiological evidence. Occup. Med. (Lond.) **61**(2), 78–89 (2011)

Introduction to Mathematical Modeling of Blood Flow Control in the Kidney

Anita T. Layton and Aurélie Edwards

Abstract Besides its best known role in the excretion of metabolic wastes and toxins, the kidney also plays an indispensable role in regulating the balance of water, electrolytes, acid–base species, blood volume, and blood pressure. To properly fulfill its functions, it is crucial for the kidney to exercise hemodynamic control. In this review, we describe representative mathematical models that have been developed to better understand the kidney's autoregulatory processes. In particular, we consider mathematical models that simulate renal blood flow regulation by means of key autoregulatory mechanisms: the myogenic response and tubuloglomerular feedback. We discuss the extent to which these modeling efforts have expanded the understanding of renal functions in health and diseases.

1 Introduction

The kidney is a major component of the excretory system. Its functions include the removal of waste products from the bloodstream, the regulation of body water and electrolyte balance, and the control of blood volume and blood pressure. Impairment of kidney function is often associated with serious health conditions such as diabetes, hypertension, and congestive heart failure.

The mammalian kidney can be divided into two regions: the outer zone, known as the renal cortex, is where blood is filtered; the inner zone, known as the renal medulla, is the region where the filtered fluid that will ultimately become urine is concentrated. An average human kidney weighs $< 0.2\%$ of body weight. Despite their small size, the kidneys receive a substantial fraction (20–25%) of the cardiac output. Blood is delivered into the kidney through the renal artery, which, following

A.T. Layton (✉)
Department of Mathematics, Duke University, Durham, NC 27708, USA
e-mail: alayton@math.duke.edu

A. Edwards
Sorbonne Universités, UPMC Univ Paris 06, Université Paris Descartes, Sorbonne Paris Cité, INSERM UMRS 1138, CNRS ERL 8228, Centre de Recherche des Cordeliers, Paris, France
e-mail: aurelie.edwards@crc.jussieu.fr

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_4

successive branching, eventually gives rise to the afferent arterioles that supply the glomeruli, i.e., the tufts of capillaries that filter plasma.

A kidney contains a large number of glomeruli, ranging from 30,000 to 40,000 in rats to ∼1 million in humans. Each glomerulus feeds into an associated nephron. The nephrons are the functional units of the kidney that are responsible for the transformation of the glomerular filtrate into urine. Structurally, a nephron is a long tubule lined by a single layer of epithelial cells. Functionally, a nephron consists of several segments each with distinct transport properties. As the filtrate flows through the nephron segments, its composition changes through reabsorption and secretion of fluid and solutes by the tubular epithelia. The composition of the final urine is adjusted so that urinary excretion roughly matches daily intake.

Normal renal function requires that the glomerular filtration rate associated with each individual nephron (called single-nephron glomerular filtration rate, SNGFR) be kept within a narrow range. When SNGFR falls outside of that range, the tubular flow rate in turn falls outside of the appropriate range; when the latter happens, the ability of the nephron to operate properly may be compromised. Glomerular hyperfiltration (i.e., SNGFR that is too high) may impose an excessive load on the nephron and cause tubular damage if the elevated metabolic demand of any tubular segment cannot be met. Glomerular hypofiltration (i.e., SNGFR that is too low) may lead to back diffusion of waste products that need to be excreted. The stabilization of SNGFR is known as renal autoregulation. One key function of renal autoregulation (i.e., blood flow regulation) is to maintain a stable glomerular filtration rate in spite of changes in systemic blood pressure (within about 80–180 mmHg). Another important function of renal autoregulation is to protect the glomerular capillaries from excessive intravascular pressure and shear stress.

We discuss below two major renal autoregulatory mechanisms: the myogenic response and tubuloglomerular feedback. The two mechanisms respond to different signals, but both act on the afferent arteriole. The goal of this review is to provide a brief introduction to the articles in this volume which concern renal hemodynamics [5, 27]. For a comprehensive review on renal hemodynamics, see [6]; for a review on mathematical modeling of renal hemodynamics, see [26].

2 Myogenic Response

When blood pressure is elevated, the smooth muscles that form the afferent arteriole respond by *constricting*. Conversely, when blood pressure is reduced, those smooth muscles dilate. This phenomenon is known as the myogenic response, and is notably the opposite of the passive response of a compliant tube. The afferent arteriole is not unique in exhibiting the myogenic response; indeed, that response is found in nearly all terminal vessels. That said, the reaction time and effectiveness of the myogenic response differ among different vessels, and the response of the afferent arteriole is distinguished by its ability to buffer large perturbations in pressure (up to 80 mmHg) and its relatively short response time (as low as 10 s).

Loutzenhiser et al. formulated a simple model of the renal myogenic response [20, 30]. The radius of the model afferent arteriole, denoted $r(t)$, is given by

$$\frac{d}{dt}r(t) = \frac{1}{\tau}\left(r^{\infty}(p) - r(t)\right) \tag{1}$$

where the target radius r^{∞} is determined by the arterial pressure p. The relationship between r^{∞} and p was chosen based on the observed vasoresponse of the afferent arteriole, such that a higher p yields a smaller r^{∞}. The time constant τ was chosen to capture the observed kinetics of the afferent arteriole myogenic response.

By construction, the above simple phenomenological model reproduces the observed kinetics and steady-state characteristics of the arteriolar vasoresponse; see Fig. 1. A major drawback of that phenomenological approach, however, is that the underlying intracellular mechanisms are not represented. Given this, Sgouralis and co-workers developed mathematical models of the myogenic response of the afferent arteriole that capture the dynamic processes within the vascular smooth muscle cell [4, 24]. These models, which are based on a model for cerebral arterioles in cat that was developed by Gonzalez-Fernandez and Ermentrout [8], represent variations in the transmembrane electric potential difference (or membrane potential), transmembrane Ca^{2+} and K^{+} fluxes, and cytosolic $[Ca^{2+}]$ handling. Key model equations are summarized below.

The rate of change of cytosolic $[Ca^{2+}]$ is given by

$$\frac{d[Ca^{2+}]}{dt} = \left(\frac{(K_d + [Ca^{2+}])^2}{(K_d + [Ca^{2+}])^2 + K_d B_T}\right)\left(\alpha I_{Ca} - k_{Ca}[Ca^{2+}]\right) \tag{2}$$

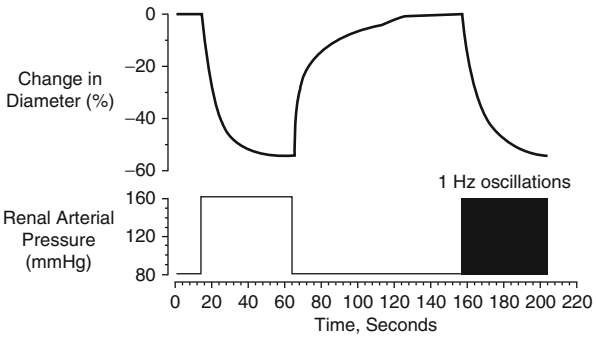


Fig. 1 Myogenic responses to a step pressure perturbation and to fast pressure oscillations. The afferent arteriole constricts at elevated arterial pressure. Due to the asymmetries in the activation times and rate constants induced by increasing/decreasing pressure, the afferent arteriole responds to rapid oscillations with sustained vasoconstriction with a radius that is determined by peak rather than mean pressure. Reproduced from [21]

where I_{Ca} is the Ca^{2+} influx from the extracellular space (see below) and α is a constant converting current to mass flux; k_{Ca} is a first-order rate constant for Ca^{2+} extrusion; K_d characterizes the reaction rates of the calcium-buffer system; and B_T is the total buffer concentration.

Calcium influx is mediated by voltage-gated membrane channels

$$I_{Ca} = g_{Ca} m_\infty (v - v_{Ca}) \tag{3}$$

where g_{Ca} is the maximum whole-cell conductance for the calcium current, and m_∞ is the fraction of channels in the open state at equilibrium, which is a function of the membrane potential (v)

$$m_\infty(v) = \frac{1}{2}\left(1 + \tanh\left(\frac{v - v_1}{v_2}\right)\right) \tag{4}$$

where v_1 is the voltage at which half of the channels are open, and v_2 determines the spread of the distribution.

Transmembrane K^+ efflux is induced by the opening of potassium channels. The model describes the rate of change of the fraction of K^+ channels in the open state, denoted n, by first-order kinetics

$$\frac{dn}{dt} = \lambda_n \left(n_\infty(v, [Ca^{2+}]) - n\right) \tag{5}$$

with the fraction at equilibrium, n_∞, given by

$$n_\infty(v, [Ca^{2+}]) = \frac{1}{2}\left(1 + \tanh\left(\frac{v - v_3}{v_4}\right)\right) \tag{6}$$

where v_3 is a function of $[Ca^{2+}]$, and v_4 is a measure of the spread of the distributions of n_∞.

The opening of Ca^{2+} and K^+ channels depends on the membrane potential v (Eqs. 4 and 6), which is in turn given by the sum of the transmembrane currents, including the Ca^{2+}, K^+, and leak currents. In addition, Sgouralis and co-workers [4, 24] represent the afferent arteriole's myogenic response based on the hypothesis that variations in hydrostatic pressure induce changes in the activity of non-selective cation channels, resulting in a transmembrane current I_{MR}. The rate of change of v is given by

$$C\frac{dv}{dt} = -I_L - I_K - I_{Ca} + I_{MR} \tag{7}$$

where C denotes the capacitance of the cell membrane.

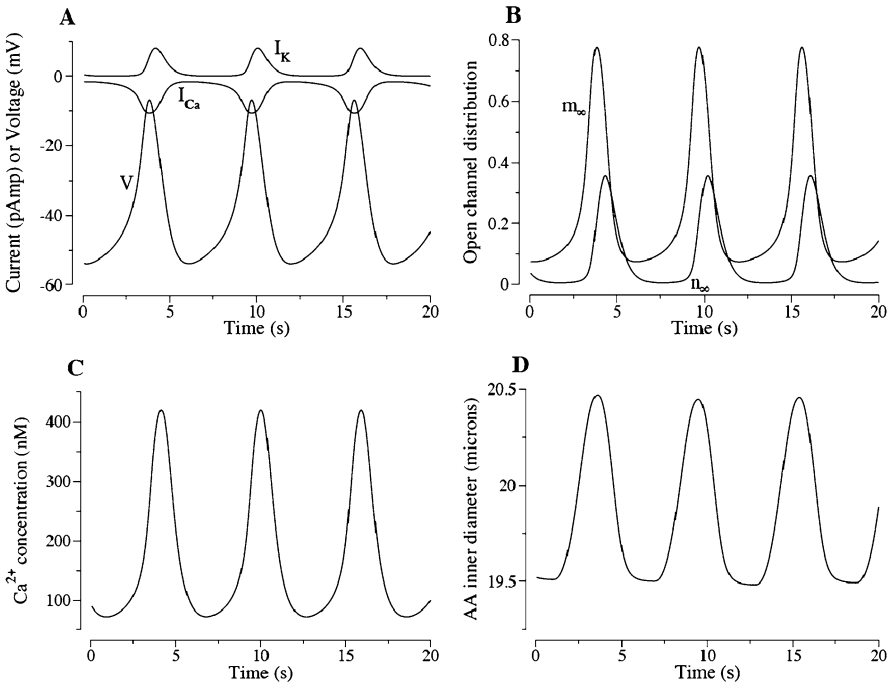


Fig. 2 Spontaneous vasomotion of the afferent arteriole. (**A**) oscillations in Ca^{2+} and K^+ currents (denoted I_{Ca} and I_K, respectively) and membrane potential v. (**B**) oscillations in the fraction of Ca^{2+} and K^+ channels in the open state at equilibrium (denoted m_∞ and n_∞, respectively). (**C** and **D**) oscillations in cytosolic Ca^{2+} concentration and arteriolar diameter. Reproduced from [4]

Cytosolic $[Ca^{2+}]$ regulates the formation of crossbridges, which in turn determines muscle tone. Muscle tone, together with luminal pressure, gives rise to hoop forces, the balance of which determines the luminal radius.

At the baseline pressure, the model [4] predicts that the smooth muscle cell exhibits periodic oscillations in luminal radius even in the absence of external stimuli, i.e., spontaneous vasomotion. Those oscillations, shown in Fig. 2, are triggered by a limit cycle developed between the membrane potential and ion channels. The inward-directed Ca^{2+} current depolarizes the cell, which results in an increase in v that triggers the opening of K^+ channels [Eq. (6)]. The subsequent outward-directed K^+ current re-polarizes the cell, thereby decreasing v. The fluctuations in I_{Ca} modulate cytosolic $[Ca^{2+}]$, which in turn regulates muscle tone and results in spontaneous vasomotion.

In the model, the afferent arteriole smooth muscle cell also exhibits the myogenic response, the signaling pathway of which is described below. At elevated pressures, I_{MR} induces depolarization, which stimulates the opening of voltage-gated Ca^{2+} channels [Eq. (4)], leading to elevations in cytosolic $[Ca^{2+}]$, muscle tone, and

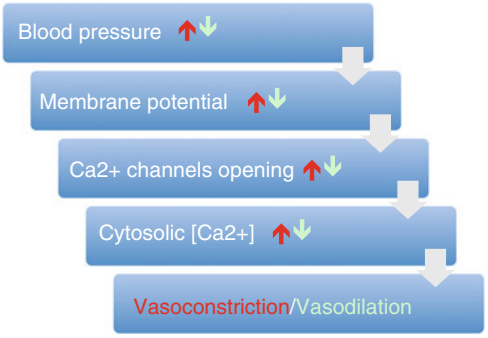


Fig. 3 Myogenic mechanism of the afferent arteriole. Pressure elevation (*red*) induces membrane depolarization, which results in the opening of Ca^{2+} channels and elevation of cytosolic [Ca^{2+}]. Increased [Ca^{2+}] stimulates the phosphorylation of crossbridges, which enhances muscle tone and leads to vasoconstriction. Decreased pressure (*light green*) induces membrane hyperpolarization, which, via the opposite signaling cascade, leads to vasodilation

vasoconstriction. Conversely, at lower pressures, I_{MR} induces hyperpolarization, which has the opposite effects and yields vasodilation. The events leading to the myogenic response are summarized in Fig. 3.

Recently, Edwards and Layton developed a more comprehensive model of intracellular Ca^{2+} signaling within afferent arteriole smooth muscle cells to study the mechanisms underlying the myogenic response [7]. In addition to accounting for the main transporters of Na^+, K^+, and Cl^-, the model also represents intracellular Ca^{2+} dynamics, including Ca^{2+} trafficking between the cytosol and the sarcoplasmic reticulum, which involves the release of Ca^{2+} by ryanodine receptors (RyR) and inositol triphosphate receptors (IP_3R). Indeed, the sarcoplasmic reticulum is an important component of the smooth muscle cell that was neglected in the earlier models [4, 24]. In addition, the model [7] represents the kinetics of myosin light chain (MLC) phosphorylation and the mechanical behavior of the cell. The contractile force (myogenic tone) depends on the fraction of MLC that are phosphorylated. A large number of model equations are used to represent these processes; interested readers are referred to [7], or to [5] in this volume which is an extension of this model.

The highly detailed model [7] yields new insight into the emergence of spontaneous vasomotion. Model results suggest that the time-periodic oscillations stem from the dynamic exchange of Ca^{2+} between the cytosol and the sarcoplasmic reticulum, coupled to the stimulation of Ca^{2+}-activated potassium and chloride channels, and the modulation of voltage-activated L-type channels. Blocking sarco/endoplasmic reticulum Ca^{2+} pumps, RyR, Ca^{2+}-activated potassium and chloride channels, or L-type channels abolishes these oscillations. These details were not, and could not be, predicted by less comprehensive models, e.g., [4, 24].

3 Tubuloglomerular Feedback

Tubuloglomerular feedback (TGF) is a negative feedback system in the kidney that seeks to balance the reabsorptive capacity of the tubules with the filtered load. This is achieved by adjusting SNGFR according to the Cl^- concentration of the fluid reaching the distal nephron. Distal fluid $[Cl^-]$ is sensed by a specialized cluster of cells, known as the macula densa, that is located in the tubular wall in the region where it comes in contact with the terminal part of the afferent arteriole that feeds the glomerulus and nephron [11].

An important tubular segment for TGF function is the thick ascending limb of the loop of Henle, which is immediately upstream of the macula densa. The epithelial cells of the thick ascending limb vigorously pump Na^+ from the tubular fluid into the surrounding interstitium by means of active transport; Cl^- efflux follows. Because the thick ascending limb walls are water impermeable, water does not follow. As a result, the active reabsorption of Na^+ dilutes the tubular fluid; in particular, tubular fluid $[Cl^-]$ decreases along the thick limb.

If SNGFR is elevated above its normal, baseline rate, then, taken in isolation, tubular fluid flow in the thick ascending limb increases. The faster-flowing fluid allows less time for the filtered NaCl to be reabsorbed. As a result, $[Cl^-]$ in the tubular fluid alongside the macula densa is increased above its target value. This concentration increase is viewed as an indication that the NaCl load delivered to the thick ascending limb has exceeded its reabsorptive capacity. Consequently, TGF is activated, which, through a sequence of signaling events, results in a constriction of the afferent arteriole, a subsequent reduction in glomerular blood pressure, and thus a reduction in SNGFR. Conversely, if SNGFR decreases below its baseline rate, $[Cl^-]$ in the tubular fluid alongside the macula densa is decreased below its target value, and TGF acts to increase SNGFR by signaling the afferent arteriole to dilate. The resulting higher flow rate reduces transit time along the thick ascending limb and raises tubular fluid $[Cl^-]$.

Some of the earliest dynamic models of TGF were developed by Holstein–Rathlou and co-workers [2, 9, 10, 12]. A common feature of these models is the representation of the afferent arteriole by a damped linear oscillator upon which TGF acts as external forcing. Motivated by the observation that tubular fluid $[Cl^-]$, the key signal for TGF, changes most substantially along the thick ascending limb of the loop of Henle, Layton and co-workers developed a family of TGF models that explicitly represent tubular transport [13–18, 22, 23] (unlike earlier models which have a much simpler representation of nephron flow). This class of model represents the conservation of Cl^- along the thick ascending limb as

$$\pi r^2 \frac{\partial}{\partial t}[Cl^-] = -Q\frac{\partial}{\partial x}[Cl^-] - 2\pi r\left(\frac{V_{\max}[Cl^-]}{K_M + [Cl^-]} + p\left([Cl^-] - [Cl^-]_{\text{ext}}\right)\right) \tag{8}$$

where $[Cl^-]_{\text{ext}}$ is the interstitial fluid $[Cl^-]$, Q is volume flow, and r is tubular radius. The chloride flux has two components: active transport, characterized by Michaelis–Menten kinetics with parameters $V_{\max}$ and K_M, and passive diffusion,

characterized by the permeability of the tubular wall to Cl^-, p. Because the thick ascending limb is water impermeable, fluid flow $Q(t)$ varies in time but not in space. TGF activation is a function of $[Cl^-]$ at the macula densa (or, the end of the thick ascending limb), denoted C_{MD}. Specifically, TGF is incorporated by assuming a sigmoidal relationship between $Q(t)$ and C_{MD} of the form

$$Q(t) = Q_0 \left(1 + K_1 \tanh\left(K_2 \left(C_{op} - C_{MD}(t-\tau)\right)\right)\right) \tag{9}$$

The TGF delay τ represents the response time of the afferent arteriole; C_{op} denotes the macula densa $[Cl^-]$ at the operating point (i.e., the "target" $[Cl^-]$); K_1 and K_2 determine the maximum range of flow that is affected by TGF and the feedback loop sensitivity.

Renal blood flow undergoes constant perturbations, by heart beats, breathing, movements, etc. For some nephrons, a transient perturbation may decay into a time-independent steady state; for others, a transient perturbation leads to sustained limit-cycle oscillations. Numerical resolution of the model equations 8 and 9 may yield qualitatively different solutions, depending on the bifurcation parameters: τ, the TGF delay, and K_1 and K_2, which determine the TGF sensitivity (or gain). Analysis and numerical simulations have suggested that, in general, for sufficiently small feedback delay and gain, the system is in a stable equilibrium. Thus, for any initial conditions, and for any transient perturbation of a steady-state solution, the model solution converges to a time-independent steady state. An example is shown in Fig. 4, panel B1. In contrast, for larger values of feedback delay and gain, the only stable solution is a regular oscillation that converges to a limit cycle with some frequency f. That frequency depends on the values of the feedback delay and gain. See Fig. 4, panels C1 and D1. A bifurcation analysis of the dynamics of the TGF model can be found in [14].

4 Applications

Together, the myogenic response and TGF maintain a generally stable SNGFR and protect the glomerular capillaries from excessive intravascular pressure and shear stress. Using an integrative model of renal hemodynamics that represents both the myogenic response and TGF, Sgouralis and Layton assessed the individual contributions of the two mechanisms to SNGFR regulation [25]. The integrative model combines an afferent arteriole, a glomerulus, and a short loop of Henle. In the model, the afferent arteriole reacts to perturbations in local luminal fluid pressure (i.e., the myogenic response) and to deviations in tubular fluid $[Cl^-]$ near the macula densa (i.e., TGF). In particular, the TGF response is modeled by adding a new current, I_{TGF}, to the membrane potential equation [Eq. (7)]. I_{TGF} is chosen so that it induces membrane depolarization, hence vasoconstriction, when macula densa $[Cl^-]$ is elevated above it target value, and vice versa. The model predicts that a stable SNGFR is maintained within a physiological range of perfusion pressures

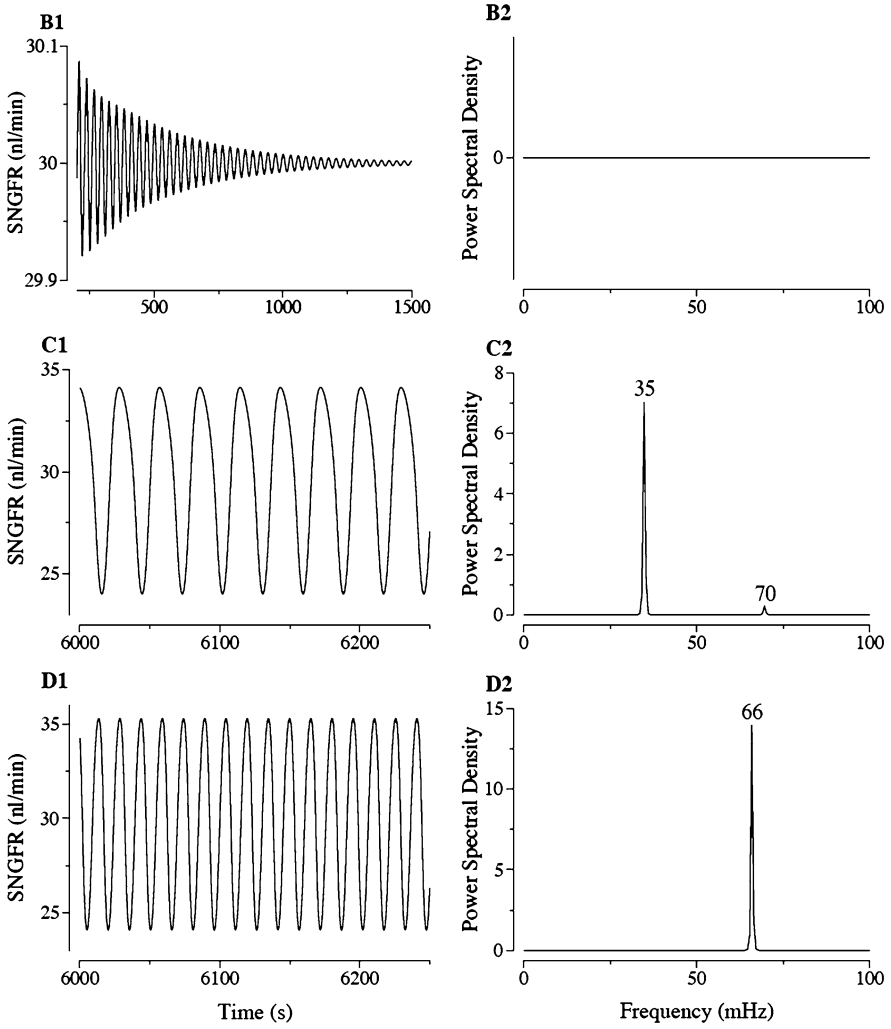


Fig. 4 Theoretical behavior of SNGFR following a transient perturbation, using different values of feedback delay and gain. Response (B1) and power spectrum (B2) for stable solution (a time-independent steady state). Stable single-frequency oscillation (C1) and corresponding power spectrum (C2). Stable double-frequency oscillation and corresponding power spectrum (D2). Reprinted from [19]

(80–180 mmHg). The contribution of TGF to overall autoregulation is significant only within a narrow band of arterial pressure values (80–110 mmHg); outside of that range, the contribution of TGF ceases to increase proportionally because of its sigmoidal response.

The model in [25] was also applied to a pathophysiological condition, namely, diabetes. The investigators conducted simulations to assess the extent to which structural changes and functional impairments observed in diabetic rats cause

glomerular hyperfiltration. To that end, model parameters were adjusted to simulate: (1) dysfunctions in afferent arteriole voltage-gated Ca^{2+} channels [3], Eq. (4), which diminish the vasoconstrictive response, (2) proximal tubule hypertrophy [29], (3) TGF resetting [28], i.e., changes in the macula densa operating point, C_{op} in Eq. (9), and (4) an increase in the glomerular ultrafiltration coefficient [1]. With these modifications, the model predicts hyperfiltration in diabetes, with a SNGFR that is $\sim$60% above baseline value. Each of these changes tends to elevate SNGFR, and from a clinical perspective, it is important to understand the relative contribution of each change. Model simulations suggest that functional impairments in the afferent arteriolar constrictive response are the most important factor.

Acknowledgements This work is the product of a workshop and short-term visits supported by the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. Support was also provided by the National Institutes of Health: National Institute of Diabetes and Digestive and Kidney Diseases and by the National Science Foundation, via grants #DK089066 and #DMS-1263995 to AT Layton.

References

1. Bank, N., Mower, P., Aynedjian, H., Wilkes, B., Silverman, S.: Sorbinil prevents glomerular hyperperfusion in diabetic rats. Am. J. Physiol. **256**, F1000–F1006 (1989)
2. Barfred, M., Mosekilde, E., Holstein-Rathlou, N.: Bifurcation analysis of nephron pressure and flow regulation. Chaos **6**, 280–287 (1996)
3. Carmines, P., Ohishi, K., Ikenaga, H.: Functional impairment of renal afferent arteriolar voltage-gated calcium channels in rats with diabetes mellitus. J. Clin. Invest. **98**, 2564–2571 (1996)
4. Chen, J., Sgouralis, I., Moore, L., Layton, H., Layton, A.: A mathematical model of the myogenic response to systolic pressure in the afferent arteriole. Am. J. Physiol. Ren. Physiol. **300**, F669–F681 (2011)
5. Ciocanel, V., Stepien, T., Edwards, A., Layton, A.: Modeling autoregulation of the afferent arteriole of the rat kidney. In: NIMBioS Proceedings, Springer (2016)
6. Cupples, W., Braam, B.: Assessment of renal autoregulation. Am. J. Physiol. Ren. Physiol. **292**, F1105–F1123 (2007)
7. Edwards, A., Layton, A.: Calcium dynamics underlying the afferent arteriole myogenic response. Am. J. Physiol. Ren. Physiol. **306**, F34–F48 (2014)
8. Gonzalez-Fernandez, J., Ermentrout, B.: On the origin and dynamics of the vasomotion of small arteries. Math. Biosci. **119**, 127–167 (1994)
9. Holstein-Rathlou, N., Leyssac, P.: Oscillations in the proximal intratubular pressure: a mathematical model. Am. J. Physiol. (Renal Fluid Electrolyte Physiol 21) **252**, F560–F572 (1987)
10. Holstein-Rathlou, N., Marsh, D.: A dynamic model of the tubuloglomerular feedback mechanism. Am. J. Physiol. (Renal Fluid Electrolyte Physiol 27) **258**, F1448–F1459 (1990)
11. Holstein-Rathlou, N., Marsh, D.: Renal blood flow regulation and arterial pressure fluctuations: a case study in nonlinear dynamics. Physiol. Rev. **74**, 637–681 (1994)
12. Holstein-Rathlou, N., Wagner, A., Marsh, D.: Tubuloglomerular feedback dynamics and renal blood flow autoregulation in rats. Am. J. Physiol. (Renal Fluid Electrolyte Physiol 29) **260**, F53–F68 (1991)

13. Layton, A.: Feedback-mediated dynamics in a model of a compliant thick ascending limb. Math. Biosci. **228**, 185–194 (2010)
14. Layton, H., Pitman, E., Moore, L.: Bifurcation analysis of TGF-mediated oscillations in SNGFR. Am. J. Physiol. (Renal Fluid Electrolyte Physiol 30) **261**, F904–F919 (1991)
15. Layton, H., Pitman, E., Moore, L.: Instantaneous and steady-state gains in the tubuloglomerular feedback system. Am. J. Physiol. Ren. Physiol. **268**, F163–F174 (1995)
16. Layton, H., Pitman, E., Moore, L.: Nonlinear filter properties of the thick ascending limb. Am. J. Physiol. (Renal Fluid Electrolyte Physiol 42) **273**, F625–F634 (1997)
17. Layton, H., Pitman, E., Moore, L.: Spectral properties of the tubuloglomerular feedback system. Am. J. Physiol. (Renal Fluid Electrolyte Physiol 42) **273**, F635–F649 (1997)
18. Layton, H., Pitman, E., Moore, L.: Limit-cycle oscillations and tubuloglomerular feedback regulation of distal sodium delivery. Am. J. Physiol. Ren. Physiol. **278**, F287–F301 (2000)
19. Layton, A., Moore, L., Layton, H.: Multistability in tubuloglomerular feedback and spectral complexity in spontaneously hypertensive rats. Am. J. Physiol. Ren. Physiol. **291**, F79–F97 (2006)
20. Loutzenhiser, R., Bidani, A., Chilton, L.: Renal myogenic response: kinetic attributes and physiologic role. Circ. Res. **90**, 1316–1324 (2002)
21. Loutzenhiser, R., Griffin, K., Williamson, G., Bidani, A.: Renal autoregulation: new perspectives regarding the protective and regulatory roles of the underlying mechanisms. Am. J. Physiol. Regul. Integr. Comp. Physiol. **290**, R1153–R1167 (2006)
22. Ryu, H., Layton, A.: Effect of tubular inhomogeneities on feedback-mediated dynamics of a model of a thick ascending limb. Med. Math. Biol. **30**, 191–212 (2012)
23. Ryu, H., Layton, A.: Tubular fluid flow and distal nacl delivery mediated by tubuloglomerular feedback in the rat kidney. J. Math. Biol. **68**, 1023–1049 (2014)
24. Sgouralis, I., Layton, A.: Autoregulation and conduction of vasomotor responses in a mathematical model of the rat afferent arteriole. Am. J. Physiol. Ren. Physiol. **303**, F229–F239 (2012)
25. Sgouralis, I., Layton, A.: Theoretical assessment of renal autoregulatory mechanisms. Am. J. Physiol. Ren. Physiol. **306**, F1357–F1371 (2014)
26. Sgouralis, I., Layton, A.T.: Mathematical modeling of renal hemodynamics in physiology and pathophysiology. Math. Biosci. **264**, 8–20 (2015)
27. Sgouralis, I., Layton, A.: Modeling blood flow and oxygenation in a diabetic rat kidney. In: NIMBioS Proceedings, Springer (2016)
28. Thomson, S., Vallon, V., Blantz, R.: Resetting protects efficiency of tubuloglomerular feedback. Kidney Int. Suppl. **67**, S65–S70 (1998)
29. Vallon, V.: The proximal tubule in the pathophysiology of the diabetic kidney. Am. J. Physiol. Regul. Integr. Comp. Physiol. **300**, R1009–R1022 (1996)
30. Williamson, G., Loutzenhiser, R., Wang, X., Griffin, K., Bidani, A.: Systolic and mean blood pressures and afferent arteriolar myogenic response dynamics: a modeling approach. Am. J. Physiol. Regul. Integr. Comp. Physiol. **295**, R1502–R1511 (2008)

Modeling Autoregulation of the Afferent Arteriole of the Rat Kidney

Maria-Veronica Ciocanel, Tracy L. Stepien, Aurélie Edwards, and Anita T. Layton

Abstract One of the key autoregulatory mechanisms that control blood flow in the kidney is the myogenic response. Subject to increased pressure, the renal afferent arteriole responds with an increase in muscle tone and a decrease in diameter. To investigate the myogenic response of an afferent arteriole segment of the rat kidney, we extend a mathematical model of an afferent arteriole cell. For each cell, we include detailed Ca^{2+} signaling, transmembrane transport of major ions, the kinetics of myosin light chain phosphorylation, as well as cellular contraction and wall mechanics. To model an afferent arteriole segment, a number of cell models are connected in series by gap junctions, which link the cytoplasm of neighboring cells. Blood flow through the afferent arteriole is modeled using Poiseuille flow. Simulation of an inflow pressure up-step leads to a decrease in the diameter for the proximal part of the vessel (vasoconstriction) and to an increase in proximal vessel diameter (vasodilation) for an inflow pressure down-step. Through its myogenic response, the afferent arteriole segment model yields approximately stable outflow

The first and second authors made equal contributions.

M.-V. Ciocanel
Division of Applied Mathematics, Brown University, Providence, RI 02912, USA
e-mail: veronica_ciocanel@alumni.brown.edu

T.L. Stepien (✉)
School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85281, USA

Department of Mathematics, University of Arizona, Tucson, AZ 85721, USA
e-mail: stepien@math.arizona.edu

A. Edwards
Sorbonne Universités, UPMC Univ Paris 06, Université Paris Descartes, Sorbonne Paris Cité, INSERM UMRS 1138, CNRS ERL 8228, Centre de Recherche des Cordeliers, Paris, France
e-mail: aurelie.edwards@crc.jussieu.fr

A.T. Layton
Department of Mathematics, Duke University, Durham, NC 27708, USA
e-mail: alayton@math.duke.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_5

pressure for a physiological range of inflow pressures (100–160 mmHg), consistent with experimental observations. The present model can be incorporated as a key component into models of integrated renal hemodynamic regulation.

1 Introduction

In addition to waste excretion, the kidney is responsible for regulating the balance of water, electrolytes, and acid–base species. To accomplish these tasks, the kidney filters a portion of its blood supply into its functional units, the nephrons. Nephrons are elongated tubules surrounded by a layer of epithelial cells. As the filtrate flows through the nephron, its fluid and solutes are selectively reabsorbed or secreted, depending on the animal's physiological state. Consequently, the composition of the tubular fluid changes significantly along the nephron, until it eventually emerges as urine.

For the kidney to properly perform its functions, the rate of filtration into the nephron must be maintained within a narrow range. Thus, blood flow in the kidney is controlled by *autoregulatory mechanisms*. One of the key autoregulatory mechanisms is the tubuloglomerular feedback system, a negative feedback loop that seeks to balance the filtered load of sodium with the reabsorptive capacity of the nephron (Eaton and Pooler [4], Schnermann and Briggs [14]). That goal is accomplished by sensing alterations in tubular fluid chloride concentration at a certain location (alongside the macula densa cells) and then adjusting the muscle tension of the afferent arteriole, and thus renal blood flow and filtration rate, appropriately.

Another key autoregulatory mechanism is the myogenic response, which is an intrinsic property of the afferent arteriole. This mechanism induces a compensatory vasoconstriction of the afferent arteriole when the vessel is presented with an increase in transmural pressure.

The afferent arteriole thus plays a critical role in renal autoregulation. Edwards and Layton [5] previously developed a very detailed mathematical model of Ca^{2+} signaling within an afferent arteriole smooth muscle cell of the rat kidney. The model represents the transmembrane transport of major ions, intracellular Ca^{2+} dynamics, the kinetics of myosin light chain phosphorylation, and the mechanical behavior of the cell. The goal of this study is to develop a multi-cell model of the afferent arteriole by connecting a series of afferent arteriole smooth muscle cells via gap junction coupling and to use the model to study the myogenic response of the vessel. The present afferent arteriole model is intended to be employed as an essential component in models of integrated renal hemodynamic regulation.

2 Mathematical Model

In this section, we summarize the model of a single afferent arteriole smooth muscle cell of the rat kidney, previously developed by Edwards and Layton [5], and then extend this model to a segment of multiple smooth muscle cells that are connected in series via gap junctions.

2.1 Single Cell Model

The main signaling pathways in a single afferent arteriole smooth muscle cell that were considered in Edwards and Layton [5] are represented in Fig. 1.

Considering the K^+, Na^+, Cl^-, and Ca^{2+} channels, the net sum of the currents flowing across the plasma membrane is

$$I_{\text{net}} = I_{\text{K,b}} + I_{\text{K,ir}} + I_{\text{K,v}} + I_{\text{K,Ca}} + I_{\text{NaK}} + I_{\text{Na,b}} + I_{\text{Na,Pres}} + I_{\text{NCX}} + I_{\text{Cl,b}} + I_{\text{Cl,Ca}} + I_{\text{Ca,b}} + I_{\text{Ca,Pres}} + I_{\text{PMCA}} + I_{\text{Ca,L}}, \quad (1)$$

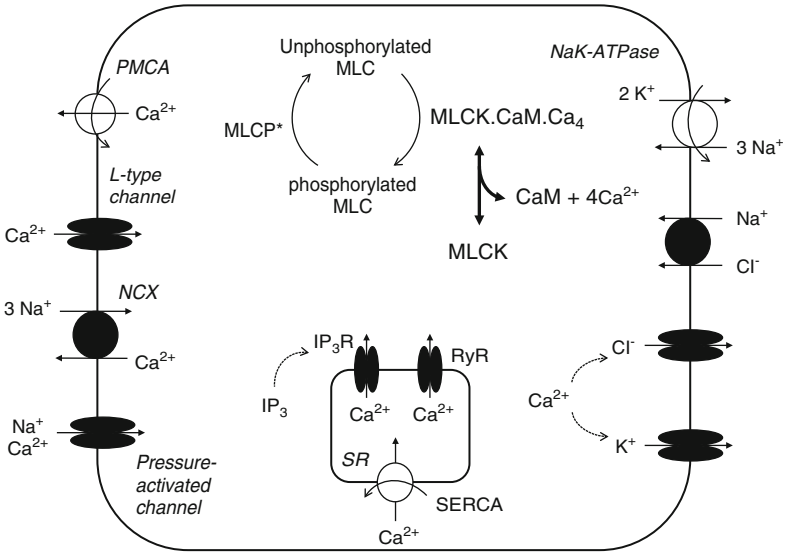


Fig. 1 Representation of a single afferent arteriole smooth muscle cell. The contractile force of the cell depends on the fraction of myosin light chains (MLC) that are phosphorylated. An increase of luminal pressure results in an influx of cations into the cytosol via pressure-activated channels. The ensuing depolarization leads to an increase in cytosolic Ca^{2+} levels, which then enhances the formation of the MLCK.CaM.Ca_4 complex (the active form of myosin light chain kinase, MLCK). Not shown in the diagram are the background currents and the inward- and delayed-rectifier K^+ channels. *MLCP* myosin light chain phosphatase, *CaM* calmodulin, *PMCA* plasma membrane Ca^{2+} pump, *NCX* Na^+/Ca^{2+} exchanger, *SERCA* sarco/endoplasmic Ca^{2+} pump, *RyR* ryanodine receptor, *IP_3R* inositol triphosphate (IP_3) receptor

Table 1 Electrochemical parameters of the cell (overall)

Parameter	Value	Unit	Definition
C_m	5.5×10^{-6}	μF	Plasma membrane capacitance
$[K]_{out}$	5.4	mM	Extracellular K^+ concentration
$[Na]_{out}$	140	mM	Extracellular Na^+ concentration
$[Cl]_{out}$	120	mM	Extracellular Cl^- concentration
$[Ca]_{out}$	2	mM	Extracellular Ca^{2+} concentration
F	96,487	C/mol	Faraday constant
R	8.341	$J \cdot mol^{-1} \cdot K^{-1}$	Ideal gas constant
T	298	K	Temperature
vol_{cyt}	1	pl	Volume of cytosol
$vol_{cyt,Ca}$	0.7	pl	Volume of cytosol accessible to Ca^{2+}
vol_{SR}	0.14	pl	Volume of sarcoplasmic reticulum
G_{gap}/C_m	950	s^{-1}	Ratio of gap junction coefficient-to-membrane capacitance

and the transmembrane potential V_m is described by the differential equation

$$\frac{dV_m}{dt} = -\frac{I_{net}}{C_m}, \tag{2}$$

where C_m is the membrane capacitance. Parameter values and definitions are given in Table 1.

The *pressure-activated ion channels* are assumed to predominately carry Na^+ and be somewhat permeable to Ca^{2+} but not to other ions. The currents across these channels are

$$I_{Na,Pres} = G_{Na,Pres}(V_m - E_{Na}), \tag{3a}$$

$$I_{Ca,Pres} = G_{Ca,Pres}(V_m - E_{Ca}), \tag{3b}$$

where the conductances $G_{Na,Pres}$ and $G_{Ca,Pres}$ depend on the luminal pressure P as

$$G_{i,Pres} = G^0_{i,Pres}\left[1 + 1.75\left(\frac{P}{\bar{P}_1} - 1\right)\left(1 + \frac{|P - \bar{P}_2|}{\bar{P}_2}\right)\right], \tag{4}$$

for $i = Na^+, Ca^{2+}$, with $\bar{P}_1 \equiv 100$ mmHg and $\bar{P}_2 \equiv 60$ mmHg. Parameter values for $G^0_{Na,Pres}$ and $G^0_{Ca,Pres}$ are given in Tables 4 and 6, respectively.

The remaining equations describing transmembrane ionic transport (including ion and charge conservation equations, background currents, and K^+, Na^+, Cl^-, and Ca^{2+} transport pathways), intracellular Ca^{2+} dynamics (including Ca^{2+} buffers such as calmodulin), the kinetics of myosin light chain (MLC) phosphorylation, and vessel mechanics for the single-cell model are described in the Appendix.

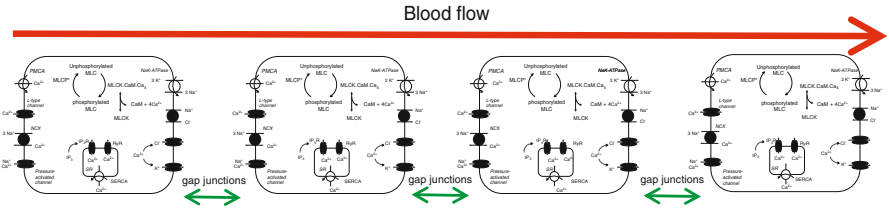


Fig. 2 Representation of a segment of multiple afferent arteriole smooth muscle cells in series. Each cell follows the dynamics of the single-cell model (as in Fig. 1), and the cells are connected to their immediate neighbors via gap junctions. Blood flow through the afferent arteriole lumen is described by Poiseuille flow (6)

2.2 *Multi-Cell Model*

To extend the afferent arteriole smooth muscle single-cell model of Edwards and Layton [5] to an afferent arteriole segment containing multiple smooth muscle cells, we assume that each cell follows the dynamics of the single-cell model as described in the previous section and the Appendix and that all the cells are connected in series via gap junctions. Gap junctions directly connect the cytoplasm of neighboring cells and allow ions to pass through them, thus coupling the cells electrically. The flow of ions carrying an electric charge causes an almost instantaneous diffusion of electrical disturbance to a neighboring cell.

Hence, we modify Eq. (2) for the transmembrane potential of a single cell to a segment of cells (see Fig. 2) where the electric charge of a given cell j may diffuse between its nearest neighbors $j+1$ and $j-1$ such that

$$\frac{dV_{\mathrm{m}}^{j}}{dt} = -\frac{I_{\mathrm{net}}^{j}}{C_{\mathrm{m}}} - \frac{G_{\mathrm{gap}}}{C_{\mathrm{m}}}\left(-V_{\mathrm{m}}^{j+1} + 2V_{\mathrm{m}}^{j} - V_{\mathrm{m}}^{j-1}\right), \qquad j = 1, \ldots, N_{\mathrm{cell}}, \tag{5}$$

where N_{cell} is the total number of cells and G_{gap} is the gap junction coefficient, which denotes the strength of the coupling; we assume that $G_{\mathrm{gap}}/C_{\mathrm{m}} = 950\ \mathrm{s}^{-1}$ (Sgouralis and Layton [15]). Per our convention, $j = 1$ is the first upstream cell and $j = N_{\mathrm{cell}}$ is the last downstream cell. We impose the homogenous Neumann boundary conditions at $j = 1$ and N_{cell}, such that $V_m^0 = V_m^1$ and $V_m^{N_{\mathrm{cell}}+1} = V_m^{N_{\mathrm{cell}}}$.

As blood flows through the segment of smooth muscle cells, the inflow and outflow pressures vary from cell to cell. The luminal blood flow is assumed to be at a quasi-steady state and is therefore described by Poiseuille flow

$$\frac{dP}{dx} = -\frac{8\mu Q}{\pi R^4}, \tag{6}$$

where P is the hydrostatic pressure, μ is the dynamic viscosity of blood, Q is the volumetric flow rate, and R is the luminal radius calculated from the diameter equation (43).

Once the radius R is determined, the hydrostatic pressure P for each cell is updated. This, in turn, affects the pressure-activated ion channel conductances $G_{\mathrm{Na,Pres}}$ and $G_{\mathrm{Ca,Pres}}$ [Eq. (4)], which are modified as

$$G_{i,\mathrm{Pres}} = \max\left\{0, \quad G^0_{i,\mathrm{Pres}}\left[1 + 1.75\left(\frac{P-\bar{P}_0}{\bar{P}_1}\right)\left(1 + \frac{\max\{0,\ P-\bar{P}_2\}}{\bar{P}_2}\right)\right]\right\}, \tag{7}$$

for $i = \mathrm{Na}^+, \mathrm{Ca}^{2+}$. In the above equation, $\bar{P}_1 \equiv 100$ mmHg and $\bar{P}_2 \equiv 60$ mmHg as in Eq. (4), and $\bar{P}_0$ is a reference pressure that decreases approximately linearly from 100 to 91.45 mmHg along the model afferent arteriole, but remains constant for each individual cell subsegment. The added supremums in Eq. (7), as compared to Eq. (4), are to ensure that the conductances are positive (which also leads to numerical stability), that there is no positive contribution to the conductances when pressure P goes below the lower bound $\bar{P}_2$, and that there are no jumps in the arteriole diameter. Both P and $\bar{P}_0$ are vector pressures that vary along the afferent arteriole, so that the myogenic response is activated at each numerical cell by the corresponding pressure in the vector.

2.3 Numerical Method

To numerically solve the multi-cell model, we implement fractional splitting: the single-cell model is solved for each cell separately in the first stage, and the diffusion of electric charge between all cells is taken into account in the second stage.

Letting $\mathscr{R}$ represent the nonlinear reaction part of Eq. (5) (the first term) and f_j^i be the solution to the 51 ordinary differential equations (solved with *ode15s* in MATLAB) of the single-cell model for cell $j = 1, \ldots, N_{\mathrm{cell}}$ at time step i with $\mathbf{f}^i = (f_1^i, f_2^i, \ldots, f_{N_{\mathrm{cell}}}^i)$, the first step in the splitting is

$$\frac{\mathbf{f}^* - \mathbf{f}^i}{\Delta t} = \mathscr{R}(\mathbf{f}^*), \tag{8}$$

where $\mathbf{f}^* = (f_1^*, f_2^*, \ldots, f_{N_{\mathrm{cell}}}^*)$ is the predicted solution that will be corrected by the second step.

Let $\hat{f}_j^*$ be the portion of the solution f_j^* that represents the transmembrane potential V_{m} for cell j with $\hat{\mathbf{f}}^* = (\hat{f}_1^*, \hat{f}_2^*, \ldots, \hat{f}_{N_{\mathrm{cell}}}^*)$. The linear diffusion part of Eq. (5) (the second term) is solved with *ode45* with no flux boundary conditions in MATLAB and is

$$\frac{\hat{\mathbf{f}}^{i+1} - \hat{\mathbf{f}}^*}{\Delta t} = \frac{G_{\mathrm{gap}}}{C_{\mathrm{m}}} \Delta \hat{\mathbf{f}}^{i+1}, \tag{9}$$

where $\hat{\mathbf{f}}^{i+1}$ replaces the transmembrane potential portion of $\mathbf{f}^*$ to obtain $\mathbf{f}^{i+1}$, the solution at time step $i+1$.

Assuming that the inflow pressure applied to the first cell, P_1, is constant throughout time, the hydrostatic pressure for the other cells, P_j, $j = 2, \ldots, N_{\text{cell}}$, is then updated from the Poiseuille flow (6), which is discretized such that

$$P_j^{i+1} = P_{j-1}^{i+1} - \frac{8\mu Q}{\pi (R_j)^4} \Delta x, \qquad j = 2, \ldots, N_{\text{cell}}, \tag{10}$$

where Δx is the width of one numerical afferent arteriole smooth muscle cell subsegment.

In the case where the inflow pressure is varied linearly between some times $\hat{t}_1$ and $\hat{t}_2$, then P_1 is discretized as

$$P_1^{i+1} = P_1^i - \hat{p} \frac{\Delta t}{\hat{t}_2 - \hat{t}_1}, \tag{11}$$

where $\hat{p}$ determines the change in perfusion pressure that is applied to the first cell from time $\hat{t}_1$ to $\hat{t}_2$.

3 Model Results

We apply our blood vessel model to an afferent arteriole of length ~ 240 μm. The parameters we use for the single-cell model are given in Table 1 and the tables in the Appendix. Additionally, the Poiseuille equation (6) parameters are given in Table 2, where we assume that the volumetric flow rate Q is known *a priori*. The model afferent arteriole is discretized into $N_{\text{cell}} = 20$ numerical cells, each of length $\Delta x = 12$ μm.

The base case corresponds to setting the inflow boundary pressure to 100 mmHg. Figure 3 shows the oscillations in membrane potential V_{m} predicted by the model for the first cell in our afferent arteriole segment. The mean value of the transmembrane potential is -36 mV, which is a good approximation of the measured value of -40 mV in the pressurized afferent arteriole. The cytosolic concentration of Ca^{2+} in this cell is shown in Fig. 4. This concentration oscillates between 220 and 320 nM, similar to the range predicted by the single-cell model in Edwards and Layton [5]. The frequency of oscillations for both V_{m} and $[Ca^{2+}]$ is about 0.15 per second. We note that while we show results for the first upstream cell as defined in Eq. (5), similar oscillations are predicted for the subsequent cells.

The diameter of the vessel at the location of the first cell also shows an oscillatory time profile, which stabilizes to an average of 20.1 μm, as can be seen in Fig. 5. We note that this value is only slightly smaller than the average luminal diameter of 20.5 μm predicted by the single-cell model in Edwards and Layton [5].

Table 2 Parameters for Poiseuille equation for blood flow through the afferent arteriole

Parameter	Value	Unit	Definition
μ	1.26×10^{-4}	mmHg · s	Dynamic viscosity of blood
Q	300	nL/min	Volumetric flow rate

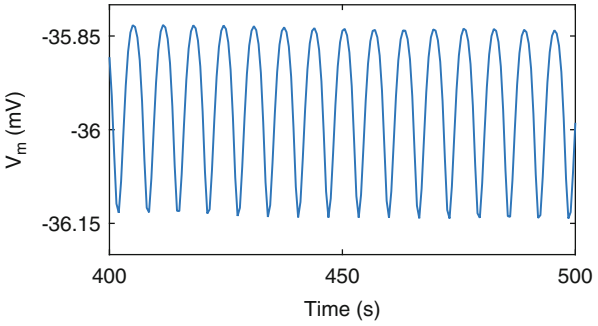


Fig. 3 Predicted oscillations of the membrane potential V_m for the first cell in the afferent arteriole at an inflow pressure of 100 mmHg

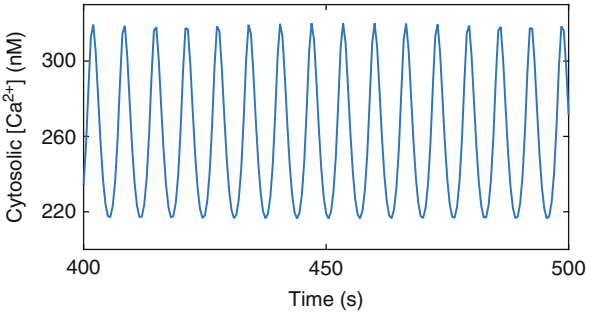


Fig. 4 Predicted oscillations of the cytosolic concentration of Ca^{2+} for the first cell in the afferent arteriole at an inflow pressure of 100 mmHg

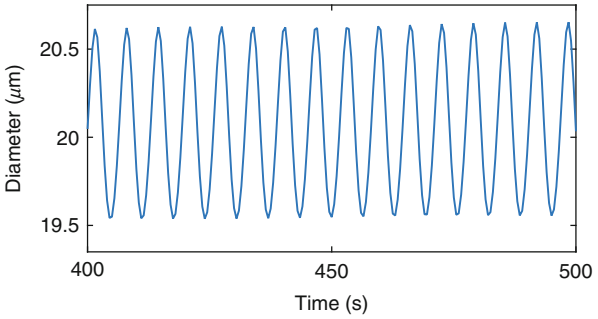


Fig. 5 Predicted oscillations of the afferent arteriole diameter at the first cell coordinate at an inflow pressure of 100 mmHg

We examined the effects of varying the inflow pressure on the afferent arteriole. The left panels in Fig. 6 correspond to model simulations where the inflow pressure starts at the reference value of 100 mmHg, then from 50 to 150 s we increase this pressure linearly to 120 mmHg, and then keep it constant at this higher level for the

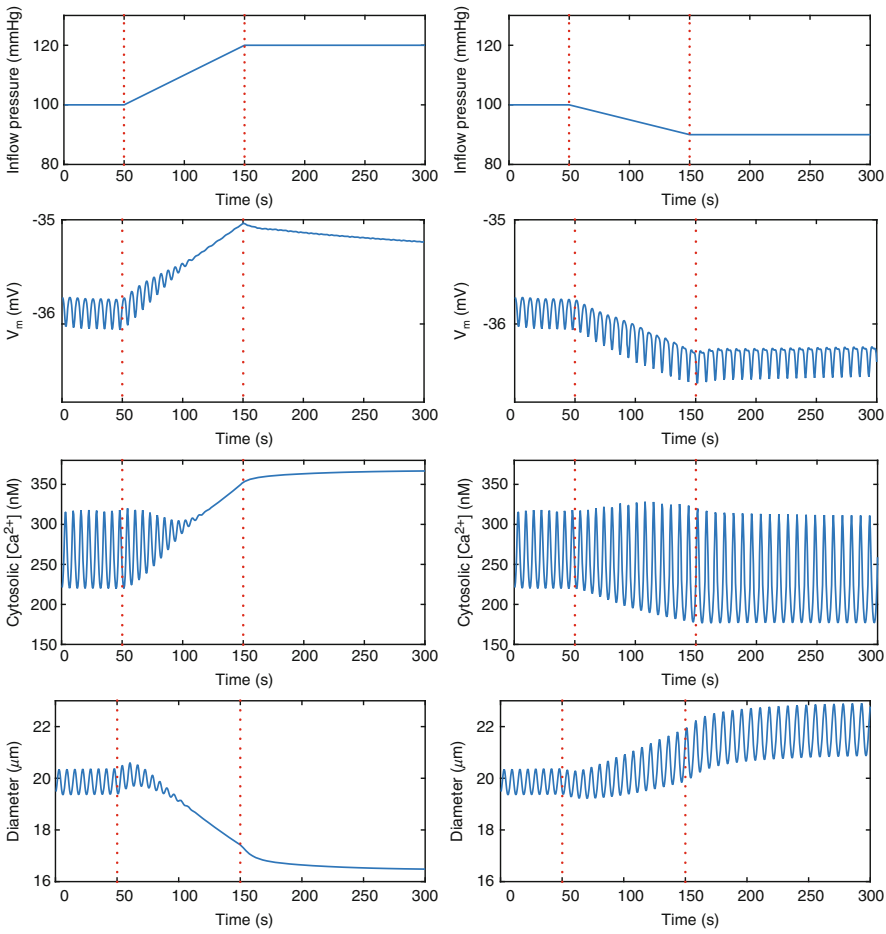


Fig. 6 Time profiles of inflow pressure, V_m, cytosolic $[Ca^{2+}]$, and vessel diameter with a pressure increase to 120 mmHg (*left*) and a pressure decrease to 90 mmHg (*right*). The *red dotted vertical lines* point to the time interval when luminal pressure is linearly increased (*left*) or decreased (*right*)

remainder of the simulation. In a separate simulation, we follow a similar protocol and decrease the luminal pressure linearly to 90 mmHg and then keep it constant at this lower level to produce the panels on the right.

In the case of a pressure increase, we note that this change opens the pressure-activated channels and triggers depolarization, thereby raising the transmembrane potential V_m and, subsequently, the cytosolic concentration of Ca^{2+}. In the pressure down-step case, we observe a reduction in V_m, which elicits a decrease in cytosolic $[Ca^{2+}]$ and leads to an oscillatory regime for this concentration. These observations are similar to single-cell model results in Edwards and Layton [5], where luminal pressure was sharply increased or reduced at a given time point.

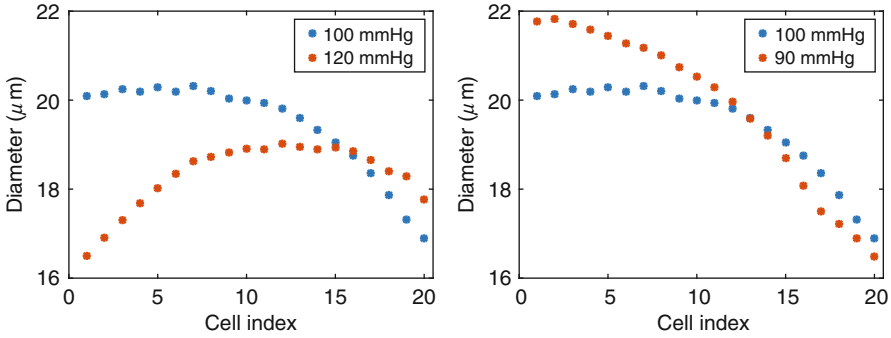


Fig. 7 Time-averaged luminal diameters along the blood vessel for base case inflow pressure 100 mmHg together with pressure increase to 120 mmHg (*left*) and pressure decrease to 90 mmHg (*right*)

The last row of Fig. 6 shows that the diameter of the vessel at the coordinate of the first cell oscillates around 20 μm before $t = 50$ s, as in Fig. 5. At $t = 50$ s, the arteriole acts as a compliant tube and dilates as pressure increases and conversely constricts as pressure decreases. This passive response was also observed in the single-cell model of Edwards and Layton [5].

This effect is quickly replaced by the myogenic response, which leads to a decrease in the vessel diameter (vasoconstriction) for the pressure up-step and an increase in diameter (vasodilation) for the pressure down-step. The vessel diameter stabilizes around 16 μm for the pressure up-step and stably oscillates around 22 μm for the pressure down-step. The plots correspond to the first cell in the vessel for both simulations, and a similar myogenic response controls the time profile of diameters of all subsequent cells (results not shown).

The effects of base case pressure, pressure up-step, and pressure down-step on all vessel cells are shown in Fig. 7. Each blue dot corresponds to the average diameter at a given cell location over a time interval where the radius stabilizes (either converges to a specific value or oscillates around it). The base case pressure of 100 mmHg shows relatively constant diameters for the first (proximal) part of the vessel and a decrease to 16.9 μm in the distal (latter) part of the segment. For the increase in luminal pressure to 120 mmHg, our model predicts vasoconstriction in the proximal part of the afferent arteriole segment, which is reflected in the left panel of Fig. 7 by the low diameters at the level of the first cells. The myogenic response stabilizes downstream pressure, thereby raising the radius in the distal part of the afferent arteriole segment to 17.8 μm despite the pressure up-step. Conversely, a pressure down-step elicits the expected vasodilation in the proximal afferent arteriole segment, as shown in the right panel; its myogenic response yields a distal afferent arteriole diameter that is smaller than the proximal afferent arteriole diameter (16.5 μm at the end cell).

It is worth noting that the diameters in the distal part of the afferent arteriole segment drop back to lower values in the case of the inflow pressure up-step (see left

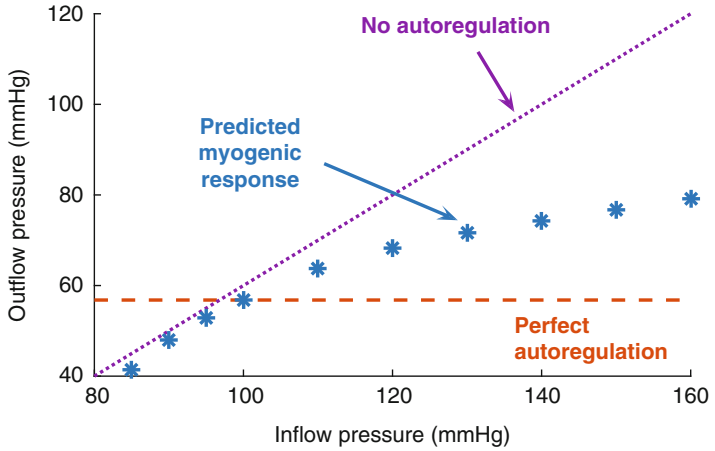


Fig. 8 Predicted myogenic response (*blue*) compared to perfect autoregulation (*orange*) and no autoregulation (*purple*) of blood flow through the vessel for a range of luminal pressures

panel of Fig. 7). In order to understand this effect, we should note that the afferent arteriole diameter represents a balance between the elastic force of the cell and the myogenic response (see Eq. (43)). As we move along the afferent arteriole, the luminal pressure decreases, so that the myogenic response is dominant and leads to the increase in diameter for the pressure up-step case. Toward the end of the vessel, the pressure becomes much lower, and thus, the elastic force (which is proportional to pressure) leads to a decrease in the distal afferent arteriole diameter.

In order to further investigate the model afferent arteriole's myogenic response, we computed the outflow pressure and time- and space-averaged diameter of the afferent arteriole given different constant inflow pressures. Figure 8 shows the predicted outflow pressures at the level of the last afferent arteriole cell. Our results suggest that there is a slow increase in the outflow pressure for reference inflow pressures between 100 and 160 mmHg. This increase is substantially slower than a slope 1 line that would represent no autoregulation and highlights the predicted myogenic response of the afferent arteriole model. Note that the predicted outflow pressure of about 80 mmHg for an inflow pressure of 160 mmHg is higher than the 50 mmHg outflow pressure predicted in Sgouralis and Layton [15]. This higher increase is likely due to the small length of the vessel modeled (240 μm), as well as to the absence of tubuloglomerular feedback in the model. Our numerical cell, chosen to be 12 μm in length for numerical stability, may also affect pressure increases.

Similarly, Fig. 9 shows the space- and time-averaged afferent arteriole diameter for different reference pressure inputs. As in Sgouralis and Layton [15], the results point to vasodilation for small inflow pressures (85 mmHg) where the mean diameter is around 20 μm, and vasoconstriction for large inflow pressures (160 mmHg), where the mean diameter is lower, around 16.5 μm.

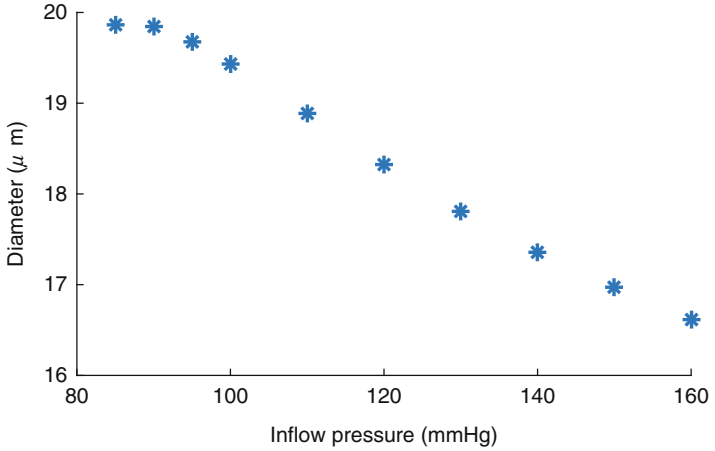


Fig. 9 Predicted space- and time-averaged afferent arteriole diameter for a range of luminal pressures

4 Discussion

We have developed a mathematical model of a segment of the afferent arteriole of the rat kidney. The model represents detailed Ca^{2+} trafficking in each of the afferent arteriole smooth muscle cells, as well as the kinetics of myosin light chain phosphorylation and the mechanical behavior of the cell. The multi-cell afferent arteriole model is an extension of our published cell model [5], which represents the Ca^{2+} dynamics and vasoresponse of a single afferent arteriole smooth muscle cell. The afferent arteriole segment model of this study was constructed by connecting 20 afferent arteriole cell models in series; each cell model is coupled to its neighbors through gap junctions, which allow the representation of electric conduction along the afferent arteriole. A fluid dynamics model was included to relate fluid pressure, fluid flow, and vascular resistance.

The model predicts spontaneous vasomotion at physiological luminal pressures, which arises from the dynamic exchange of Ca^{2+} between the cytosol and the sarcoplasmic reticulum, coupled to the stimulation of Ca^{2+}-activated potassium (K_{Ca}) and chloride (Cl_{Ca}) channels, and the modulation of voltage-activated L-type channels. These spontaneous oscillations of the afferent arteriole muscle tone result in oscillations in fluid pressure and flow.

It is well known that the renal afferent arteriole exhibits the *myogenic response*, wherein it reacts to an elevation in blood pressure with an increase in muscle tone and a decrease in luminal diameter. The myogenic response is believed to stabilize glomerular filtration and to protect the glomerulus from exceedingly high systolic blood pressure, especially in hypertension. The model represents the myogenic response by assuming that the response is initiated by pressure-induced changes in the activity of nonselective cation channels. Through its myogenic response,

the model afferent arteriole stabilizes, to a significant degree, outflow pressure for a range of steady-state inflow pressures, from 100 to 160 mmHg (see Fig. 8). The observed increase in outflow pressure is likely due to the dimension of the vessel considered, as well as to the fact that tubuloglomerular feedback is not included in this model.

With its representation of the myogenic response, the present afferent arteriole segment model can be used as an essential component in models of integrated renal hemodynamic regulation. By coupling a number of afferent arteriole segment models, one can investigate how vasomotor responses propagate among a vascular tree. Furthermore, as we have previously done using a simpler model of an afferent arteriole segment (Sgouralis and Layton [16]), the present model could be combined with a model of glomerular filtration (e.g., Deen et al. [3], Sgouralis and Layton [16]) and a model of the tubuloglomerular feedback mechanism (e.g., Layton [12], Sgouralis and Layton [16]), which is another key renal autoregulatory mechanism. The resulting integrative model of renal hemodynamics could then serve to investigate the interactions between the myogenic and TGF mechanisms in the context of renal autoregulation.

Appendix

This appendix contains the remaining equations besides the ones given in Sect. 2.1 for the afferent arteriole smooth muscle single-cell model of Edwards and Layton [5]. For further details and kinetic diagrams, refer to [5].

Transmembrane Ionic Transport

Ion and Charge Conservation Equations

The cytosolic concentrations of K^+, Na^+, Cl^-, and Ca^{2+} are determined by considering the net sum of their respective fluxes into the cytosol (described in subsequent sections) and integrating

$$\frac{d[\mathrm{K}]_{\mathrm{cyt}}}{dt} = -\frac{(I_{\mathrm{K,b}} + I_{\mathrm{K,ir}} + I_{\mathrm{K,v}} + I_{\mathrm{K,Ca}} - 2I_{\mathrm{NaK}})}{F \cdot \mathrm{vol}_{\mathrm{cyt}}}, \tag{12a}$$

$$\frac{d[\mathrm{Na}]_{\mathrm{cyt}}}{dt} = -\frac{(I_{\mathrm{Na,b}} + I_{\mathrm{Na,Pres}} + 3I_{\mathrm{NaK}} + 3I_{\mathrm{NCX}})}{F \cdot \mathrm{vol}_{\mathrm{cyt}}} + \frac{J_{\mathrm{NaCl}}}{\mathrm{vol}_{\mathrm{cyt}}}, \tag{12b}$$

$$\frac{d[\mathrm{Cl}]_{\mathrm{cyt}}}{dt} = \frac{I_{\mathrm{Cl,b}} + I_{\mathrm{Cl,Ca}}}{F \cdot \mathrm{vol}_{\mathrm{cyt}}} + \frac{J_{\mathrm{NaCl}}}{\mathrm{vol}_{\mathrm{cyt}}}, \tag{12c}$$

$$\frac{d[\mathrm{Ca}]_{\mathrm{cyt}}}{dt} = -\frac{\left(I_{\mathrm{Ca,b}} + I_{\mathrm{Ca,Pres}} + I_{\mathrm{PMCA}} + I_{\mathrm{Ca,L}} - 2I_{\mathrm{NCX}} + I_{\mathrm{SERCA}} - I_{\mathrm{RyR}} - I_{\mathrm{IP3R}}\right)}{2F \cdot \mathrm{vol}_{\mathrm{cyt,Ca}}} + R^{\mathrm{cyt}}_{\mathrm{CaM}} + R^{\mathrm{cyt}}_{\mathrm{Bf}}. \tag{12d}$$

Parameter values and definitions are given in Table 1.

In the sarcoplasmic reticulum (SR),

$$\frac{d[\mathrm{Ca}]_{\mathrm{SR}}}{dt} = \frac{I_{\mathrm{SERCA}} - I_{\mathrm{RyR}} - I_{\mathrm{IP3R}}}{2F \cdot \mathrm{vol}_{\mathrm{SR}}} + R^{\mathrm{SR}}_{\mathrm{Calseq}}. \tag{13}$$

The reaction terms $R^{\mathrm{cyt}}_{\mathrm{CaM}}$, $R^{\mathrm{cyt}}_{\mathrm{Bf}}$, and $R^{\mathrm{SR}}_{\mathrm{Calseq}}$ account for the buffering of Ca^{2+} by calmodulin, other cytosolic buffers, and calsequestrin, respectively, and are described in Eq. (35) below.

Background Currents

The background current of ion i, for $i = K^+, Na^+, Cl^-, Ca^{2+}$, is

$$I_{i,\mathrm{b}} = G_{i,\mathrm{b}} \left(V_{\mathrm{m}} - E_i\right), \tag{14}$$

where the Nernst potential of ion i with valence z_i is

$$E_i = \frac{RT}{z_i F} \ln\left(\frac{[i]_{\mathrm{out}}}{[i]_{\mathrm{cyt}}}\right). \tag{15}$$

Parameter values and definitions are given in Table 1.

Potassium Transport Pathways

The potassium current across inward-rectifier (K_{ir}) channels is determined as

$$I_{\mathrm{K,ir}} = G_{\mathrm{Kir}} P_{\mathrm{Kir}} \left(\frac{[K]_{\mathrm{out}}}{[K]_{\mathrm{ref}}}\right)^{0.9} \left(V_{\mathrm{m}} - E_{\mathrm{K}}\right), \tag{16a}$$

$$P_{\mathrm{Kir}} = \frac{1}{1 + \exp\left(\dfrac{V_{\mathrm{m}} - V_{\mathrm{Kir}}}{s_{\mathrm{Kir}}}\right)}, \tag{16b}$$

where the exponential factor, 0.9, the potential of half-maximal activation, V_{Kir}, and the slope, s_{Kir}, were obtained by fitting K_{ir} currents in cerebral arterial smooth muscle cells (Wu et al. [18]). Parameter values are given in Tables 1 and 3.

Table 3 Parameters for potassium currents

Parameter	Value	Unit
$G_{K,b}$	0	nS
G_{Kir}	0.50	nS
$[K]_{ref}$	5.0	mM
V_{Kir}	−80/−65	mV[a]
s_{Kir}	20/5	mV[a]
G_{Kv}	9.83	nS
G_{KCa}	5.0	nS
τ_{Pf}	0.5	mS
τ_{Ps}	11.5	mS

[a]Values for V_m below/above −60 mV

The potassium current across delayed-rectifier (K_v) channels is given by

$$I_{K,v} = G_{Kv}\,(P_{Kv})^2\,(V_m - E_K)\,, \tag{17a}$$

$$P_{Kv} = 0.58P_{Kv1} + 0.42P_{Kv2}, \tag{17b}$$

$$\frac{dP_{Kv1}}{dt} = \frac{\bar{P}_{Kv} - P_{Kv1}}{\tau_{Kv1}}, \tag{17c}$$

$$\frac{dP_{Kv2}}{dt} = \frac{\bar{P}_{Kv} - P_{Kv2}}{\tau_{Kv2}}, \tag{17d}$$

$$\bar{P}_{Kv} = \frac{1}{1 + \exp\left(-\dfrac{V_m + 1.77}{14.52}\right)}, \tag{17e}$$

$$\tau_{kv1} = 210.99\exp\left[-\left(\frac{V_m + 214.34}{195.35}\right)^2\right] - 20.59, \tag{17f}$$

$$\tau_{kv2} = 821.39\exp\left[-\left(\frac{V_m + 31.59}{27.46}\right)^2\right] + 0.189, \tag{17g}$$

where P_{Kv1} and P_{Kv2} are the two components of the channel activation process and τ_{Kv1} and τ_{Kv2} are the respective time constants (in ms) (Yang et al. [19]). Variable $\bar{P}_{Kv}$ is voltage-dependent and represents the steady-state value of P_{Kv1} and P_{Kv2}.

The potassium current across Ca^{2+}-activated K^+ (K_{Ca}) channels is computed as

$$I_{K,Ca} = G_{KCa}P_{KCa}\,(V_m - E_K)\,, \tag{18a}$$

$$P_{KCa} = 0.65P_f + 0.35P_s, \tag{18b}$$

$$\frac{dP_f}{dt} = \frac{\bar{P}_{KCa} - P_f}{\tau_{Pf}}, \tag{18c}$$

$$\frac{dP_s}{dt} = \frac{\bar{P}_{KCa} - P_s}{\tau_{Ps}}, \tag{18d}$$

$$\bar{P}_{KCa} = \frac{1}{1 + \exp\left(-\dfrac{V_m - V_{KCa}}{21.70}\right)}, \tag{18e}$$

$$V_{KCa} = -45.0 \ \log_{10}\left([Ca]_{cyt}\right) - 63.55, \tag{18f}$$

where P_f and P_s are the fast and slow components of the channel activation process, respectively, and τ_{Pf} and τ_{Ps} are the corresponding time constants (Yang et al. [19]). The steady-state open probability of the channels is given by $\bar{P}_{KCa}$.

The ATP-dependent K^+ channels are not considered in the Edwards and Layton model [5] since it assumed that their conductance is negligible in well-perfused and oxygenated arterioles.

Sodium Transport Pathways

The current across Na^+/K^+-ATPase pumps is determined as

$$I_{NaK} = I_{NaK,max}\left(\frac{[K]_{out}}{[K]_{out} + K^K_{m,NaK}}\right)^2\left(\frac{[Na]_{cyt}}{[Na]_{cyt} + K^{Na}_{m,NaK}}\right)^3. \tag{19}$$

The current across Na^+/Ca^{2+} (NCX) exchangers is given by

$$I_{NCX} = I_{NCX,max}A_{NCX}\left(\frac{\Phi_F[Na]^3_{cyt}[Ca]_{out} - \Phi_R[Na]^3_{out}[Ca]_{cyt}}{G(1 + k_{sat}\Phi_R)}\right), \tag{20a}$$

$$A_{NCX} = \frac{[Ca]^2_{cyt}}{\left(K^{Ca}_{m,NCX}\right)^2 + [Ca]^2_{cyt}}, \tag{20b}$$

$$\Phi_F = \exp\left(\frac{\gamma V_m F}{RT}\right), \tag{20c}$$

$$\Phi_R = \exp\left(\frac{(\gamma - 1)V_m F}{RT}\right), \tag{20d}$$

$$\begin{aligned} G =& [Na]^3_{out}[Ca]_{cyt} + [Na]^3_{cyt}[Ca]_{out} + K^3_{mNao}[Ca]_{cyt} \\ &+ K_{mCao}[Na]^3_{cyt} + K^3_{mNai}[Ca]_{out}\left(1 + [Ca]_{cyt}/K_{mCai}\right) \\ &+ K_{mCai}[Na]^3_{out}\left(1 + [Na]^3_{cyt}/K^3_{mNai}\right) \end{aligned} \tag{20e}$$

(Shannon et al. [17]). The flux across NaCl cotransporters is computed as

$$J_{\mathrm{NaCl}} = J_{\mathrm{NaCl,max}} \frac{(E_{\mathrm{Na}} - E_{\mathrm{Cl}})^4}{(E_{\mathrm{Na}} - E_{\mathrm{Cl}})^4 + R_{\mathrm{NaCl}}^4} \tag{21}$$

(Kneller et al. [11]). Parameter values are given in Tables 1 and 4.

Chloride Transport Pathways

The current across Ca^{2+}-activated Cl^- (Cl_{Ca}) channels is

$$I_{\mathrm{Cl,Ca}} = G_{\mathrm{ClCa}} P_{\mathrm{ClCa}} (V_{\mathrm{m}} - E_{\mathrm{Cl}}), \tag{22a}$$

$$\frac{dP_{\mathrm{ClCa}}}{dt} = \frac{\bar{P}_{\mathrm{ClCa}} - P_{\mathrm{ClCa}}}{\tau_{\mathrm{ClCa}}}, \tag{22b}$$

$$\bar{P}_{\mathrm{ClCa}} = \frac{[\mathrm{Ca}]_{\mathrm{cyt}}^3}{[\mathrm{Ca}]_{\mathrm{cyt}}^3 + K_{\mathrm{ClCa}}^3}, \tag{22c}$$

where $\bar{P}_{\mathrm{ClCa}}$ is the steady-state open probability of the channel (Jacobson et al. [8]). Parameter values are given in Table 5.

Table 4 Parameters for sodium currents

Parameter	Value	Unit
$G_{\mathrm{Na,b}}$	0.007	nS
$G^0_{\mathrm{Na,Pres}}$	0.020	nS
$I_{\mathrm{NaK,max}}$	3.75	μA/μF
$K^{\mathrm{K}}_{\mathrm{m,NaK}}$	1.5	mM
$K^{\mathrm{Na}}_{\mathrm{m,NaK}}$	12	mM
$I_{\mathrm{NCX,max}}$	1.5	μA/μF
$K^{\mathrm{Ca}}_{\mathrm{m,NCX}}$	0.125×10^{-3}	mM
k_{sat}	0.27	Dimensionless

Parameter	Value	Unit
$K_{\mathrm{m,Nao}}$	87.5	mM
$K_{\mathrm{m,Cao}}$	1.3	mM
$K_{\mathrm{m,Nai}}$	12.29	mM
$K_{\mathrm{m,Cai}}$	3.59×10^{-3}	mM
γ	0.35	Dimensionless
$J_{\mathrm{NaCl,max}}$	1.08×10^{-16}	mM/s
R_{NaCl}	87.825	mV

Table 5 Parameters for chloride currents

Parameter	Value	Unit
$G_{\mathrm{Cl,b}}$	0.007	nS
G_{ClCa}	0.80	nS
τ_{ClCa}	0.050	s
K_{ClCa}	140	mM

Table 6 Parameters for calcium currents and buffers

Parameter	Value	Unit
$G_{\mathrm{Ca,b}}$	0.007	nS
$G^{0}_{\mathrm{Ca,Pres}}$	0.020	nS
G_{CaL}	2.75	nS
$I_{\mathrm{PMCA,max}}$	0.9	μA/μF
$K^{\mathrm{Ca}}_{\mathrm{m,PMCA}}$	170×10^{-6}	mM
$I_{\mathrm{SERCA,max}}$	11.76	pA
$K^{\mathrm{Ca}}_{\mathrm{m,SERCA}}$	310×10^{-6}	mM
ν_{RyR}	12	s^{-1}
K_{a}	3.7222×10^{-6}	mM
K_{b}	6.3601×10^{-6}	mM
K_{c}	0.0571	Dimensionless
k^{-}_{c}	0.1	s^{-1}
ν_{IP3R}	12	s^{-1}
K^{IP3}_{+}	1.85	s^{-1}
$[\mathrm{IP}_3]_{\mathrm{ref}}$	240	mM
α_4	0.5	Dimensionless

Parameter	Value	Unit
k_4	1.1×10^{-3}	mM
k^{IP3}_{-}	1.0	s^{-1}
a_1	400×10^{3}	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
a_3	400×10^{3}	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
a_4	0.2×10^{3}	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
a_5	20×10^{3}	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
d_1	0.13×10^{-3}	mM^{-1}
d_3	943.4×10^{-6}	mM^{-1}
d_4	144.5×10^{-6}	mM^{-1}
d_5	82.34×10^{-6}	mM^{-1}
$K^{\mathrm{Bf}}_{\mathrm{on}}$	10^{3}	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
$K^{\mathrm{Bf}}_{\mathrm{off}}$	5	s^{-1}
$[\mathrm{Bf}]^{\mathrm{tot}}_{\mathrm{cyt}}$	0.50	mM
$k^{\mathrm{Calseq}}_{\mathrm{on}}$	10^{5}	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
$k^{\mathrm{Calseq}}_{\mathrm{off}}$	65	s^{-1}
$[\mathrm{Calseq}]^{\mathrm{tot}}_{\mathrm{SR}}$	0.14	mM

Calcium Transport Pathways

Calcium is exchanged between the cytosol and the extracellular space, and between the cytosol and the SR, which acts as a storage compartment. Parameter values for calcium currents and buffer reactions are given in Tables 1 and 6.

The current through plasma membrane Ca^{2+} (PMCA) pumps is determined as

$$I_{\mathrm{PMCA}} = I_{\mathrm{PMCA,max}} \left(\frac{[\mathrm{Ca}]_{\mathrm{cyt}}}{K^{\mathrm{Ca}}_{\mathrm{m,PMCA}} + [\mathrm{Ca}]_{\mathrm{cyt}}} \right). \tag{23}$$

The $\mathrm{Ca_V}1.2$ model of Faber et al. [6] is used for the current across L-type Ca^{2+} channels. The voltage-dependent gating mode of the channel is considered, which includes four closed states (c_0, c_1, c_2, and c_3), one open state (p_{o}), and fast (i_{vf}) and slow (i_{vs}) inactivated states. The corresponding equations are

$$I_{\mathrm{Ca,L}} = G_{\mathrm{CaL}} p_{\mathrm{o}} \left(V_{\mathrm{m}} - E_{\mathrm{Ca}} \right), \tag{24a}$$

$$\frac{dc_0}{dt} = \beta c_1 - (4\alpha) c_0, \tag{24b}$$

$$\frac{dc_1}{dt} = (4\alpha) c_0 + (2\beta) c_2 - (3\alpha + \beta) c_1, \tag{24c}$$

$$\frac{dc_2}{dt} = (3\alpha) c_1 + (3\beta) c_3 - (2\alpha + 2\beta) c_2, \tag{24d}$$

$$\frac{dc_3}{dt} = (2\alpha)c_2 + (4\beta)p_o + \omega_f i_{vf} + \omega_s i_{vs} - (\alpha + 3\beta + \gamma_f + \gamma_s)c_3, \tag{24e}$$

$$\frac{dp_o}{dt} = \alpha c_3 + \lambda_f i_{vf} + \lambda_s i_{vs} - (4\beta + \phi_f + \phi_s)p_o, \tag{24f}$$

$$\frac{di_{vf}}{dt} = \gamma_f c_3 + \phi_f p_o + \omega_{sf} i_{vs} - (\omega_f + \lambda_f + \omega_{fs})i_{vf}, \tag{24g}$$

$$\frac{di_{vs}}{dt} = \gamma_s c_3 + \phi_s p_o + \omega_{fs} i_{vf} - (\omega_s + \lambda_s + \omega_{sf})i_{vs}, \tag{24h}$$

where

$$\alpha = 0.925\exp(V_m/30), \qquad \beta = 0.390\exp(V_m/40),$$
$$\gamma_f = 0.245\exp(V_m/10), \qquad \gamma_s = 0.005\exp(-V_m/40),$$
$$\phi_f = 0.020\exp(V_m/500), \qquad \phi_s = 0.030\exp(-V_m/280),$$
$$\lambda_f = 0.035\exp(-V_m/300), \qquad \lambda_s = 0.0011\exp(V_m/500),$$
$$\omega_f = (4\beta\lambda_f\gamma_f)/(\alpha\phi_f), \qquad \omega_s = (4\beta\lambda_s\gamma_s)/(\alpha\phi_s),$$
$$\omega_{sf} = (\lambda_s\phi_f)/\lambda_f, \qquad \omega_{fs} = \phi_s.$$

T-type Ca^{2+} channels are not considered in the Edwards and Layton [5] model.

The current across sarco/endoplasmic reticulum Ca^{2+} (SERCA) pumps is given by

$$I_{SERCA} = I_{SERCA,max}\left(\frac{[Ca]_{cyt}^2}{\left(K_{m,SERCA}^{Ca}\right)^2 + [Ca]_{cyt}^2}\right). \tag{25}$$

The RyR model of Keizer and Levine [10] is used to determine the RyR-mediated release current into the cytosol,

$$I_{RyR} = v_{RyR}P_{RyR}\left([Ca]_{SR} - [Ca]_{cyt}\right)(2F \cdot vol_{SR}), \tag{26}$$

where v_{RyR} is the RyR rate constant. The open probability of RyR (P_{RyR}) is calculated as

$$P_{RyR} = \frac{\omega\left(1 + \left([Ca]_{cyt}/K_b\right)^3\right)}{1 + \left(K_a/[Ca]_{cyt}\right)^4 + \left([Ca]_{cyt}/K_b\right)^3}, \tag{27a}$$

$$\frac{d\omega}{dt} = \frac{k_c^-(\omega^\infty - \omega)}{\omega^\infty}, \tag{27b}$$

$$\omega^\infty = \frac{1 + \left(K_a/[Ca]_{cyt}\right)^4 + \left([Ca]_{cyt}/K_b\right)^3}{1 + 1/K_c + \left(K_a/[Ca]_{cyt}\right)^4 + \left([Ca]_{cyt}/K_b\right)^3}. \tag{27c}$$

The IP_3R model of De Young and Keizer [2] is used to determine the IP_3R-mediated release current in the cytosol,

$$I_{IP3R} = v_{IP3R}(x_{110})^3 \left([Ca]_{SR} - [Ca]_{cyt}\right)(2F \cdot vol_{SR}), \tag{28}$$

where v_{IP3R} is the IP_3R rate constant and x_{110} is the fraction of receptors bound by one activated Ca^{2+} and one IP_3, calculated as described below. The cytosolic concentration of IP_3 is calculated as

$$\frac{d[IP_3]_{cyt}}{dt} = k_+^{IP3}[IP_3]_{ref}\left(\frac{[Ca]_{cyt} + (1-\alpha_4)k_4}{[Ca]_{cyt} + k_4}\right) - k_-^{IP3}[IP_3]_{cyt}, \tag{29}$$

where k_+^{IP3} and k_-^{IP3} are the rate constants for IP_3 formation and consumption, respectively, $[IP_3]_{ref}$ is a reference IP_3 concentration, and α_4 determines the strength of the Ca^{2+} feedback on IP_3 production.

Three equivalent and independent IP_3R subunits are assumed to be involved in conduction, and each subunit has one IP_3 binding site (denoted as *site 1*) and two Ca^{2+} binding sites, one for activation (*site 2*) and one for inhibition (*site 3*). The fraction of receptors in state $S_{i_1 i_2 i_3}$ is denoted by $x_{i_1 i_2 i_3}$, where i_j equals 0 if the j-th binding site is unoccupied or 1 if it is occupied. All three subunits must be in the state S_{110} (corresponding to the binding of one IP_3 and one activating Ca^{2+}) for the IP_3R channel to be open. Assuming rapid equilibrium for IP_3 binding,

$$a_1[Ca]_{cyt}x_{0k0} = b_1 x_{1k0}, \qquad k = 0, 1, \tag{30a}$$

$$a_3[Ca]_{cyt}x_{0k1} = b_3 x_{1k1}, \qquad k = 0, 1. \tag{30b}$$

Defining $d_k = b_k/a_k$, the conservation equations for $x_{0i_2i_3}$ are

$$\frac{dx_{000}}{dt} = -a_4\left([Ca]_{cyt}x_{000} - d_4 x_{001}\right) - a_5\left([Ca]_{cyt}x_{000} - d_5 x_{010}\right), \tag{31a}$$

$$\frac{dx_{001}}{dt} = +a_4\left([Ca]_{cyt}x_{000} - d_4 x_{001}\right) - a_5\left([Ca]_{cyt}x_{001} - d_5 x_{011}\right), \tag{31b}$$

$$\frac{dx_{010}}{dt} = +a_5\left([Ca]_{cyt}x_{000} - d_5 x_{010}\right) - a_4\left([Ca]_{cyt}x_{010} - d_4 x_{011}\right), \tag{31c}$$

$$x_{011} = 1 - (x_{000} + x_{001} + x_{010} + x_{100} + x_{101} + x_{110} + x_{111}). \tag{31d}$$

Intracellular Ca^{2+} Dynamics

Calcium Buffers

Calcium buffering by cytosolic Ca^{2+}-binding proteins other than calmodulin is described as a first-order dynamic process,

$$\frac{d[\mathrm{Bf}\cdot\mathrm{Ca}]_{\mathrm{cyt}}}{dt} = k_{\mathrm{on}}^{\mathrm{Bf}}[\mathrm{Ca}]_{\mathrm{cyt}}\left([\mathrm{Bf}]_{\mathrm{cyt}}^{\mathrm{tot}} - [\mathrm{Bf}\cdot\mathrm{Ca}]_{\mathrm{cyt}}\right) - k_{\mathrm{off}}^{\mathrm{Bf}}[\mathrm{Bf}\cdot\mathrm{Ca}]_{\mathrm{cyt}}, \tag{32}$$

where $[\mathrm{Bf}]_{\mathrm{cyt}}^{\mathrm{tot}}$ is the total concentration of Ca^{2+}-binding proteins other than calmodulin in the cytosol and $[\mathrm{Bf}\cdot\mathrm{Ca}]_{\mathrm{cyt}}$ is the concentration of the calcium-bound sites of these other buffering elements.

Calcium buffering by calsequestrin in the SR is described as

$$\frac{d[\mathrm{Calseq}\cdot\mathrm{Ca}]_{\mathrm{SR}}}{dt} = k_{\mathrm{on}}^{\mathrm{Calseq}}[\mathrm{Ca}]_{\mathrm{SR}}\left([\mathrm{Calseq}]_{\mathrm{SR}}^{\mathrm{tot}} - [\mathrm{Calseq}\cdot\mathrm{Ca}]_{\mathrm{SR}}\right) - k_{\mathrm{off}}^{\mathrm{Calseq}}[\mathrm{Calseq}\cdot\mathrm{Ca}]_{\mathrm{SR}}, \tag{33}$$

where $[\mathrm{Calseq}]_{\mathrm{SR}}^{\mathrm{tot}}$ is the total concentration of calsequestrin sites available for Ca^{2+} binding in the SR and $[\mathrm{Calseq}\cdot\mathrm{Ca}]_{\mathrm{SR}}$ is the concentration of Ca^{2+}-bound calsequestrin sites in that compartment. Parameter values are given in Table 6.

Kinetics of Myosin Light Chain Phosphorylation

CaM Activation of MLCK

Calmodulin (CaM) has four Ca^{2+} binding sites, with two at the NH_2 terminus (low affinity) and two at the COOH terminus (high affinity). Binding of Ca^{2+} to those sites yields the $CaM.Ca_4$ complex, and $CaM.Ca_4$ binds to myosin light chain kinase (MLCK) to form $MLCK.CaM.Ca_4$, which is the active form of MLCK that phosphorylates MLCs.

The scheme proposed by Fajmut et al. [7] is used to determine the kinetics of formation of $MLCK.CaM.Ca_4$. Subscripts *N* and *C* represent two binding sites each for Ca^{2+} at the NH_2 and COOH terminus of CaM, respectively, and the subscript *M* represents the CaM binding site occupied by MLCK. An underscore (_) denotes an unoccupied site for each of these binding sites. For example, $\mathrm{CaM_{NCM}}$ designates $MLCK.CaM.Ca_4$. The kinetic equations for the formation of MLCK are

$$\frac{d[\mathrm{CaM_{-C-}}]}{dt} = \left(-k_{-1}^{\mathrm{CaM}} - k_{4}^{\mathrm{CaM}}[\mathrm{Ca}]^2 - k_{5}^{\mathrm{CaM}}[\mathrm{MLCK}]_{\mathrm{free}}\right)[\mathrm{CaM_{-C-}}] + k_{1}^{\mathrm{CaM}}[\mathrm{Ca}]^2[\mathrm{CaM_{---}}] + k_{-4}^{\mathrm{CaM}}[\mathrm{CaM_{NC-}}] + k_{-5}^{\mathrm{CaM}}[\mathrm{CaM_{-CM}}], \tag{34a}$$

$$\frac{d[\mathrm{CaM_{N--}}]}{dt} = \left(-k_{-2}^{\mathrm{CaM}} - k_{3}^{\mathrm{CaM}}[\mathrm{Ca}]^2\right)[\mathrm{CaM_{N--}}] + k_{2}^{\mathrm{CaM}}[\mathrm{Ca}]^2[\mathrm{CaM_{---}}] + k_{-3}^{\mathrm{CaM}}[\mathrm{CaM_{NC-}}], \tag{34b}$$

$$\frac{d[\mathrm{CaM_{NC-}}]}{dt} = \left(-k_{-3}^{\mathrm{CaM}} - k_{-4}^{\mathrm{CaM}} - k_{7}^{\mathrm{CaM}}[\mathrm{MLCK}]_{\mathrm{free}}\right)[\mathrm{CaM_{NC-}}] + k_{3}^{\mathrm{CaM}}[\mathrm{Ca}]^2[\mathrm{CaM_{N--}}] + k_{4}^{\mathrm{CaM}}[\mathrm{Ca}]^2[\mathrm{CaM_{-C-}}] + k_{-7}^{\mathrm{CaM}}[\mathrm{Ca}]^2[\mathrm{CaM_{NCM}}], \tag{34c}$$

$$\frac{d[\mathrm{CaM_{-CM}}]}{dt} = \left(-k_{-5}^{\mathrm{CaM}} - k_{6}^{\mathrm{CaM}}[\mathrm{Ca}]^2\right)[\mathrm{CaM_{-CM}}] + k_{5}^{\mathrm{CaM}}[\mathrm{MLCK}]_{\mathrm{free}}[\mathrm{CaM_{-C-}}] + k_{-6}^{\mathrm{CaM}}[\mathrm{Ca}]^2[\mathrm{CaM_{NCM}}], \tag{34d}$$

$$\frac{d[\mathrm{CaM_{NCM}}]}{dt} = \left(-k_{-6}^{\mathrm{CaM}} - k_{7}^{\mathrm{CaM}}[\mathrm{Ca}]^2\right)[\mathrm{CaM_{NCM}}] + k_{6}^{\mathrm{CaM}}[\mathrm{Ca}]^2[\mathrm{CaM_{-CM}}] + k_{7}^{\mathrm{CaM}}[\mathrm{MLCK}]_{\mathrm{free}}[\mathrm{CaM_{NC-}}], \tag{34e}$$

$$[\mathrm{CaM}]^{\mathrm{tot}} = [\mathrm{CaM_{---}}] + [\mathrm{CaM_{-C-}}] + [\mathrm{CaM_{N--}}] + [\mathrm{CaM_{NC-}}], + [\mathrm{CaM_{-CM}}] + [\mathrm{CaM_{NCM}}], \tag{34f}$$

$$[\mathrm{MLCK}]^{\mathrm{tot}} = [\mathrm{CaM_{-CM}}] + [\mathrm{CaM_{NCM}}] + [\mathrm{MLCK}]_{\mathrm{free}}, \tag{34g}$$

where the on- and off-rate constants are denoted by k_i^{CaM} and k_{-i}^{CaM}, respectively, $[\mathrm{MLCK}]_{\mathrm{free}}$ is the concentration of free (unbound) MLCK, $[\mathrm{CaM}]^{\mathrm{tot}}$ is the total concentration of calmodulin, $[\mathrm{MLCK}]^{\mathrm{tot}}$ is the total concentration of MLCK, and the subscript "cyt" that denotes the cytosolic compartment is omitted for simplicity. Parameter values are given in Table 7.

The buffering terms in the Ca^{2+} conservation equations (12d) and (13) are given by

$$R_{\mathrm{CaM}}^{\mathrm{cyt}} = -2\frac{d[\mathrm{CaM_{N--}}]}{dt} - 2\frac{d[\mathrm{CaM_{-C-}}]}{dt} - 2\frac{d[\mathrm{CaM_{-CM}}]}{dt} - 4\frac{d[\mathrm{CaM_{NC-}}]}{dt} - 4\frac{d[\mathrm{CaM_{NCM}}]}{dt}, \tag{35a}$$

$$R_{\mathrm{Bf}}^{\mathrm{cyt}} = -\frac{d[\mathrm{Bf}\cdot\mathrm{Ca}]_{\mathrm{cyt}}}{dt}, \tag{35b}$$

$$R_{\mathrm{Calseq}}^{\mathrm{SR}} = -\frac{d[\mathrm{Calseq}\cdot\mathrm{Ca}]_{\mathrm{SR}}}{dt}. \tag{35c}$$

Table 7 Parameters for MLCK and MLCP kinetics

Parameter	Value	Unit
k_1^{CaM}	2.8×10^6	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
k_{-1}^{CaM}	6	s^{-1}
k_2^{CaM}	10^8	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
k_{-2}^{CaM}	800	s^{-1}
k_3^{CaM}	2.8×10^6	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
k_{-3}^{CaM}	6	s^{-1}
k_4^{CaM}	10^8	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
k_{-4}^{CaM}	800	s^{-1}
k_5^{CaM}	10^9	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
k_{-5}^{CaM}	20	s^{-1}
k_6^{CaM}	1.25×10^7	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$
k_{-6}^{CaM}	5	s^{-1}
k_7^{CaM}	10^9	$\mathrm{mM}^{-1} \cdot \mathrm{s}^{-1}$

Parameter	Value	Unit
k_{-7}^{CaM}	1	s^{-1}
$[\mathrm{CaM}]^{\mathrm{tot}}$	10×10^{-3}	mM
$[\mathrm{MLCK}]^{\mathrm{tot}}$	2×10^{-3}	mM
$[\mathrm{MLCP}]^{\mathrm{tot}}$	2×10^{-3}	mM
[RhoK]	30×10^{-6}	mM
$k_{\mathrm{cat}}/k_+^{\mathrm{MLCP}}$	0.33×10^6	mM^{-1}
$k_{\mathrm{MLCK}}^{\mathrm{Myo}}$	0.537	s^{-1}
$k_{\mathrm{MLCP}}^{\mathrm{Myo}}$	1.62	s^{-1}
k_3^{Myo}	1.8	s^{-1}
k_4^{Myo}	0.1	s^{-1}
k_7^{Myo}	0.045	s^{-1}
$[\mathrm{Myo}]^{\mathrm{tot}}$	30×10^{-3}	mM

Rho-Kinase Inhibition of MLCP

Myosin light chain phosphatase (MLCP) consists of three subunits, one of which, MYPT1, can be phosphorylated by Rho kinase (RhoK). Rho-K-induced phosphorylation of MYPT1 inactivates MLCP, which promotes contraction. The cytosolic concentration of active MLCP (denoted MLCP*) is given by

$$\frac{d[\mathrm{MLCP}^*]}{dt} = k_+^{\mathrm{MLCP}} \left([\mathrm{MLCP}]^{\mathrm{tot}} - [\mathrm{MLCP}^*] \right) - k_-^{\mathrm{MLCP}} [\mathrm{MLCP}^*], \tag{36}$$

where $[\mathrm{MLCP}]^{\mathrm{tot}}$ is the total concentration of MLCP in the cytosol and the inactivation of MLCP by RHoK is given by

$$k_-^{\mathrm{MLCP}} = k_{\mathrm{cat}} [\mathrm{RhoK}] \tag{37}$$

(Mbikou et al. [13]). The concentration of RhoK, [RhoK], is assumed to be fixed at 30 nM (Kaneko-Kawano et al. [9]) except in the presence of specific inhibitors. Parameter values are given in Table 7.

MLCK- and MLCP-Dependent Phosphorylation of Myosin

The contractile force of the vessels is determined by the fraction of myosin cross-bridges that are phosphorylated. The four types of cross-bridges considered are free cross-bridges (Myo), phosphorylated cross-bridges (MyoP), attached phosphorylated cycling cross-bridges (AMyoP), and attached dephosphorylated, non-cycling cross-bridges (AMyo), and the corresponding equations for their concentrations are

$$\frac{d[\text{Myo}]}{dt} = -k_1^{\text{Myo}}[\text{Myo}] + k_2^{\text{Myo}}[\text{MyoP}] + k_7^{\text{Myo}}[\text{AMyo}], \tag{38a}$$

$$\frac{d[\text{MyoP}]}{dt} = +k_1^{\text{Myo}}[\text{Myo}] - (k_2^{\text{Myo}} + k_3^{\text{Myo}})[\text{MyoP}] + k_4^{\text{Myo}}[\text{AMyoP}], \tag{38b}$$

$$\frac{d[\text{AMyoP}]}{dt} = +k_3^{\text{Myo}}[\text{MyoP}] - (k_4^{\text{Myo}} + k_5^{\text{Myo}})[\text{AMyoP}] + k_6^{\text{Myo}}[\text{AMyo}], \tag{38c}$$

$$[\text{Myo}]^{\text{tot}} = [\text{Myo}] + [\text{MyoP}] + [\text{AMyoP}] + [\text{AMyo}]. \tag{38d}$$

The rate constants k_3^{Myo}, k_4^{Myo}, and k_7^{Myo} are fixed (Yang et al. [19]). Parameter values are given in Table 7.

The rate constants k_1^{Myo} and k_6^{Myo} represent the activity of MLCK and are assumed to be proportional to the fraction of the fully activated form of the enzyme, while the rate constants k_2^{Myo} and k_5^{Myo} represent the activity of MLCP. The corresponding equations are

$$k_1^{\text{Myo}} = k_6^{\text{Myo}} = k_{\text{MLCK}}^{\text{Myo}} \frac{[\text{CaM}_{\text{NCM}}]}{[\text{CaM}]^{\text{tot}}}, \tag{39a}$$

$$k_2^{\text{Myo}} = k_5^{\text{Myo}} = k_{\text{MLCP}}^{\text{Myo}} \frac{[\text{MLCP}^*]}{[\text{MLCP}]^{\text{tot}}}, \tag{39b}$$

where $k_{\text{MLCK}}^{\text{Myo}}$ and $k_{\text{MLCP}}^{\text{Myo}}$ are fixed.

Mechanical Behavior of Cell

The vasomotion of the afferent arteriole is affected by the variations in the number of cross-bridges, which induce variations in the contractile force and thus alter the diameter of the vessel. Edwards and Layton [5] implemented the model of Carlson et al. [1] that represents vessel wall tension as the sum of a passive component and an active myogenic component. The passive component, T_{pass}, is a function of the vessel diameter, D,

$$T_{\text{pass}} = C_{\text{pass}} \exp\left[C'_{\text{pass}}\left(\frac{D}{D_0} - 1\right)\right], \tag{40}$$

where D_0 is the reference vessel diameter.

The active myogenic component is the product of the maximal active tension generated at a given vessel circumference, $T_{\text{act}}^{\max}$, given by

$$T_{\text{act}}^{\max} = C_{\text{act}} \exp\left[-\left(\frac{D/D_0 - C'_{\text{act}}}{C''_{\text{act}}}\right)^2\right], \tag{41}$$

Table 8 Parameters for smooth muscle cell mechanics

Parameter	Value	Unit
C_{pass}	223	dyn/cm
C'_{pass}	20.2	Dimensionless
C_{act}	500	dyn/cm
C'_{act}	0.985	Dimensionless
C''_{act}	0.500	Dimensionless
D_0	27.5	μm
D_{c}	20.0	μm
τ_{d}	1.71	s

and the fraction of myosin light chains that are phosphorylated, Ψ. Therefore, the total tension in the wall, T_{wall}, is

$$T_{\text{wall}} = T_{\text{pass}} + \Psi T_{\text{act}}^{\max}. \tag{42}$$

The change in vessel diameter depends on the difference between the tension resulting from intravascular pressure p, $T_{\text{pres}} = pD/2$, and the tension generated in the wall, T_{wall}, so that

$$\frac{dD}{dt} = \frac{1}{\tau_{\text{d}}} \frac{D_{\text{c}}}{T_{\text{c}}} (T_{\text{pres}} - T_{\text{wall}}), \tag{43}$$

where D_{c} is a reference diameter, T_{c} is the tension at a pressure of 100 mmHg and diameter D_{c}, and τ_{d} is a time constant. Parameter values are given in Table 8.

Acknowledgements This work is the product of a workshop and short-term visits supported by the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from the University of Tennessee, Knoxville. Support was also provided by the National Institutes of Health: National Institute of Diabetes and Digestive and Kidney Diseases and by the National Science Foundation, via grants #DK089066 and #DMS-1263995 to AT Layton.

References

1. Carlson, B.E., Arciero, J.C., Secomb, T.W.: Theoretical model of blood flow autoregulation: roles of myogenic, shear-dependent, and metabolic responses. Am. J. Physiol. Heart Circ. Physiol. **295**, H1572–H1579 (2008)
2. De Young, G.W., Keizer, J.: A single-pool inositol 1,4,5-triphosphate-receptor-based model for agonist-stimulated oscillations in Ca^{2+} concentrations. Proc. Natl. Acad. Sci. U. S. A. **89**, 9895–9899 (1992)
3. Deen, W., Robertson, C., Brenner, B.: A model of glomerular ultrafiltration in the rat. Am. J. Physiol. **223**(5), 1178–1183 (1972)
4. Eaton, D., Pooler, J.: Vander's Renal Physiology, 6th edn. McGraw-Hill Medical, New York (2004)

5. Edwards, A., Layton, A.: Calcium dynamics underlying the afferent arteriole myogenic response. Am. J. Physiol. Ren. Physiol. **306**, F34–F48 (2014)
6. Faber, G.M., Silva, J., Livshitz, L., Rudy, Y.: Kinetic properties of the cardiac L-type Ca^{2+} channel and its role in myocyte electrophysiology: a theoretical investigation. Biophys. J. **92**, 1522–1543 (2007)
7. Fajmut, A., Jagodič, M., Brumen, M.: Mathematical modeling of the myosin light chain kinase activation. J. Chem. Inf. Model. **45**, 1605–1609 (2005)
8. Jacobsen, J.C.B., Aalkjær, C., Nilsson, H., Matchkov, V.V., Freiberg, J., Holstein-Rathlou, N.H.: Activation of a cGMP-sensitive calcium-dependent chloride channel may cause transition from calcium waves to whole cell oscillations in smooth muscle cells. Am. J. Physiol. Heart Circ. Physiol. **293**, H215–H228 (2007)
9. Kaneko-Kawano, T., Takasu, F., Naoki, H., Sakumura, Y., Ishii, S., Ueba, T., Eiyama, A., Okada, A., Kawano, Y., Suzuki, K.: Dynamic regulation of myosin light chain phosphorylation by Rho-kinase. PLoS One **7**, e39269 (2012)
10. Keizer, J., Levine, L.: Ryanodine receptor adaptation and Ca^{2+}-induced Ca^{2+} release dependent Ca^{2+} oscillations. Biophys. J. **71**, 3477–3487 (1996)
11. Kneller, J., Ramirez, R.J., Chartier, D., Courtemanche, M., Nattel, S.: Time-dependent transients in an ionically based mathematical model of the canine atrial action potential. Am. J. Physiol. Heart Circ. Physiol. **282**, H1437–H1451 (2002)
12. Layton, A.: Feedback-mediated dynamics in a model of a compliant thick ascending limb. Math. Biosci. **228**, 185–194 (2010)
13. Mbikou, P., Fajmut, A., Brumen, M., Roux, E.: Contribution of Rho kinase to the early phase of the calcium-contraction coupling in airway smooth muscle. Exp. Physiol. **96**, 240–258 (2011)
14. Schnermann, J., Briggs, J.: Function of the juxtaglomerular apparatus: control of glomerular hemodynamics and renin secretion. In: Alpern, R.J., Hebert, S.C. (eds.) Seldin and Giebisch's The Kidney: Physiology and Pathophysiology, 4th edn., pp. 589–626. Elsevier Academic Press, Amsterdam, Boston (2008)
15. Sgouralis, I., Layton, A.: Autoregulation and conduction of vasomotor responses in a mathematical model of the rat afferent arteriole. Am. J. Physiol. Ren. Physiol. **303**, F229–F239 (2012)
16. Sgouralis, I., Layton, A.: Theoretical assessment of renal autoregulatory mechanisms. Am. J. Physiol. Ren. Physiol. **306**, F1357–F1371 (2014)
17. Shannon, T.R., Wang, F., Puglisi, J., Weber, C., Bers, D.M.: A mathematical treatment of integrated Ca dynamics within the ventricular myocyte. Biophys. J. **87**, 3351–3371 (2004)
18. Wu, B.N., Luykenaar, K.D., Brayden, J.E., Giles, W.R., Corteling, R.L., Wiehler, W.B., Welsh, D.G.: Hyposmotic challenge inhibits inward rectifying K^+ channels in cerebral arterial smooth muscle cells. Am. J. Physiol. Heart Circ. Physiol. **292**, H1085–H1094 (2007)
19. Yang, J., Clark Jr., J.W., Bryan, R.M., Robertson, C.: The myogenic response in isolated rat cerebrovascular arteries: smooth muscle cell model. Med. Eng. Phys. **25**, 691–709 (2003)

Modeling Blood Flow and Oxygenation in a Diabetic Rat Kidney

Ioannis Sgouralis and Anita T. Layton

Abstract We use a highly detailed mathematical model of renal hemodynamics and solute transport to simulate medullary oxygenation in the kidney of a diabetic rat. Model simulations suggest that alterations in renal hemodynamics, which include diminished vasoconstrictive response of the afferent arteriole as a major factor, lead to glomerular hyperfiltration in diabetes. The resulting higher filtered Na^+ load increases the reabsorptive work of the nephron, but by itself does not significantly elevate medullary oxygen consumption. The key explanation for diabetes-related medullary hypoxia may be impaired renal metabolism. Tubular transport efficiency is known to be reduced in diabetes, leading to increased medullary oxygen consumption, despite relatively unchanged active Na^+ transport. The model predicts that interstitial fluid oxygen tension of the inner stripe, which is a particularly oxygen-poor region of the medulla, decreases by 18.6% in a diabetic kidney.

1 Introduction

Diabetes has become increasingly prevalent in the developed world. In 2012, almost 30 million Americans, or ∼10% of the population, had diabetes. Diabetes is one of the most common causes of end-stage renal diseases [12]. But despite intense research, the underlying mechanisms remain incompletely understood. It is generally believed that renal hypoxia is an important pathway in the development of chronic kidney disease. Typically, renal hypoxia is caused by a mismatch between changes in renal oxygen delivery and consumption [10]. Thus, to assess the degree of hypoxia under a given set of circumstances, one would consider the balance between renal oxygen delivery and consumption. Renal oxygen delivery is primarily

I. Sgouralis (✉)
NIMBioS, University of Tennessee Knoxville, Knoxville, TN 37996, USA
e-mail: sgouralis@nimbios.org

A.T. Layton
Department of Mathematics, Duke University, Durham, NC 27708, USA
e-mail: alayton@math.duke.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_6

determined by renal blood flow [10]. Renal oxygen consumption is mainly driven by the metabolic work of tubular Na^+ reabsorption, which in turn is largely driven by the filtered Na^+ load, and thus glomerular filtration rate (GFR) [11].

In diabetes, several pathophysiological processes lead to a significant increase in renal oxygen consumption. First, intrarenal angiotensin II is elevated, which via the activation of NADPH oxidase leads to increased oxidative stress [16]. This results in increased mitochondrial uncoupling via UCP-2, which stimulates O_2^- formation but decreases transport efficiency and increases oxygen consumption. Furthermore, proximal tubule reabsorption is enhanced in hyperglycemia due to increased glomerular filtration of glucose [32]. In addition, tubular growth occurs shortly after the onset of diabetes, leading to a further increase in proximal tubule reabsorption and to glomerular hyperfiltration [31]. It is noteworthy that glomerular hyperfiltration and kidney growth are both risk factors for the development of diabetic nephropathy [31, 33]. Proximal tubule reabsorption is further enhanced in hyperglycemia due to increased glomerular filtration of glucose, which increases reabsorption of both glucose and Na^+ through the cotransporters SGLT2 and SGLT1. Without a compensatory increase in oxygen delivery, these events may result in intrarenal tissue hypoxia and the development of clinical hallmarks of diabetic nephropathy.

The goal of this study is to use a detailed mathematical model of solute transport and blood flow control to study intrarenal oxygenation in a diabetic rat kidney.

2 Mathematical Model

The model is based on two of our previously applied models that represent different aspects of renal function: (1) A model of the medulla of the rat kidney that describes the urine concentrating mechanism and oxygen transport; this model encompasses the tubular and vascular architecture of both the outer and inner medullas [14]. (2) A model of the myogenic response and tubuloglomerular feedback in the afferent arteriole that regulates renal blood flow [26]; the afferent arteriole is located in the cortex. To connect these two components, the present model includes a proximal tubule segment that exhibits pressure-dependent reabsorption of the filtered load, and a cortical thick ascending limb segment that terminates at the site of macula densa and provides the input signal for the tubuloglomerular feedback response [23]. Model structure is shown in Fig. 1.

The medullary component of the model represents loops of Henle, collecting ducts, and vasa recta as rigid tubes. Two-thirds of the model loops of Henle are "short loops" that are assumed to turn at the outer–inner medullary boundary, while the other one-third, the "long loops," turn at all depths of the inner medulla. The model medullary collecting ducts extend from the corticomedullary boundary to the papillary tip, and exhibit coalescence in the inner medulla. The vasa recta terminate or originate at all depths of the medulla, and are assumed to peel off and supply the capillary plexus. The model separates blood flow in the vasa recta into two compartments—plasma and red blood cells (RBCs)—which are both also represented by rigid tubes along the corticomedullary axis.

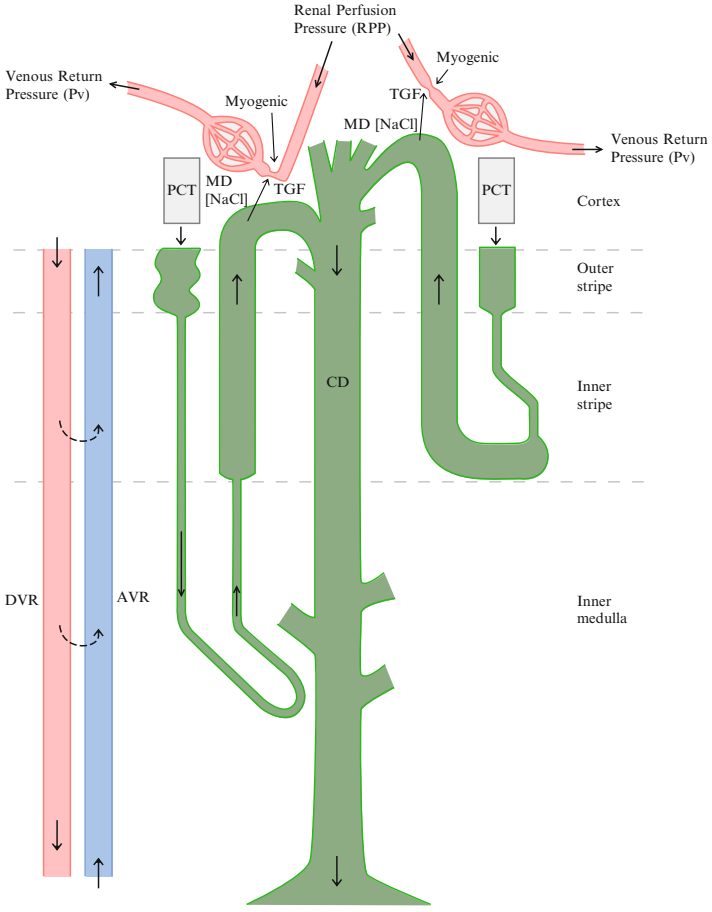


Fig. 1 An overview of the vascular and tubular structure of the model. For simplicity, only one short and one long loop of Henle are depicted. Reprinted from [23]

The model predicts fluid flows in the tubules and vessels, as well as concentrations of NaCl, urea, O_2, NO, O_2^-, deoxy-hemoglobin (Hb), oxy-hemoglobin (HbO_2), and nitrosyl-heme (HbNO), in each tubule and vessel, as well as in the interstitium. Model equations are based on transmural transport and conservation of water and solutes. Below we summarize key components of the model that represent conservation of water and solute and consumption of oxygen. Full model equations can be found in [5, 9, 21].

2.1 Renal Autoregulation

Here we describe how the model represents renal autoregulation. The key model component is the afferent arteriole, which is located in the cortex. The afferent arteriole exhibits autoregulatory responses mediated by two mechanisms: the myogenic response (MR) and tubuloglomerular feedback (TGF). Each model afferent arteriole delivers blood to its associated glomerulus. Model equations that represent glomerular filtration and post-glomerular blood flow can be found in [26].

The model afferent arteriole adopts our previously applied model of renal autoregulation [25, 26]. The vessel is represented as a series of smooth muscle cells, coupled via their gap junction. The model predicts arteriolar vasoresponses by representing, for each smooth muscle cell, its membrane potential, Ca^{2+} and K^+ transmembrane fluxes, cytosolic Ca^{2+} handling, and muscle mechanics. We describe key model equations; for the full set of model equations see [25, 26].

For each smooth muscle cell, the rate of change of cytosolic $[Ca^{2+}]$ is given by

$$\frac{d[Ca^{2+}]}{dt} = \left(\frac{(K_d + [Ca^{2+}])^2}{(K_d + [Ca^{2+}])^2 + K_d B_T} \right) \left(\alpha I_{Ca} - k_{Ca}[Ca^{2+}] \right) \tag{1}$$

where I_{Ca} is the Ca^{2+} influx from the extracellular space (see below) and α is a constant converting charge to mass flux; k_{Ca} is a first-order rate constant for Ca^{2+} extrusion; K_d is the ratio of the forward and backward reaction rates of the calcium-buffer system; and B_T is the total buffer concentration.

Calcium influx is provided by voltage-gated membrane channels

$$I_{Ca} = g_{Ca} m_\infty (v - v_{Ca}) \tag{2}$$

where g_{Ca} is the maximum whole-cell membrane conductance for the calcium current, and m_∞ is the equilibrium distribution of open calcium channel states which is a function of the membrane potential (v)

$$m_\infty(v) = \frac{1}{2} \left(1 + \tanh \left(\frac{v - v_1}{v_2} \right) \right) \tag{3}$$

where v_1 is the voltage at which half of the channels are open, and v_2 determines the spread of the distribution.

For each afferent arteriole smooth muscle cell, the opening of Ca^{2+} channels (and that of K^+ channels, represented in the model but not described here) depends on the membrane potential v [Eq. (3)], which is in turn given by the sum of the transmembrane currents:

$$C \frac{dv}{dt} = -I_L - I_K - I_{Ca} + I_{MR} + I_{TGF} + I_{coupling} \tag{4}$$

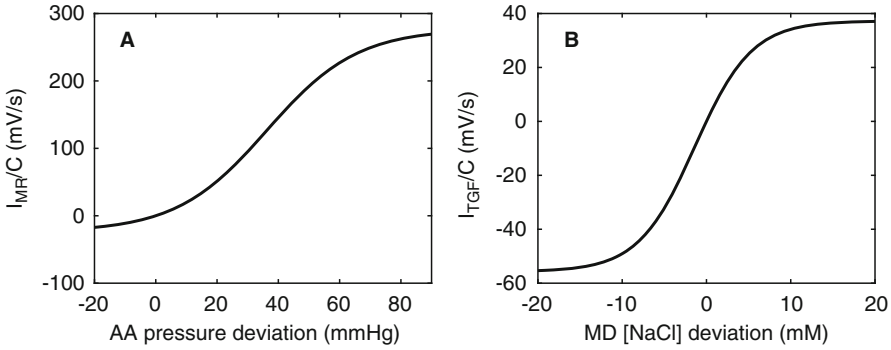


Fig. 2 Afferent arteriole (AA) smooth muscle transmembrane currents mediating the myogenic response (panel **a**) and tubuloglomerular feedback (panel **b**). The currents are normalized by smooth muscle membrane capacitance *C*. Elevated arteriolar blood pressure (panel **a**) or variations in macula densa (MD) [NaCl] (panel **b**) induce depolarization of the smooth muscles, leading to changes in muscle tone

where C denotes the capacitance of the cell membrane. The transmembrane currents include the leak current (I_{L}), K^+ current (I_{K}), Ca^{2+} current (I_{Ca}), and gap-junction current between adjacent smooth muscle cells (I_{coupling}).

The representation of the myogenic response is based on the hypothesis that changes in luminal blood pressure induce changes in the activity of non-selective cation channels, resulting in a transmembrane current I_{MR}. I_{MR} is chosen so that an increase in luminal blood pressure results in membrane depolarization, Fig. 2a, which increases intracellular [Ca^{2+}], leading to enhanced cross-bridge formation (see below) and vasoconstriction. Conversely, a decrease in pressure leads to membrane polarization and vasodilation.

The TGF response is modeled by I_{TGF}, which is a function of the macula densa [NaCl] (i.e., tubular fluid [NaCl] at the outflow of the thick ascending limbs). Deviation in macula densa [NaCl] from its operating point induces a TGF response, Fig. 2b, such that an elevated macula densa [NaCl] results in membrane depolarization and vasoconstriction, and vice versa. The TGF current is only applied to the distal smooth muscle cells (i.e., those proximal to the glomerulus).

Cytosolic [Ca^{2+}] determines the formation of crossbridges, which in turn determines the muscle tone [6]. Muscle tone, together with luminal pressure, gives rise to hoop forces, the balance of which determines the luminal radius; for details see [26].

2.2 Solute Conservation

In the glomerulus, a fraction of the supplied blood is filtered through the capillaries to feed the associated nephron. Along the nephron, water and solutes are reabsorbed or secreted; as a result, tubular solute concentrations are adjusted. To predict solute

concentrations and flows along the different segments of the nephron, the model represents conservation of each solute k in each tubule or vas rectum, indexed i. At steady state, solute conservation is given by

$$\frac{\partial}{\partial x}\left(Q_i C_{i,k}\right) = 2\pi r_i J_{i,k} + R_{i,k} \tag{5}$$

where Q_i denotes volume flow, $C_{i,k}$ denotes the concentration of solute k, r_i denotes luminal radius, $J_{i,k}$ denotes transmural solute fluxes, and $R_{i,k}$ denotes reactions (set to 0 for non-reactive solutes such as NaCl and urea). Components for Na^+ flux and O_2 consumption are given below. Full expressions for $J_{i,k}$ and $R_{i,k}$ can be found in [5, 13, 14, 19].

2.3 Oxygen Consumption

Because the present study focuses on medullary oxygenation, we describe how oxygen consumption is represented. The model considers two types of O_2 consumption: *basal consumption* and *active consumption* due to active Na^+ transport. In the model, basal O_2 consumption (i.e., for processes other than Na^+ transport) is assumed to take place in the vascular endothelial cells, tubular epithelial cells, and interstitial cells. The volumetric rate of basal O_2 consumption in tubule or vas rectum i is given by

$$R^{\text{basal}}_{i,\text{O}_2} = R^{\text{basal}}_{\text{max,O}_2}\left[\frac{C_{i,\text{O}_2}(x)}{K_{\text{M, O}_2} + C_{i,\text{O}_2}(x)}\right], \tag{6}$$

where C_{i,O_2} is the O_2 concentration in tubule or vessel i, $K_{\text{M, O}_2}$ is the Michaelis constant that depends on local NO concentration (see [14]), and $R^{\text{basal}}_{\text{max,O}_2}$ is the maximal volumetric rate of O_2 consumption. Here, $R^{\text{basal}}_{\text{max,O}_2}$ is taken to be 10 μM/s [5]. $R^{\text{basal}}_{i,\text{O}_2}$ contributes to the reaction term R_{i,O_2} in Eq. (5).

In the proximal straight tubules, thick ascending limbs, and collecting ducts, Na^+ is actively reabsorbed via basolateral Na^+-K^+-ATPase pumps, resulting in active O_2 consumption. The volumetric active Na^+ transport rate is modeled assuming Michaelis–Menten kinetics

$$\Psi^{\text{active}}_{i,\text{Na}} = V_{\text{max},i,\text{Na}}\left[\frac{C_{i,\text{Na}}(x)}{K_{\text{M,Na}} + C_{i,\text{Na}}(x)}\right], \tag{7}$$

where $C_{i,\text{Na}}$ is the Na^+ concentration in tubule i, $V_{\text{max},i,\text{Na}}$ is the maximal rate of Na^+ transport, and $K_{\text{M,Na}}$ is the Michaelis constant (taken here to be 70 mM [15, 20]). When luminal P_{O2} drops below some critical level, the maximum rate of Na^+ transport may become limited by O_2 concentration. In addition, the model represents the inhibition and stimulation of Na^+ transport by NO and O_2^-, respectively; i.e., $V_{\text{max},i,\text{Na}}$ depends on local NO and O_2^- concentrations (see [14] for details). $\Psi^{\text{active}}_{i,\text{Na}}$ contributes to the flux term $J_{i,\text{Na}}$ in Eq. (5).

To relate active Na^+ transport to O_2 consumption, we must define the T_{Na}-to-Q_{O_2} ratio (or, TQ ratio, where T_{Na} denotes the moles of Na^+ transported, and Q_{O_2} denotes the amount of O_2 consumed). The stoichiometry of metabolic reactions suggests that under maximal efficiency, the number of Na^+ moles actively reabsorbed per mole of O_2 consumed is 18. That is, the TQ ratio is 18 for active transport. Under favorable thermodynamic conditions, additional moles of Na^+ may be reabsorbed paracellularly, which would increase the TQ ratio; on the other hand, unfavorable thermodynamic conditions would decrease the TQ ratio. In the baseline (i.e., non-diabetic) model, TQ ratios are fixed at 18 for the thick ascending limb and proximal straight tubule, and 12 for the collecting duct, based on tubular epithelial transport simulations [24, 35]. We further assume that below some critical P_{O2} (denoted $P_{i,c}$ for tubule i), anaerobic metabolism maintains some fraction (denoted F_{AN}) of the energy required to actively transport Na^+. The rate of active O_2 consumption in the epithelia of the proximal straight tubules, thick ascending limbs, and collecting ducts is then given by

$$R_{i,\mathrm{O_2}}^{\mathrm{active}} = \frac{2\pi r_i}{A_i^{\mathrm{epi}}} \frac{\Psi_{i,\mathrm{Na}}^{\mathrm{active}}}{\mathrm{TQ}_i} \Theta(P_{i,\mathrm{O_2}}) \tag{8}$$

where r_i is the inner radius of tubule i. $R_{i,\mathrm{O_2}}^{\mathrm{acative}}$ contributes to the reaction term $R_{i,\mathrm{O_2}}$ in Eq. (5). TQ_i is the TQ ratio, $\Psi_{i,\mathrm{Na}}^{\mathrm{active}}$ is the Na^+ active transport rate, and $\Theta(P_{i,\mathrm{O_2}})$ is the fraction of that rate that is supported by aerobic respiration, given by

$$\Theta(P_{i,\mathrm{O_2}}) = \begin{cases} 1, & \mathrm{P_{O2}} \geq P_{i,c} \\ \frac{\mathrm{P_{O2}}/P_{i,c}}{F_{AN}+(1-F_{AN})(\mathrm{P_{O2}}/P_{i,c})}, & \mathrm{P_{O2}} < P_{i,c} \end{cases} \tag{9}$$

We assume that $P_{i,c} = 10\,\mathrm{mmHg}$ [13] in all tubules. The value of F_{AN} is taken to be 0.5 in the outer medullary collecting duct [29, 36], 0.4 in the inner medullary collecting duct [27], 0.1 in the thick ascending limb [1], and 0.14 in the proximal straight tubule [8].

2.4 Modeling a Diabetic Kidney

Our previous models [14, 26] are formulated for a baseline, or non-diabetic, kidney. To simulate a diabetic kidney, we incorporate the functional impairment in afferent arteriole voltage-gated Ca^{2+} channels (VGCC), observed in diabetic rats [4], which is believed to diminish the vasoconstrictive response, thereby causing glomerular hyperfiltration. Recall that the model represents the voltage-dependent distribution of open Ca^{2+} channel states using Eq. (3). In the baseline (i.e., non-diabetic) model the voltage at which half of VGCC channels are open in the unimpaired state (v_1) is set to -22.5 mV. To represent VGCC impairment, we increase v_1 to -21 mV.

At the onset of diabetes, the proximal tubule is known to hypertrophize [30]. As a result, reabsorption along the proximal tubule increases, a process which we simulate by increasing fractional proximal tubule water reabsorption to 0.81 (compare with baseline 0.67). Diabetes is also associated with TGF resetting [28, 33], which we represent by lowering the operating macula densa [NaCl] from their baseline values of 32 and 42 mM (short and long nephrons, respectively) to 25 and 33 mM, respectively. Finally, the ultrafiltration coefficients K_f of all nephrons are increased by 20% from their baseline values [2].

To model impairment in metabolic efficiency in a diabetic kidney, the TQ ratio is assumed to decrease by 20% in all tubules.

3 Model Results

3.1 Renal Autoregulation in Diabetes

To assess the model's ability to stabilize SNGFR (single-nephron glomerular filtration rate) in health and disease, we computed afferent arteriole blood flow and SNGFR for a range of steady renal perfusion pressure, from 60 to 200 mmHg. Simulations were conducted for the baseline model and for the diabetic model. These results are shown in Fig. 3 for the short loop (solid curves) and for the longest loop (i.e., the model loop of Henle that reaches to the papillary tip, dashed curves).

For the baseline model, SNGFR for both the short and long nephrons exhibits a wide plateau (Fig. 3, blue curves), where it stays within 5% of the value obtained for a reference perfusion pressure of 100 mmHg [7, 17]. The afferent arterioles associated with the long nephrons are assumed to be slightly larger than those of the

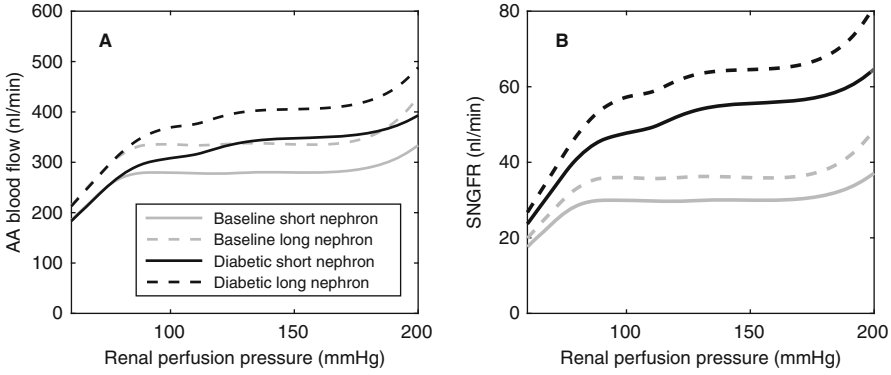


Fig. 3 Afferent arteriole blood flow (panel **a**) and SNGFR (panel **b**) associated with the short and long nephrons (*solid and dashed curves*, respectively) for perfusion pressures ranging in 60–200 mmHg. Results are shown for the baseline model (*blue curves*) and for the diabetes model (*red curves*)

short nephrons (baseline luminal diameter of 22 versus 20 μm at perfusion pressure of 100 mmHg, [3]); thus, the long nephrons are predicted to have larger arteriolar blood flow and SNGFR.

Autoregulation in the model diabetic kidney is predicted to be significantly impaired. Unlike the baseline model, the autoregulation curves no longer exhibit a plateau (compared blue and yellow curves in Fig. 3). Furthermore, the model predicts hyperfiltration in diabetes, with a SNGFR of 47.72 and 57.31 nl/min for the short and long nephrons, respectively, predicted for a perfusion pressure of 100 mmHg. These SNGFR are ~60% above the corresponding baseline values (30 and 36 nl/min for the short and long nephrons, respectively).

Additional comparison between the two models is provided in Fig. 4. Owing to the functional impairment in the VGCC in the afferent arteriole of a diabetic kidney, the vascular smooth muscle intracellular [Ca^{2+}] is predicted to be lower in the diabetic model (panel a1), resulting in larger arteriolar luminal diameters (panel a2) and higher arteriolar blood flows (panel a3). Together with the higher glomerular filtration coefficients, SNGFR are predicted to be substantially higher in the diabetic model (panel b1). However, with enhanced proximal reabsorption induced by tubular hyperfiltration, thick ascending limb flow is predicted to be lower in diabetes (panel b2), as is macula densa [NaCl] (panel b3), consistent with experimental results by Vallon and Thomson [33, 34]. Similar trends are exhibited by both the short and long nephrons.

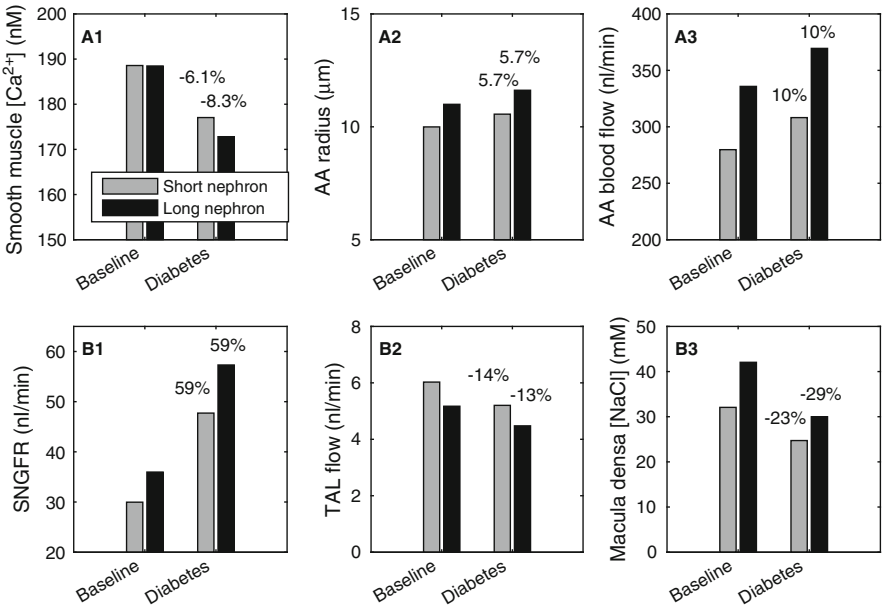


Fig. 4 Comparison of key model predictions for the baseline and diabetic models: (**a1**) afferent arteriole smooth muscle intracellular [Ca^{2+}]; (**a2**) afferent arteriole luminal radius; (**a3**) afferent arteriolar blood flow; (**b1**) SNGFR; (**b2**) thick ascending limb (TAL) tubular fluid flow; (**b3**) tubular fluid [NaCl] in thick limb outflow. *Blue bars*, short nephron; *yellow bars*, long nephron

3.2 Renal Oxygenation in Diabetes

In diabetes the enhanced proximal tubular reabsorption results in reduced NaCl delivery to the thick ascending limbs, which, taken in isolation, would result in reduced oxygen consumption. However, other competing factors may increase oxygen consumption. First, elevated oxidative stress (i.e., increased O_2^- production) enhances thick ascending limb active Na^+ transport and thus oxygen consumption. Furthermore, mitochondrial uncoupling decreases transport efficiency (represented in the model as the 20% reduction in TQ ratios) and increases oxygen consumption.

To determine the prevailing effects that alter medullary oxygenation in diabetes, we summarize key predictions of both models in Fig. 5. Medullary oxygen supply is provided by descending vasa recta inflow, which is assumed to be unchanged in early diabetes (panel a1). Enhanced proximal reabsorption reduces Na^+ delivery to the thick ascending limbs; as a result, active Na^+ reabsorption by the ascending limbs is slightly reduced (panel a2). Despite the reduced Na^+ reabsorption, medullary O_2 consumption increases due to the decreased transport efficiency (i.e., the reduced TQ ratio; see panel b1). The increased O_2 consumption, which is

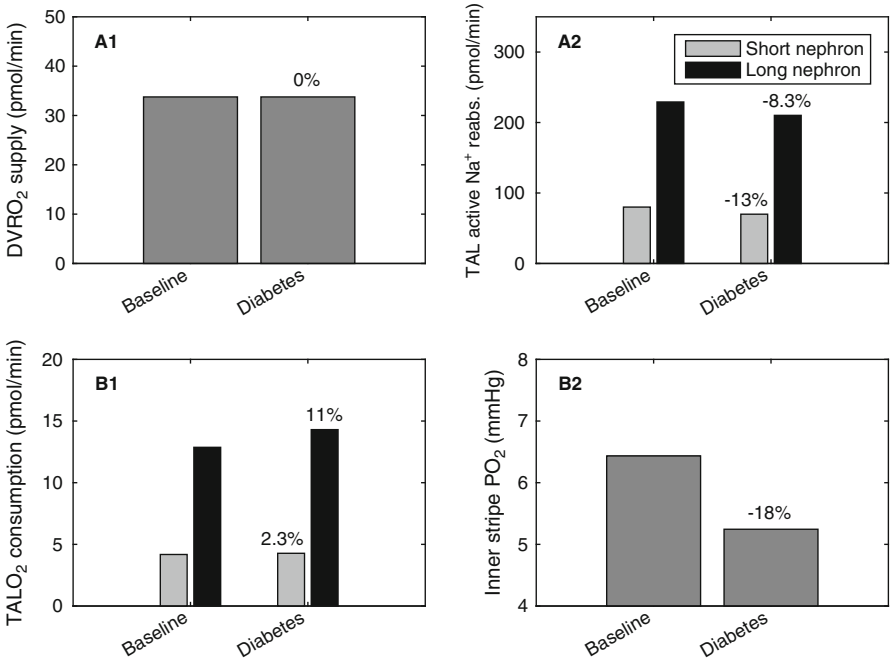


Fig. 5 Comparison of medullary oxygenation for the baseline and diabetic models: (**a1**) medullary oxygen supply via the descending vasa recta (DVR); (**a2**) active Na^+ reabsorption by the medullary thick ascending limb (TAL); (**b1**), medullary oxygen consumption; (**b2**) interstitial P_{O2} along the inner stripe. *Blue bars*, short nephron; *yellow bars*, long nephron

not compensated by an increase in O_2 supply, results in a significantly reduced medullary P_{O2}. In the inner stripe, which is a particularly O_2-poor region, P_{O2} decreases ~18.6%, from the baseline value of 6.45 to 5.25 mmHg in the diabetic kidney.

4 Discussion

Changes in glomerular hemodynamics critically contribute to the pathophysiology of diabetes. Carmines et al. [4] reported that depolarization-induced Ca^{2+} influx and the resulting increase in intracellular $[Ca^{2+}]$ are attenuated in the afferent arteriole of diabetic rats. Combined with other pathophysiological changes, the impaired afferent arteriole vasoresponse yields a substantially higher glomerular hyperfiltration. The resulting larger filtered Na^+ load increases the reabsorptive load of the nephron, and may have an impact on medullary oxygenation. However, in early diabetes, the proximal tubule hypertrophies and its reabsorptive capacity increases. The present model predicts that these competing factors result in a slightly lower Na^+ flow into the medullary thick ascending limbs. Taken in isolation, a lower Na^+ load should reduce thick limb transport work and oxygen consumption.

Besides changes in hemodynamics, metabolism is also altered in diabetes. Mitochondrial uncoupling via UCP-2 decreases transport efficiency. The model assumes that the TQ ratio is reduced by 20% in the diabetic kidney. Consequently, the model predicts that, even though active Na^+ transport is slightly reduced in diabetes, medullary O_2 consumption increases substantially, resulting in significantly lower medullary P_{O2}. This result suggests that alterations in metabolism may be an important pathway in the development of intrarenal tissue hypoxia and diabetic nephropathy.

Other pathophysiological mechanisms that are not represented in the present model may also contribute to the alterations in glomerular hemodynamics that are observed in diabetes mellitus. For instance, NOS-1 mediated NO production in macula densa is elevated in diabetes [18], which lowers vascular resistance. Also, findings by Leehey et al. [22] suggest that hyperglycemia activates the intrarenal renin–angiotensin system, leading to stimulation of locally produced angiotensin II, a potent vasoconstrictor. Elevated local angiotensin II may contribute to the development of glomerular hypertension by preferentially constricting efferent arterioles as compared to the afferent arterioles. The impacts of these processes may be investigated in a future more comprehensive model of a diabetic kidney.

Acknowledgements This work is the product of a workshop and short-term visits supported by the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. Support was also provided by the National Institutes of Health: National Institute of Diabetes and Digestive and Kidney Diseases and by the National Science Foundation, via grants #DK089066 and #DMS-1263995 to AT Layton.

References

1. Bagnasco, S., Good, D., Balaban, R., Burg, M.: Lactate production in isolated segments of the rat nephron. Am. J. Physiol. Ren. Physiol. **248**, F522–F526 (1985)
2. Bank, N., Mower, P., Aynedjian, H.S., Wilkes, B.M., Silverman, S.: Sorbinil prevents glomerular hyperperfusion in diabetic rats. Am. J. Physiol. **256**, F1000–F1006 (1989)
3. Carmines, P., Inscho, E., Gensure, R.: Arterial pressure effects on preglomerular microvasculature of juxtamedullary nephrons. Am. J. Physiol. **258**, F94–F102 (1990)
4. Carmines, P.K., Ohishi, K., Ikenaga, H.: Functional impairment of renal afferent arteriolar voltage-gated calcium channels in rats with diabetes mellitus. J. Clin. Invest. **98**, F386–F395 (1996)
5. Chen, J., Layton, A., Edwards, A.: A mathematical model of oxygen transport in the rat outer medulla: I. model formulation and baseline results. Am. J. Physiol. Ren. Physiol. **297**, F517–F536 (2009)
6. Chen, J., Sgouralis, I., Moore, L., Layton, H., Layton, A.: A mathematical model of the myogenic response to systolic pressure in the afferent arteriole. Am. J. Physiol. Ren. Physiol. **300**, F669–F681 (2011)
7. Cupples, W., Braam, B.: Assessment of renal autoregulation. Am. J. Physiol. Ren. Physiol. **292**, F1105–F1123 (2007)
8. Dickman, K., Mandel, L.: Differential effects of respiratory inhibitors on glycolysis in proximal tubules. Am. J. Physiol. **258**, F1608–F1615 (1990)
9. Edwards, A., Layton, A.: Modulation of outer medullary NaCl transport and oxygenation by nitric oxide and superoxide. Am. J. Physiol. Ren. Physiol. **301**, F979–F996 (2011)
10. Evans, R., Ince, C., Joles, J., Smith, D., May, C., O'Connor, P., Gardiner, B.: Haemodynamic influences on kidney oxygenation: the clinical implications of integrative physiology. Clin. Exp. Pharmacol. Physiol. **40**, 106–122 (2013)
11. Evans, R., Harrop, G., Ngo, J., Ow, C., O'Connor, P.: Basal renal oxygen consumption and the efficiency of oxygen utilization for sodium reabsorption. Am. J. Physiol. Ren. Physiol. **306**, F551–F560 (2014)
12. Foley, R.N., Collins, A.J.: End-stage renal disease in the United States: an update from the United States Rrenal Data System. J. Am. Soc. Nephrol. **18**, 2644–2648 (2007)
13. Fry, B., Edward, A., Sgouralis, I., Layton, A.: Impact of renal medullary three-dimensional architecture on oxygen transport. Am. J. Physiol. Ren. Physiol. **307**, F263–F272 (2014)
14. Fry, B., Edward, A., Layton, A.: Impacts of nitric oxide and superoxide on renal medullary oxygen transport and urine concentration. Am. J. Physiol. Ren. Physiol. **308**, F967–F980 (2015)
15. Greger, R., Schlatter, E., Lang, F.: Evidence for electroneutral sodium chloride cotransport in the cortical thick ascending limb of Henle's loop of rabbit kidney. Pflügers. Arch. **396**, 308–314 (1983)
16. Hansell, P., Welch, W., Blantz, R., Palm, F.: Determinants of kidney oxygenation and their relationship to tissue oxygen tension in diabetes and hypertension. Clin. Exp. Pharmacol. Physiol. **40**, 123–137 (2013)
17. Holstein-Rathlou, N., Marsh, D.: Renal blood flow regulation and arterial pressure fluctuations: a case study in nonlinear dynamics. Physiol. Rev. **74**, 637–681 (1994)
18. Komers, R., Lindsley, J., Oyama, T., Allison, K., Anderson, S.: Role of neuronal nitric oxide synthesis (NOS1) in the pathoegensis of renal hemodynamic changes in diabetes. Am. J. Physiol. **279**, F573–F583 (2000)
19. Layton, A.: A mathematical model of the urine concentrating mechanism in the rat renal medulla: I. formulation and base-case results. Am. J. Physiol. Ren. Physiol. **300**, F356–F371 (2011)
20. Layton, H., Pitman, E., Moore, L.: Bifurcation analysis of TGF-mediated oscillations in SNGFR. Am. J. Physiol. (Renal Fluid Electrolyte Physiol 30) **261**, F904–F919 (1991)

21. Layton, A., Dantzler, W., Pannabecker, T.: Urine concentrating mechanism: impact of vascular and tubular architecture and a proposed descending limb urea-Na^+ cotransporter. Am. J. Physiol. Ren. Physiol. **302**, F591–F605 (2012)
22. Leehey, D., Singh, A., Alavi, N., Singh, R.: Role of angiotensin II in diabetic nephropathy. Kidney Int. **58**(Suppl 77), S93–S98 (2000)
23. Moss, R., Layton, A.: Dominant factors that govern pressure natriuresis in diuresis and antidiuresis: a mathematical model. Am. J. Physiol. Ren. Physiol. **306**, F952–F969 (2014)
24. Nieves-Gonzalez, A., Clausen, C., Layton, A., Layton, H., Moore, L.: Transport efficiency and workload distribution in a mathematical model of the thick ascending limb. Am. J. Physiol. Ren. Physiol. **304**, F653–F664 (2013)
25. Sgouralis, I., Layton, A.: Autoregulation and conduction of vasomotor responses in a mathematical model of the rat afferent arteriole. Am. J. Physiol. Ren. Physiol. **303**, F229–F239 (2012)
26. Sgouralis, I., Layton, A.: Theoretical assessment of renal autoregulatory mechanisms. Am. J. Physiol. Ren. Physiol. **306**, F1357–F1371 (2014)
27. Stokes, J., Grupp, C., Kinne, R.: Purification of rat papillary collecting duct cells: functional and metabolic assessment. Am. J. Physiol. **253**, F251–F262 (1987)
28. Thomson, S.C., Vallon, V., Blantz, R.C.: Resetting protects efficiency of tubuloglomerular feedback. Kidney Int. Suppl. **67**, S65–S70 (1998)
29. Uchida, S., Endou, H.: Substrate specificity to maintain cellular ATP along the mouse nephron. Am. J. Physiol. Ren. Physiol. **255**, F977–F983 (1988)
30. Vallon, V.: The proximal tubule in the pathophysiology of the diabetic kidney. Am. J. Physiol. Regul. Integr. Comp. Physiol. **300**, R1009–R1022 (1996)
31. Vallon, V.: The proximal tubule in the pathophysiology of the diabetic kidney. Am. J. Physiol. Regul. Integr. Comp. Physiol. **300**, 1009–1022 (2011)
32. Vallon, V.: The mechanisms and therapeutic potential of SGLT2 inhibitors in diabetes mellitus. Annu. Rev. Med. **66**, 255–270 (2015)
33. Vallon, V., Thomson, S.: Renal function in diabetic disease models: the tubular system in the pathophysiology of the diabetic kidney. Annu. Rev. Physiol. **74**, 351–375 (2012)
34. Vallon, V., Blantz, R., Thomson, S.: Homeostatic efficiency of tubuloglomerular feedback is reduced in established diabetes mellitus in rats. Am. J. Physiol. (Renal Fluid Electrolyte Physiol 38) **269**, F876–F883 (1995)
35. Weinstein, A.: A mathematical model of the inner medullary collecting duct of the rat: pathways for Na and K transport. Am. J. Physiol. (Renal Physiol 43) **274**, F841–F855 (1998)
36. Zeidel, M., Silva, P., Seifter, J.: Intracellular pH regulation and proton transport by rabbit renal medullary collecting duct cells: role of plasma membrane proton adenosine triphosphatase. J. Clin. Invest. **77**, 113–120 (1986)

Tracking the Distribution of a Solute Bolus in the Rat Kidney

Anita T. Layton

Abstract The goal of this study is to develop a detailed mathematical model that tracks filtered solutes in the rat kidney. A better understanding of intrarenal solute distribution, and its cycling by way of countercurrent exchange and preferential tubular interactions, may yield new insights into fundamental principles of concentrating mechanism function. This is a complex problem, however, in part because of the marked heterogeneity exhibited in the transport properties of different nephron segments, and in the organization of tubules and vessels in the renal medulla, which likely gives rise to preferential interactions among neighboring tubules and vessels. The present model represents renal tubules in both the cortex and the medulla, the medullary vasculature, and their spatial relationship. By simulating the fate a marked bolus, we obtain the distribution of that solute as a function of time. In addition, we characterize the residence time of a solute by computing the portion of that solute remaining in the model kidney as a function of time. Model simulations of an anti-diuretic rat kidney predict that, owing to the different tubular transport properties to NaCl and urea, and to the more effective urea cycling mechanism in the inner medulla, the residence time of urea is substantially longer than that of NaCl. Simulation results also suggest that urea cycling is disrupted in the diuretic state, resulting in a significantly shorter residence time for urea.

1 Introduction

The kidney is an essential organ that serves a number of regulatory functions. The most well-understood role of the kidney is probably its function as a filter, removing metabolic wastes and toxins from blood and excreting them through the urine. Nonetheless, that is by no means the kidney's only crucial function. Through a number of regulatory mechanisms, the kidney also helps maintain the body's water balance, electrolyte balance, and acid-base balance [4].

A.T. Layton (✉)
Department of Mathematics, Duke University, Durham, NC 27708, USA
e-mail: alaytaon@math.duke.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_7

For example, to maintain water balance, the kidneys of a mammal produce a urine having an osmolality much higher than that of blood plasma during periods of water deprivation [5]. In other words, disproportionally more water is reabsorbed than solutes from the blood that has been filtered through the kidneys, thereby allowing the kidneys to serve its excretory function while preserving body water. This regulation of water balance is provided by the kidney's urine concentrating mechanism, which is localized in the innermost part of the kidney, known as the renal medulla. The renal medulla is divided into the outer medulla and the inner medulla. In the outer medulla, the concentrating mechanism is driven by active NaCl transport from the thick ascending limbs of loops of Henle. This concentrating effect, coupled with a countercurrent flow configuration of tubules and vessels, generates an axial osmolality gradient along all structures in the outer medulla, from the cortico-medullary boundary to the outer–inner medullary boundary. In contrast, the underlying mechanism in the inner medulla remains one of the long-standing mysteries in traditional physiology, despite much experimental and modeling effort.

Anatomic studies have revealed a highly structured organization of tubules and vasa recta in the outer medulla of some mammalian kidneys [1, 7]. Tubules and vessels are organized concentrically around vascular bundles, which are tightly packed clusters of parallel vessels and tubules containing mostly vasa recta. The countercurrent arrangement and proximity of the descending and ascending vasa recta is believed to facilitate diffusion exchange among these vessels. Recent studies of three-dimensional architecture of rat inner medulla have revealed transport and structural properties that may impact solute transport and have implications for the inner medullary urine concentrating mechanism in the mammalian kidney. In particular, these studies have shown that clusters of collecting ducts form the organizing motif through the initial 3–3.5 mm of the inner medulla [17–19]. Tubules and vessels that occupy nearby positions likely interact preferentially.

Investigation of intra-renal solute distribution, and its cycling by way of countercurrent exchange and preferential tubular interactions, allows model predictions to be compared with micropuncture data and with solute cycling paths inferred from anatomy and permeability data. Thus, such investigation may yield new insights into the fundamental principles of the urine concentrating mechanism and other kidney functions. Models of the solute and water transport in the kidney have usually been formulated as steady-state boundary-value problems involving differential equations expressing mass conservation (e.g., [9, 10, 15, 22]). Traditionally, solute and water cycling patterns have been predicted via an examination of steady-state model solutions [9, 10, 15, 22]. In this study, we propose an alternative and more powerful approach for examining intra-renal solute distribution and cycling: by tracking the dynamic distribution of a marked bolus of solute. (This is similar to tracer experiments in which a small amount of fluid containing a radioactive solute is injected into a renal tubule or a group of tubules [3].)

The method is implemented in a dynamic formulation of a concentrating mechanism model that represents renal tubules in both the cortex and the medulla of a rat kidney [16]. By computing the fate of a solute originally belonging to a marked bolus in the proximal tubule, we obtain the distribution of that solute as a

function of time throughout the cortex and medulla. In addition, we characterize the residence time of that solute in the model kidney by computing the portion of that solute remaining in the model as a function of time. In the context of the urine concentrating mechanism, the residence time provides a measure of the kidney's efficacy in trapping that solute. Residence time is also a useful measure in pharmaceutical applications, as it yields a measure of the effects of a drug or toxin on the kidney.

2 Mathematical Model

We implemented a time-dependent version of a rat kidney model [16] and extended the model to track the distribution of a solute bolus. The model represents the tubular and vascular architecture of the renal medulla, as well as the cortical thick ascending limbs and distal tubules; see Fig. 1.

Model Structure The model medulla includes loops of Henle, a composite collecting duct, and vasa recta. All nephron segments and vasa recta are represented as rigid tubules. The model loops of Henle extend from the cortico-medullary boundary to differing depths of the medulla (more below). The cortical nephron segments connect the medullary thick ascending limbs to the medullary collecting duct, which extends from the cortico-medullary boundary to the papillary tip; see Fig. 1.

Loops of Henle can be distinguished into short loops, which belong to superficial nephrons and which account for 2/3 of the loop population, and long loops, which are associated with the juxtamedullary nephrons. Short loops turn at the outer–inner medullary boundary, whereas long loops turn throughout the inner medulla. Thus, the medullary length of a long loop can be just over 2 mm (length of the outer medulla), 7 mm (length of entire medulla), or somewhere in between. The loop population decreases approximately exponentially in the inner medulla. To capture that population distribution, the model represents a distributed loop that turns continuously along the inner medulla. In that formulation, $C_{\mathrm{DL},k}(x, y, t)$ denotes the concentration of solute k at medullary level x and time t in a descending limb that turns at y. The same notation is used for other variables. A population density function $w_l(x)$ is assigned for the loops [9]. The fraction of loops remaining at medullary level x is given by

$$w_l(x) = \begin{cases} 1, & 0 \le x \le L_{\mathrm{OM}}, \\ \frac{1}{3} e^{-\left(a_1 + a_2\left(0.5\frac{(x-L_{\mathrm{OM}})}{L_{\mathrm{IM}}}\right)\right)0.5\frac{(x-L_{\mathrm{OM}})}{L_{\mathrm{IM}}}}, & L_{\mathrm{OM}} \le x \le L_{\mathrm{OM}} + 0.85L_{\mathrm{IM}}, \\ w_l(L_{\mathrm{OM}} + 0.85L_{\mathrm{IM}}) \times e^{-\left(a_3 + a_4\left(a_5\frac{(x-L_{\mathrm{OM}})}{L_{\mathrm{IM}}}\right)\right)a_5\frac{(x-L_{\mathrm{OM}})}{L_{\mathrm{IM}}}}, & L_{\mathrm{OM}}+0.85L_{\mathrm{IM}} < x \le L_{\mathrm{OM}}+L_{\mathrm{IM}}, \end{cases} \tag{1}$$

where L_{OM} and L_{IM} are the lengths of the model outer and inner medulla, taken to be 2 and 5 mm, respectively. $a_1 = 4.42$, $a_2 = 8.016$, $a_3 = 21.08$, $a_4 = 304.8$, and $a_5 = 0.7782$. The jump discontinuity of w_l at $x = L_{\mathrm{OM}}$ arises from the turning of the short loops, which account for 2/3 of the loop population.

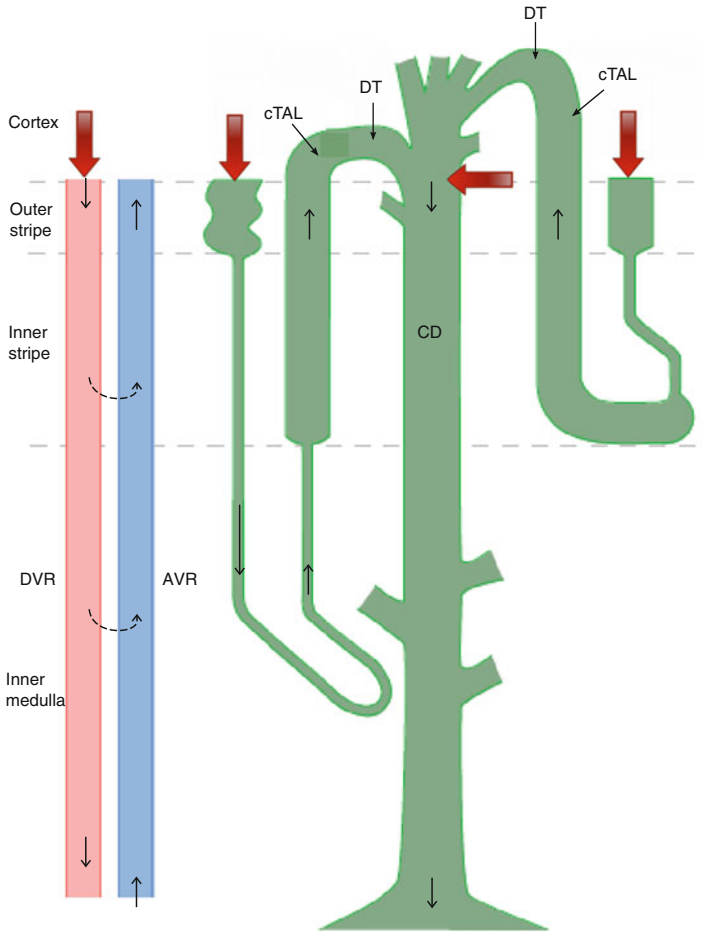


Fig. 1 Schematic diagram of model configuration. *DVR/AVR* descending/ascending vasa recta, *cTAL* cortical thick ascending limb, *DT* distal tubule, *CD* collecting duct. *Solid arrows* not associated with a label indicate flow directions. *Dashed arrows* indicate countercurrent exchange between vasa recta. *Red arrows* identify possible bolus injection sites. Figure modified from [16]

The medullary thick ascending limbs of the loops of Henle feed into the cortical thick ascending limbs. The model represents one superficial and one juxtamedullary cortical thick ascending limb, assumed to be contiguous with the medullary thick ascending limbs of the short and long loops, respectively. Each cortical think ascending limb is joined with a "distal tubule," which represents the distal convoluted tubule, connecting tubule, and cortical collecting duct. The distal tubules empty into the outer-medullary collecting duct.

In the inner medulla, the collecting ducts undergo successive coalescences, resulting in a decrease in the population of collecting ducts as a function of increasing inner-medullary depth. We assume that the fraction of collecting ducts remaining at a given medullary depth x is approximated by [13, 14]

$$w_{\mathrm{CD}}(x) = \begin{cases} 1, & 0 \le x \le L_{\mathrm{OM}}, \\ \left(1 - 0.97\left(\frac{x - L_{\mathrm{OM}}}{L_{\mathrm{IM}}}\right)^2\right) e^{-2.75\left(\frac{(x - L_{\mathrm{OM}})}{L_{\mathrm{IM}}}\right)}. & L_{\mathrm{OM}} \le x \le L. \end{cases} \tag{2}$$

Descending and ascending vasa recta are represented in the medulla. Descending vasa recta branch off into capillaries at every depth of the medulla, with a rate that can be computed from the population functions specified in [15]. The descending vasa recta branch off and empty their contents into "regions," which represent capillaries, and extra-tubular and extra-vascular structures like interstitial cells and interstitial spaces (more below). No axial flow is assumed along the regions; instead, any net fluid accumulation is drained by the model ascending vasa recta. One ascending vas rectum is represented for each region.

Four interstitial regions are included in the model medulla to represent the inhomogeneous distributions of renal tubules and vessels revealed in anatomic studies of mammalian kidneys [1, 7]. That structural organization may give rise to preferential interactions among tubules and vessels. In the outer medulla, tubules and vessels are organized concentrically around vascular bundles, which are tightly packed clusters of parallel vessels and tubules containing mostly of vasa recta. This organization is particularly pronounced in the inner stripe and continues into the initial inner medulla [17–19]. To represent this organization, we assume that at any given medullary level each region is a well-mixed compartment with which tubules and vasa recta interact. To specify the relative positions or distributions of the tubules and vasa recta and to simulate the potential preferential interactions among them, each tubule or vas rectum is assigned to a particular region [11, 12]. Tubules and vasa recta that are in contact with different regions are influenced by differing interstitial environments.

The model represents three solutes that are believed to play a key role in the mammalian urine concentrating mechanism: NaCl, urea, and a non-reabsorbable solute that represents all solutes that are neither reabsorbed nor secreted along the nephron. To track a NaCl or urea bolus, we further differentiate these solutes into two classes: background and bolus. The background and bolus solutes have the same transport properties.

Model Equations The steady-state version of key model equations can be found in [12, 14, 15]. The present model differs from these previous studies in that (1) we have adopted the dynamic formulation of those equations and (2) we have extended these models, which simulate solute transport in the renal medulla, to include the cortex. Model equations that describe the cortical segments are discussed below. We also summarize key equations and highlight any modifications from published models.

Water conservation in a tubule or vas rectum i (except the collecting ducts) is the same as in the steady-state model:

$$\frac{\partial}{\partial x} F_{i,v} = J_{i,v} \tag{3}$$

where $F_{i,v}$ denotes volume flow and $J_{i,v}$ is the transmural water flux. Solute conservation is given by

$$A_i \frac{\partial}{\partial t} C_{i,k} + \frac{\partial}{\partial x} \left(F_{i,v} C_{i,k} \right) = J_{i,k} \tag{4}$$

where $C_{i,k}$ is the concentration of solute k in the tubule or vas rectum, where k may denote background or original NaCl (denoted o-NaCl), bolus NaCl (denoted b-NaCl), background or original urea (denoted o-urea), bolus urea (denoted b-urea), or non-reabsorbable; A_i is the luminal cross-sectional area; and $J_{i,k}$ is the transepithelial solute flux.

Water conservation for the collecting ducts takes into account their coalescences, and is given by

$$\frac{\partial}{\partial x} \left(w_{\mathrm{CD}} F_{\mathrm{CD},v} \right) = w_{\mathrm{CD}} J_{\mathrm{CD},v} \tag{5}$$

where $F_{\mathrm{CD},v}$ and $J_{\mathrm{CD},v}$ denote volume flow and water flux, respectively, associated with a single collecting duct. Similarly, solute conservation is given by

$$A_{\mathrm{CD}} \frac{\partial}{\partial t} C_{\mathrm{CD},k} + \frac{\partial}{\partial x} \left(w_{\mathrm{CD}} F_{\mathrm{CD},v} C_{\mathrm{CD},k} \right) = w_{\mathrm{CD}} J_{\mathrm{CD},k} \tag{6}$$

The transmural solute flux $J_{i,k}$ into the loops of Henle and collecting ducts comprises of a passive diffusion and an active transport components, i.e.,

$$J_{i,k} = J_{i,k}^{\mathrm{pass}} + J_{i,k}^{\mathrm{act}} \tag{7}$$

Passive diffusion is characterized by transmural permeability and the interstitial region that the tubule interacts with; for details see [9, 12]. Active transport, represented only for NaCl, is characterized by Michaelis–Menten kinetics that takes into account the competition between original and bolus NaCl:

$$J_{i,\mathrm{NaCl}}^{\mathrm{act}} = \frac{V_{\max,\mathrm{NaCl}} C_{i,\mathrm{NaCl}}}{K_{M,\mathrm{NaCl}} + C_{i,\text{o-NaCl}} + C_{i,\text{b-NaCl}}} \tag{8}$$

Given that many studies of interest to us (e.g., urine concentrating mechanism) focus on solute distribution within the renal medulla, a relatively simple representation of the cortex is employed. The model cortical thick ascending limb is formulated similarly as its medullary counterpart. Because the cortex is well perfused, the cortical interstitial fluid composition is assumed to be the same as plasma. As previously noted, the distal convoluted tubule, connecting tubule, and cortical collecting duct are represented collectively as one "distal tubule." We assume that a fixed fraction ζ_v of the fluid entering the distal tubule is reabsorbed; similarly, a fixed fraction ζ_{NaCl} of the NaCl flow (both background and bolus NaCl) into the distal tubule is also reabsorbed. The values of ζ_v and ζ_{NaCl} depend on the

hydration state of the animal. Water and solute fluxes are assumed to be constant along each distal tubule but varying in time. Specifically, for the superficial distal tubule (denoted by the subscript DT-SF),

$$J_{\text{DT-SF},v}(t) = \zeta_v F_{\text{cTAL-SF},v}(x_{\text{cTAL-SF}}, t)/L_{\text{DT}} \tag{9}$$

where the subscript cTAL-SF denotes superficial cortical thick ascending limb, $x_{\text{cTAL-SF}}$ denotes the end of the superficial cortical thick ascending limb, and L_{DT} denotes the length of the distal tubule, taken to be 3 mm. Analogous expressions can be obtained for the juxtamedullary distal tubule and for NaCl fluxes.

We assume that the bolus urea is not reabsorbed along the distal tubule, and that the concentration of the background urea is adjusted to achieve a desired tubular fluid osmolality, which is assumed to increase linearly from the osmolality of cortical thick ascending limb outflow (allowed to be time varying) to plasma osmolality assumed at the medullary collecting duct inflow (constant and known a priori).

Boundary Conditions To complete the system, fluid flow and solute concentrations must be specified for the descending limbs and descending vasa recta at the cortico-medullary boundary, i.e., at $x = x_{\text{C-M}}$. At each loop bend, the associated descending and ascending limbs are assumed to be contiguous; thus, $F_{\text{DL},v}(y, y, t) = -F_{\text{AL},v}(y, y, t)$ (opposite flow directions) and $C_{\text{DL},k}(y, y, t) = C_{\text{AL},k}(y, y, t)$ for each k. Superficial medullary and cortical thick ascending limbs are assumed to be contiguous; thus, $F_{\text{cTAL-SF},v}(x_{\text{C-M}}, t) = F_{\text{AL},v}(x_{\text{C-M}}, x_{\text{OM-IM}}, t)$ and $C_{\text{cTAL-SF},k}(x_{\text{C-M}}, t) = C_{\text{AL},k}(x_{\text{C-M}}, x_{\text{OM-IM}}, t)$ for each k. The treatment for juxtamedullary thick ascending limbs are more complicated, inasmuch as long loops have differing lengths and thus differing composition and flow rates. Boundary flows for the juxtamedullary cortical ascending limb are given by the aggregates of the juxtamedullary medullary ascending limbs, i.e.,

$$F_{\text{cTAL-JM},v}(x_{\text{C-M}}, t) = 3 \int_{x_{\text{OM-IM}}}^{L} w_l'(y) F_{\text{AL},v}(x_{\text{C-M}}, y, t)\, dy \tag{10}$$

$$C_{\text{cTAL-JM},k}(x_{\text{C-M}}, t) = \frac{3 \int_{x_{\text{OM-IM}}}^{L} w_l'(y) F_{\text{AL},v}(x_{\text{C-M}}, y, t) C_{\text{C-M},k}(x_{\text{C-M}}, y, t)\, dy}{F_{\text{cTAL-JM},v}(x_{\text{C-M}}, t)} \tag{11}$$

The factor 3 arises from the need to yield a per-tubule value for $F_{\text{cTAL-JM}}$ given that $\int_{x_{\text{OM-IM}}}^{L} w_l'(y)\, dy = 1/3$ (only 1/3 of all loops are juxtamedullary).

Each cortical ascending limb is assumed to be contiguous with the corresponding distal tubule; i.e., $F_{\text{cTAL-SF},v}(x_{\text{cTAL-SF}}, t) = F_{\text{DT-SF},v}(x_{\text{cTAL-SF}}, t)$ and $C_{\text{cTAL-SF},k}(x_{\text{cTAL-JM}}, t) = C_{\text{DT-JM},v}(x_{\text{cTAL-JM}}, t)$ for each k; analogous relations are assumed for the juxtamedullary nephron segments.

To represent the merging of multiple connecting tubules into one cortical collecting duct, the model assumes that on average 6.1 distal tubules converge into one medullary collecting duct [6]. Two-thirds of these distal tubules are associated

with superficial nephrons and the rest are juxtamedullary. With these assumptions, collecting duct flows at the cortico-medullary boundary are given by:

$$F_{\mathrm{CD},v}(x_{\mathrm{C\text{-}M}},t) = -6.1 \times \left(\frac{2}{3}F_{\mathrm{DT\text{-}SF},v}(x_{\mathrm{C\text{-}M}},t) + \frac{1}{3}F_{\mathrm{DT\text{-}JM},v}(x_{\mathrm{C\text{-}M}},t)\right) \tag{12}$$

$$C_{\mathrm{CD},k}(x_{\mathrm{C\text{-}M}},t) = \frac{\left(\frac{2}{3}F_{\mathrm{DT\text{-}SF},v}(x_{\mathrm{C\text{-}M}},t)C_{\mathrm{DT\text{-}SF},k}(x_{\mathrm{C\text{-}M}},t) + \frac{1}{3}F_{\mathrm{DT\text{-}JM},v}(x_{\mathrm{C\text{-}M}},t)C_{\mathrm{DT\text{-}JM},k}(x_{\mathrm{C\text{-}M}},t)\right)}{F_{\mathrm{CD},v}(x_{\mathrm{C\text{-}M}},t)} \tag{13}$$

The minus sign in (12) is needed because the model distal tubules and collecting duct are assumed to have opposite flow directions.

Together with these boundary conditions and parameter values found in [15] and Table 1, model equations can be solved to predict fluid flow, solute concentrations (background and bolus) and transepithelial fluxes along each tubule and vessel at every depth of the medulla, and the interstitial solute concentrations (background and bolus) for each region, as functions of time.

3 Model Results

3.1 Steady-State Results

We first solved model equations to obtain steady-state solutions for two parameter sets, one corresponding to a rat kidney in an anti-diuretic state, and another in a mildly diuretic state. No bolus was represented in these steady-state simulations; only background solute concentrations were computed. Table 1 highlights key model parameters that differ between the two cases. Key results are displayed graphically in Fig. 2, which shows $[Na^+]$ and [urea] profiles along the nephrons.

An adequate understanding of the steady-state solution is essential for interpreting the dynamic bolus distribution results below. Thus, we first describe the steady-state solute concentration profiles and how they are generated. In the anti-diuretic model, a generally increasing osmolality gradient is obtained along the

Table 1 Key differences in transport parameters between the anti-diuretic and diuretic models

Parameter	Anti-diuretic	Diuretic	Unit
OMCD water permeability	463	23	μm/s
IMCD water permeability (initial)	750	125	μm/s
IMCD water permeability (terminal)	750	300	μm/s
IMCD urea permeability	110	46	$\times 10^{-5}$ cm/s
DCT fractional water reabs. ξ_v	83.5	63.3	%
DCT fractional urea reabs. ξ_{NaCl}	89.0	67.8	%

Water and urea permeabilities based on [20]

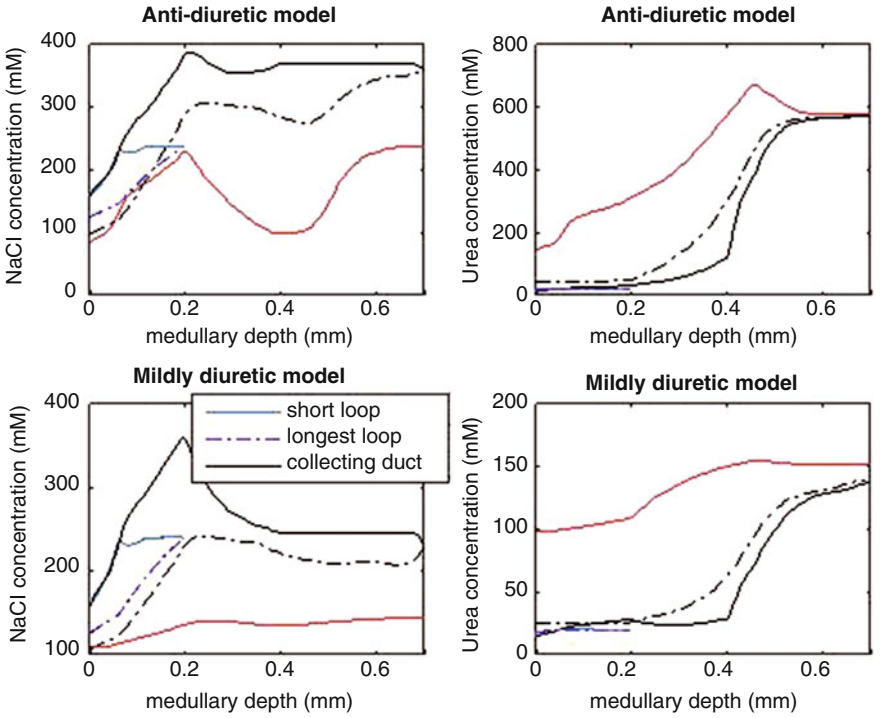


Fig. 2 Steady-state NaCl and urea tubular fluid concentration profiles, obtained using the anti-diuretic model (*top panels*) and the mildly diuretic model (*bottom panels*). *Solid* and *dashed lines* denote the descending and ascending limbs, respectively. *Dashed-dotted lines* denote ascending flows. Note the different *y*-axis scaling in the urea concentration plots

cortico-medullary axis in most structures. In the outer medulla, that axial gradient is generated by means of active Na^+ transport by the thick ascending limbs. The Na^+ reabsorbed progressively reduces the [Na^+] of the ascending limb luminal fluid, but elevates the interstitial fluid osmolality. As a result, water is continuously withdrawn from the tubular fluid of the water-permeable tubular segments and vessels, i.e., the collecting duct, the proximal straight tubules, segments of the descending limbs, and the descending vasa recta, thereby raising their fluid osmolality.

The concentrating mechanism in the rat inner medulla has yet to be fully elucidated. The model assumes that the axial osmolality gradient is generated in the inner medulla by means of the "solute-separation, solute-mixing" mechanism, the details of which are described in [9]. Here we summarize key steps of that mechanism. As a consequence of the concentrating mechanism in the outer medulla (see above), the collecting duct delivers a fluid into the inner medulla that is rich in urea. Thus, urea is reabsorbed from the collecting duct. Owing to the high collecting duct water permeability, water reabsorption follows, thereby diluting the interstitial fluid [Na^+]. Urea then enters the segments of the descending limbs that have high

urea permeability. Unlike the thick ascending limbs in the outer medulla, the thin ascending limbs in the inner medulla do not have active Na^+ transport; instead, those thin limbs are highly permeable to Na^+. Thus, Na^+ is reabsorbed from the thin ascending limbs, elevating the interstitial fluid [Na^+] and osmolality. That facilitates more water reabsorption from the collecting ducts, raising its tubular fluid urea concentration and osmolality.

A major difference between the diuretic and anti-diuretic models is that in diuresis water reabsorption is assumed to be much reduced along the cortical collecting ducts; thus, collecting duct flow is substantially higher at the cortico-medullary boundary in the diuretic model (14.7 versus 6.4 nl/min per collecting duct). The high collecting duct flow imposes a large load on the concentrating mechanism; i.e., a substantially larger amount of descending tubular fluid flow must now be concentrated by the concentrating mechanism. As a result, the diuretic model predicts a much more dilute urine with a osmolality of 435 mosmol/(kg H_2O), compared to 1192 mosmol/(kg H_2O) in the anti-diuretic model. (For further comparison, note that blood plasma has an osmolality of ∼300 mosmol/(kg H_2O).)

3.2 Bolus Simulations in an Anti-Diuretic Kidney

We then simulated the injection of a NaCl bolus and a urea bolus into the proximal straight tubules of an anti-diuretic kidney. Simulations were conducted separately for superficial and juxtamedullary nephrons. The injections were applied at the cortico-medullary boundary ($x = x_{\text{C-M}}$) for 5 s. The composition of the bolus is given by 16 mM of marked NaCl (bolus), 154 mM of unmarked NaCl (background), 1 mM of marked urea (bolus), 14 mM of unmarked urea (background), 0.1 mM of the non-reabsorbable solute. (In the absence of a bolus, descending limb boundary concentrations are set to be 160 mM NaCl, 15 mM urea, and 0.1 non-reabsorbable.)

Concentration Profiles Figure 3 shows the concentrations profiles of the NaCl bolus, injected into a superficial proximal straight tubule, at selected times. The bolus NaCl is rapidly pumped out of the thick ascending limb and carried by blood flow out of the system. Bolus NaCl concentration in the collecting duct tubular fluid is relatively low, with $< 5\%$ of the original NaCl bolus reaching the medullary collecting duct.

The fate of a urea bolus injected into a superficial proximal straight tubule differs significantly from that of NaCl. Because the short loop of Henle has low to moderate urea permeability and no active urea transport, a much bigger fraction (∼52%) of the urea bolus reaches the outer-medullary collecting duct, and ∼50% reaches the inner-medullary collecting duct. See Fig. 4 for urea bolus concentration profiles at selected times. Urea permeability remains low along the initial segment of the inner-medullary collecting duct (0.25×10^{-5} cm/s), but increases exponentially in the terminal inner medulla to 110×10^{-5} cm/s. Thus, much of the urea bolus is reabsorbed from the terminal inner-medullary collecting duct; it then enters the

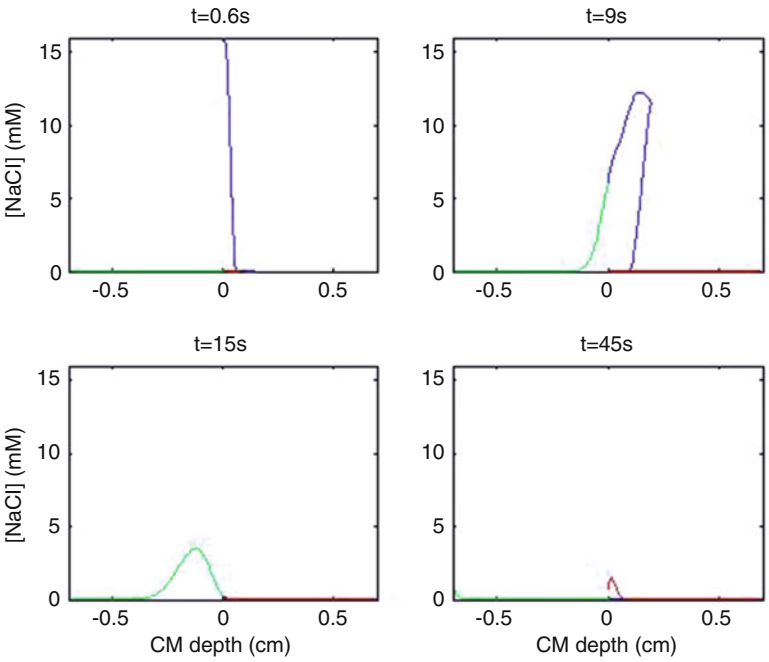


Fig. 3 Concentration profiles of a NaCl bolus injected into the superficial proximal straight tubule of the anti-diuretic model. *Blue*, loop of Henle; *green*, cortical nephron segments; *red*, collecting duct

descending and ascending limb segments of the long loops, which are highly urea permeable. As we will see below, this "urea cycling" greatly increases the residence time of urea within the medulla, especially in the inner medulla.

Figure 5 shows concentration profiles at selected times for a NaCl bolus injected into a juxtamedullary proximal tubule at the cortico-medullary boundary. The corresponding results for a urea bolus are shown in Fig. 6. Compared to the profiles shown in Figs. 3 and 4 for the superficial nephron, here the bolus concentration decreases rapidly along the long descending limb. This observation can be attributed to the tortuosity of the juxtamedullary proximal straight tubule; in contrast, the superficial proximal straight tubule is relatively straight. That tortuosity increases the transit time of the bolus as well as the tubular transport area. Thus, much of the NaCl is reabsorbed via active transport along the proximal straight tubule, with $< 4\%$ remaining in the loops of Henle after 30 s. Indeed, the NaCl bolus barely reaches the inner medulla.

Because the outer medullary segment of the long descending limb has low urea permeability (0.25×10^{-5} cm/s) and does not actively reabsorb urea, tubular fluid concentration of the urea bolus decreases much more slowly than the NaCl bolus. The largest decrease in urea bolus concentration can be observed along the terminal 60% of the inner medullary segment of the descending limbs. For the longest loop

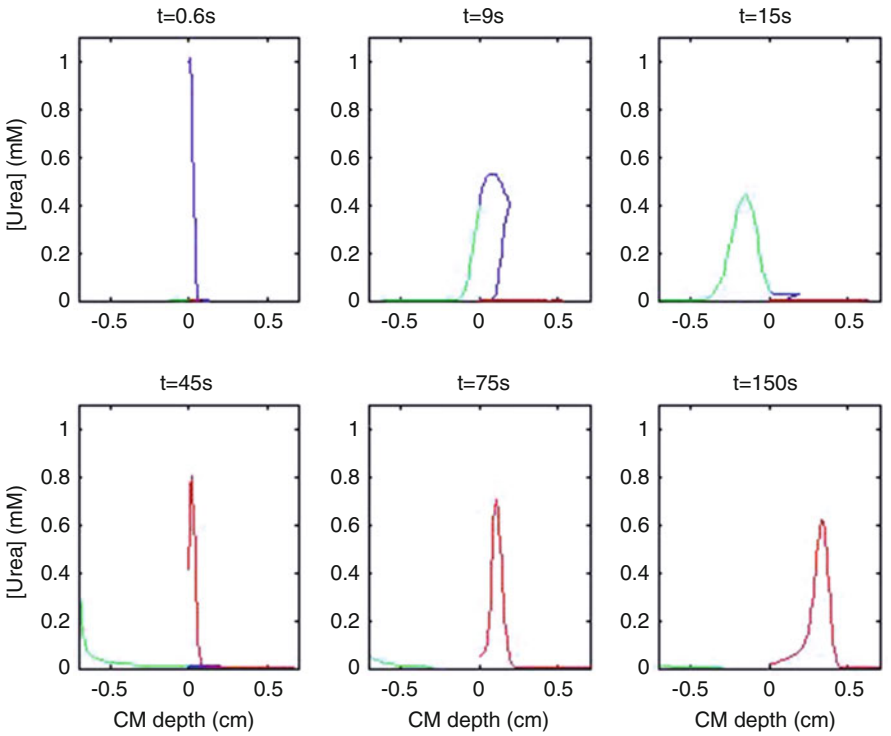


Fig. 4 Concentration profiles of a urea bolus injected into the superficial proximal straight tubule of the anti-diuretic model. Notations are analogous to Fig. 3

shown in Figs. 5 and 6 which reaches the papillary tip, that segment corresponds to the terminal 3 mm of the descending limb. That nephron segment exhibits a high urea permeability of 180×10^{-5} cm/s; as a result, the urea bolus is rapidly reabsorbed.

Residence Times To quantify the residence time of a solute within the kidney, we measure the fraction of a solute remaining in the medulla and the model cortical segments, at a given time t. The amount of bolus solute remaining at time t is given by the product of the luminal (or interstitial) cross-sectional area and solute concentration, taken over all structures, then integrated axially along the medulla. That is,

$$f_k(t) = \int_0^L \Bigg(M_{\mathrm{DL},k}(x,t) + M_{\mathrm{AL},k}(x,t) + n_{\mathrm{CD}} w_{\mathrm{CD}}(x) A_{\mathrm{CD}}(x) C_{\mathrm{CD},k}(x,t) + n_{\mathrm{DVR}} M_{\mathrm{DVR},k}(x,t) + \sum_{i=1}^{4} \left(n_{\mathrm{AVR_i}} A_{\mathrm{AVR_i}}(x) C_{\mathrm{AVR_i},k}(x,t) + n_{\mathrm{R}i} A_{\mathrm{R}i}(x) C_{\mathrm{R}i,k}(x,t) \right) \Bigg) \, dx$$

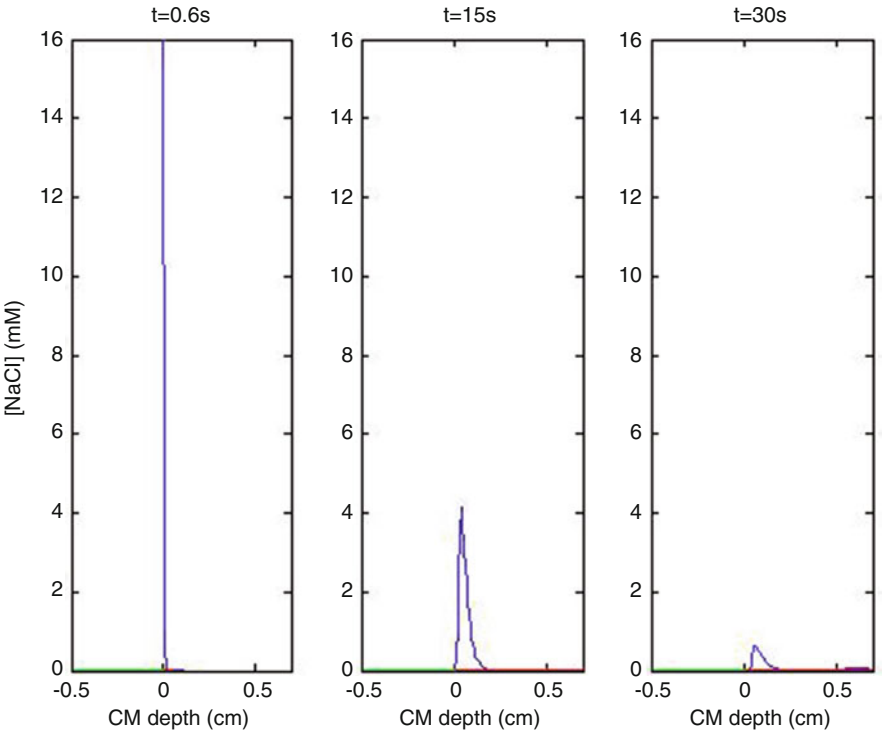


Fig. 5 Concentration profiles of a NaCl bolus injected into the juxtamedullary proximal straight tubules of the anti-diuretic model. Profiles are shown for a loop that reaches the papillary tip, although the bolus is administered to all loops that reach into the inner medulla. *Blue*, loop of Henle; *green*, cortical nephron segments; *red*, collecting duct

$$+\sum_{j=\mathrm{SF,JM}} n_j \left(\int_0^{L_{\mathrm{cTAL}\text{-}j}} A_{\,\mathrm{cTAL}\text{-}j}(x) C_{\mathrm{cTAL}\text{-}j,k}(x,t)\, dx + \int_0^{L_{\mathrm{DT}\text{-}j}} A_{\,\mathrm{DT}\text{-}j}(x) C_{\mathrm{DT}\text{-}j,k}(x,t)\, dx \right) \tag{14}$$

where n_i denotes the number of structure i per loop of Henle. A_i denotes luminal cross-sectional area, except for A_{Ri}, for Ri = R1, R2, R3, R4, which denotes interstitial area in the renal medulla not occupied by tubules or vessels. R1 and R2 correspond to the vascular bundles; R3 and R4 correspond to the interbundle regions. We assume that the amount of bolus in the cortical interstitium is negligible, due to the large blood supply in the cortex, which quickly washes out any bolus reabsorbed. As previously noted, the model represents four ascending vasa recta (denoted AVR), one for each region. The computation of the bolus found in the ascending vasa recta is relatively simple, inasmuch as the model ascending vasa recta neither terminate nor merge. On the other hand, the collecting duct population decreases within the inner medulla. Thus, the collecting duct area-concentration product is scaled by w_{CD}, the number of collecting ducts remaining at x. A fraction

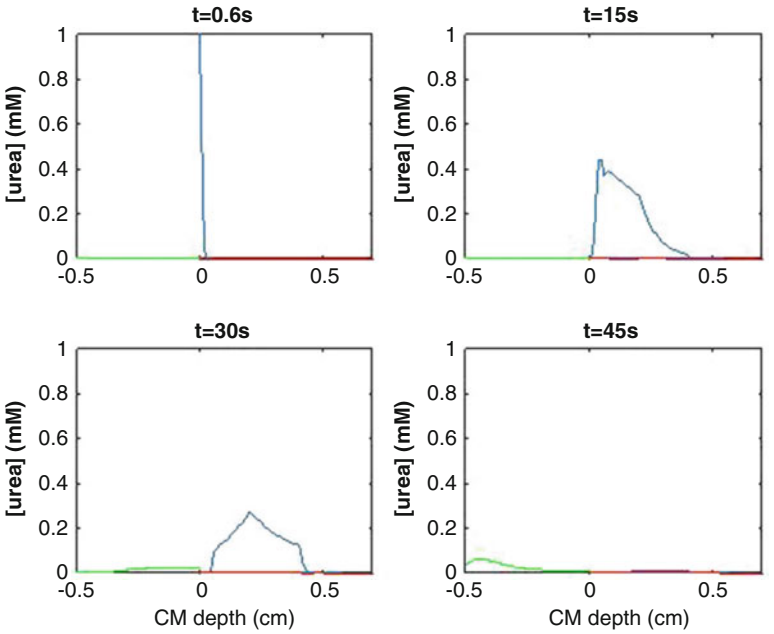


Fig. 6 Concentration profiles of a NaCl bolus injected into the juxtamedullary proximal straight tubule of the anti-diuretic model. Notations are analogous to Fig. 5

of the loops turn at level x, as do the descending vasa recta (DVR, which terminate into capillaries instead of turning like the loops do). Thus, we compute $M_{i,k}(x,t)$, which denotes the amount of solute at medullary level x, computed over all the instances of structure i that reach beyond x. For example, if x is in the outer medulla, then $M_{\mathrm{DL,NaCl}}(x,t)$ denotes the total amount of NaCl in all descending limbs, because none of the loops turn within the outer medulla. If x is the mid-inner medulla, then $M_{\mathrm{DL,NaCl}}(x,t)$ denotes the total amount of NaCl in the descending limbs of the long loop that turn within the lower half of the inner medulla. Specifically, $M_{\mathrm{DL},k}(x,t)$ is given by

$$M_{\mathrm{DL},k}(x,t) = \int_x^L w_l'(y) A_{\mathrm{DL}}(x,y) C_{\mathrm{DL},k}(x,y,t)\, dy \tag{15}$$

$M_{\mathrm{AL},k}$ and $M_{\mathrm{DVR},k}$ are defined analogously.

We first consider the residence time of a NaCl bolus injected into a superficial proximal straight tubule. Figure 7a shows the fraction of the bolus remaining within the model kidney as a function of time. The half-life of the NaCl bolus is 124 s, or $\sim$2 min. Initially, the majority of the bolus is found within the nephron (Fig. 7b), but as the bolus NaCl is reabsorbed along the loops of Henle, a progressively larger fraction of the remaining bolus is found in the interstitium (Fig. 7d). Much of the NaCl bolus is found in the interstitial region where most of the thick ascending limbs are found (R3).

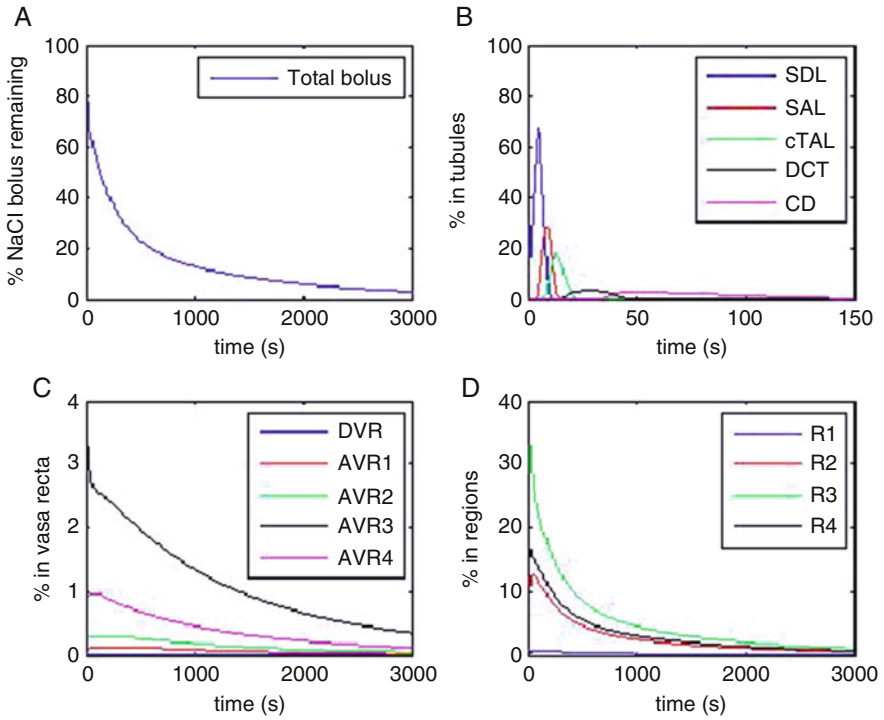


Fig. 7 Distribution of a NaCl bolus injected into the proximal straight tubule of a superficial nephron in the anti-diuretic model. (**a**) fraction of total bolus remaining over time. (**b**) fraction of bolus found in renal tubules as a function of time. SDL/SDL, short descending/ascending limb; cTAL, cortical thick ascending limb; DCT, distal connecting tubule; CD, collecting duct. (**c**) fraction of bolus found in vasa recta. DVR/AVR descending/ascending vasa recta. (**d**) fraction of bolus found in each of the four interstitial regions. Note the different *x*-axis scale in panel **b**

Intra-renal distribution of a urea bolus injected into a superficial proximal straight tubule is exhibited in Fig. 8. The half-life of a urea bolus is markedly longer than that of a NaCl bolus, 32.9 min compared to 2.1 min. That longer half-life can be attributed, in large part, to the differing tubular transport properties with respect to NaCl and urea. See discussion concerning the urea profile results in Fig. 4. Unlike NaCl, urea is not actively reabsorbed along the nephron, which explains why a larger fraction of the bolus reaches the distal tubule (47% of the urea bolus, compared to 5.2% of the NaCl bolus). Upon reaching the medullary collecting duct, only a minor fraction of the urea bolus is reabsorbed in the outer medulla or the initial inner medulla, where the collecting duct segments have low urea permeabilities (0.25×10^{-5} cm/s, see the plateau of the collecting duct curve in Fig. 8b). Collecting duct urea permeability is assumed to be high in the terminal inner medulla; thus, the urea bolus is rapidly reabsorbed instead of exiting the system in the urine. Urea reabsorbed from the renal tubules, especially the collecting ducts, is trapped

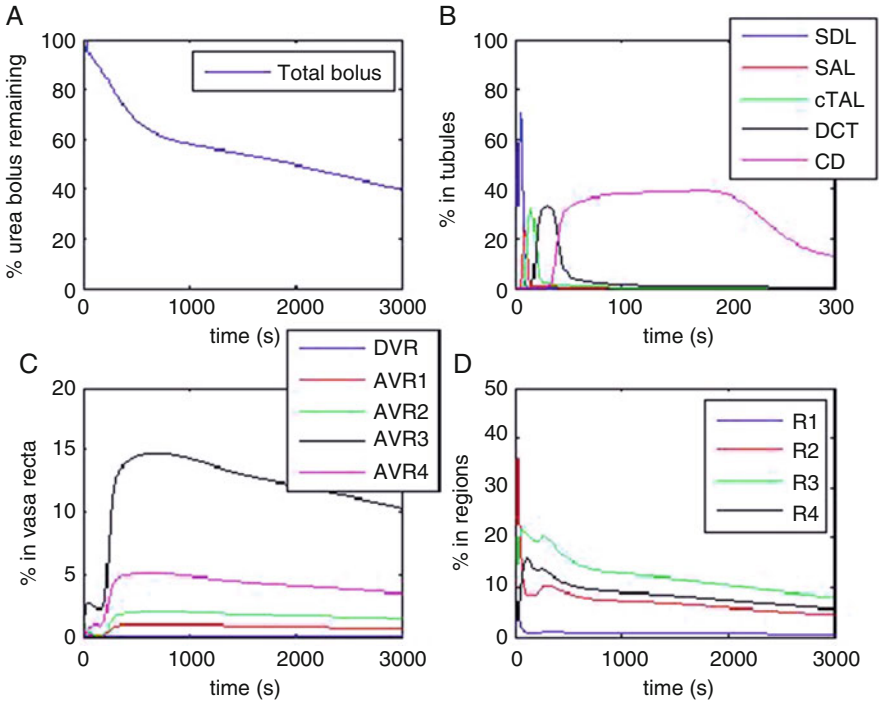


Fig. 8 Distribution of a urea bolus injected into the proximal straight tubule of a superficial nephron in the anti-diuretic model. Notations are analogous to Fig. 7

within the inner medulla by its urea cycling system, via the countercurrent vascular flows (the highly urea permeable descending and ascending vasa recta) or tubular flows (descending and ascending thin limbs, also highly urea permeable). The effectiveness of that urea cycling system is evidenced by the slow decrease in the fraction of urea bolus remaining in the interstitial regions (Fig. 8d).

We then consider a NaCl bolus injected into the proximal straight tubule of a juxtamedullary nephron. As previously noted, owing to its larger transport area, the majority of the NaCl bolus is pumped out along the juxtamedullary proximal straight tubule, and only a negligible fraction reaches the downstream nephron segments (see Fig. 9b). The half-life of the NaCl bolus is 140 s or $2\frac{1}{3}$ min. As in the case for the superficial nephron, a urea bolus has a substantially longer half-life (36 min, see Fig. 10), owing, in part, to the more effective cycling mechanism in the inner medulla.

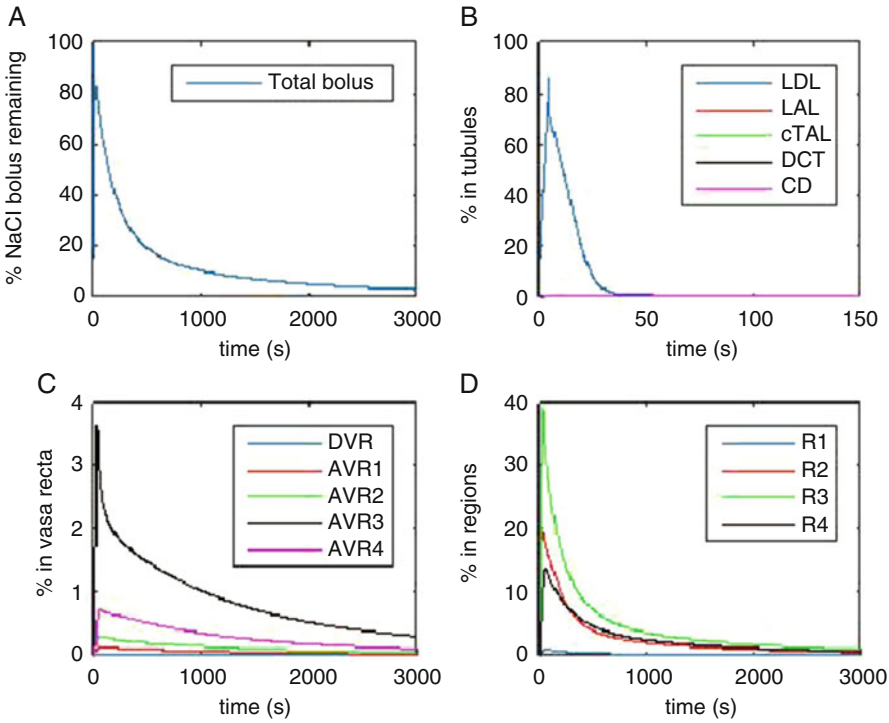


Fig. 9 Distribution of a NaCl bolus injected into the proximal straight tubule of a juxtamedullary nephron in the anti-diuretic model. Notations are analogous to Fig. 7. *LDL/LAL* long descending/ascending limb

3.3 Bolus Simulations in a Mildly Diuretic Kidney

We repeated the bolus simulations for a mildly diuretic kidney. The distribution of the NaCl bolus is similar to that in an anti-diuretic kidney (results not shown), because that distribution is determined primarily by the active NaCl transport in the outer medulla and cortex, which is assumed to be the same in the anti-diuretic and diuretic kidneys. In contrast, marked differences can be found in the distribution of a urea bolus. A comparison between the anti-diuretic and diuretic curves in Fig. 11 reveals a substantially shorter residence time of the urea bolus (injected into the superficial proximal straight tubule) in a diuretic kidney, especially in the collecting duct. That discrepancy arises because of the much faster collecting duct flow in the mildly diuretic kidney (14.8 versus 6.4 nl/min at the cortico-medullary boundary), owing to the assumption of low water permeability and thus much reduced water reabsorption along the distal and connecting tubules (see Table 1). As a result, the urea bolus reaches the terminal inner medullary collecting duct much faster in the diuretic kidney. Furthermore, unlike the anti-diuretic kidney, that terminal segment

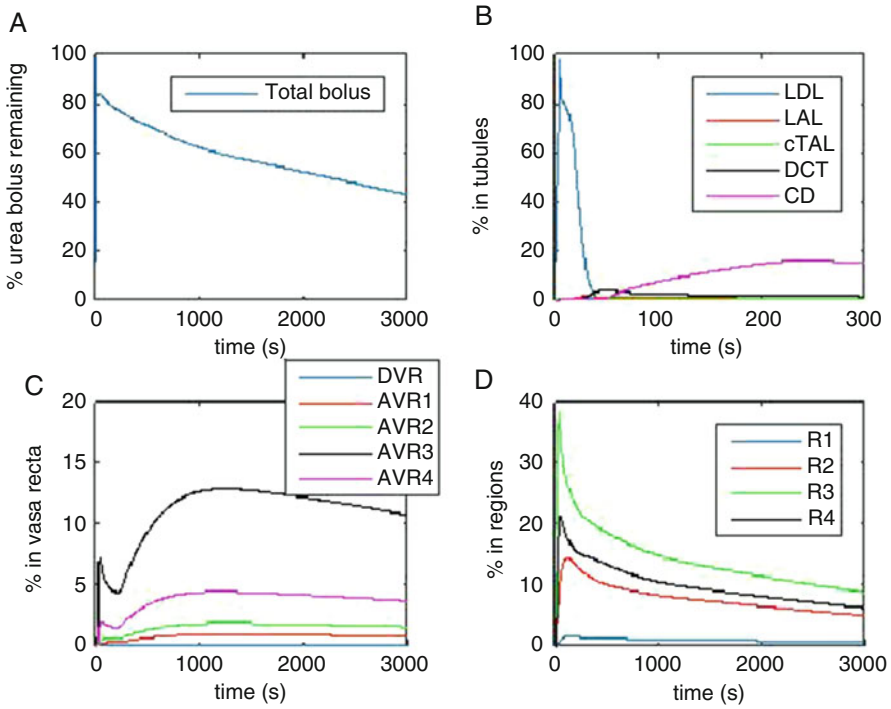


Fig. 10 Distribution of a urea bolus injected into the proximal straight tubule of a juxtamedullary nephron in the anti-diuretic model. Notations are analogous to Figs. 7 and 9

is assumed to have low urea permeability. Thus, instead of being reabsorbed, much of the urea bolus quickly exits the system, resulting in much shorter half-lives of 256 and 425 s, or 4.25 and 7.08 min, for injections into superficial and juxtamedullary proximal straight tubules, respectively, of the mildly diuretic kidney.

Danielson and Schmidt-Nielsen injected C^{14}-labelled urea into the distal tubules of rats, and measured its excretion and cumulative recovery rate [3]. We simulated the injection of a urea bolus into the collecting duct at the cortico-medullary boundary. The predicted excretion rate of the urea bolus is shown in Fig. 12. The model predicts that a significant amount of the marked urea appears in the urine about a minute after its injection, and that appears lasts about a minute. That prediction is consistent with measurements reported by Danielson and Schmidt-Nielsen [3].

4 Discussion

We have developed a model to predict the time-dependent distributions and residence times of filtered solutes in a mathematical model of the rat urine concentrating mechanism. Model simulations illustrate how the different tubular

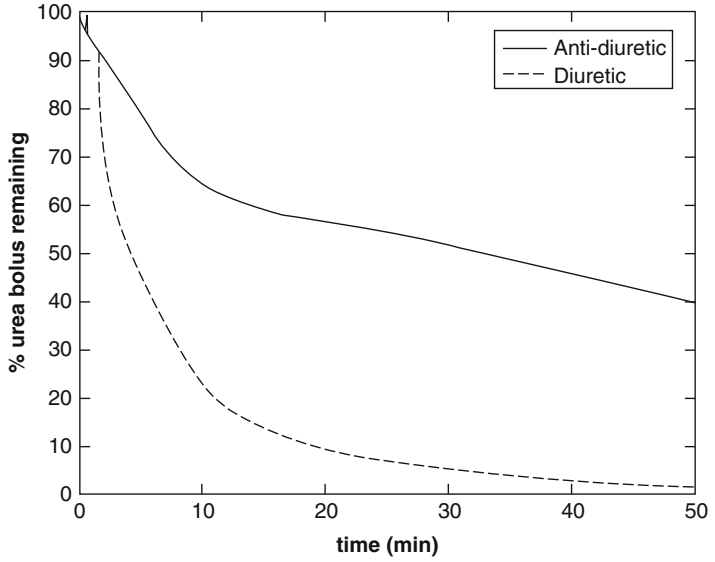


Fig. 11 Fraction of urea bolus remaining in the system. Urea residence time is substantially longer in the anti-diuretic model

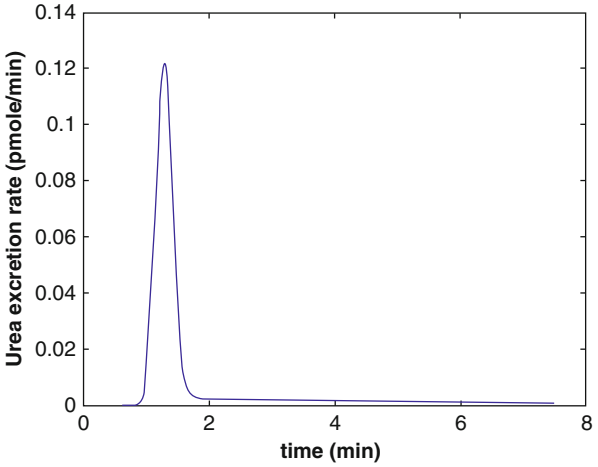


Fig. 12 Excretion rate of a urea bolus injected into the collecting ducts at the cortico-medullary boundary of a mildly diuretic kidney

transport properties of NaCl and urea impact the two types of solute boluses. Because NaCl is actively and briskly reabsorbed along some segments of the nephron, a NaCl bolus leaves the kidney at a significantly faster rate than a urea bolus. A large fraction of a bolus, NaCl or urea, does not traverse through the entire nephron (which, of course, would have been the case where tubular permeabilities

and active transport rate set to zero or some very low values). The model predicts excretion rate of a marked urea bolus consistent with experiment observations [3].

Traditionally, models of solute and water transport in the kidney are formulated as steady-state boundary-value problems involving differential equations expressing mass conservation (e.g., [9, 10, 15, 22]). About a decade ago, our group published the first urine concentrating mechanism model that focuses on dynamic solutions [8]. That model was based on the simple "central core formulation," and represented a single nephron. A so-called central core model typically represents only renal tubules explicitly, but not the vasa recta [21]. The renal tubules interact in a common compartment called the central core, which represents all extra-tubular structures, including the vasa recta, capillaries, interstitium, and interstitial spaces. Thus, unlike the present model, the central core model [8] did not capture the countercurrent vascular flows and their effects on solute cycling. Another important difference is that the present model contains a much more realistic representation of the anatomy of the rat kidney. In particular, loops of Henle with different lengths are represented. In contrast, the earlier model represented only a single nephron [8] and thus could not differentiate between boluses injected into the superficial versus juxtamedullary nephrons.

The cycling, or trapping, of solutes by way of countercurrent exchange and preferential tubular interactions is considered an important aspect of the mammalian urine concentrating mechanism [20]. Because the vast majority of the concentrating mechanism models are formulated for the steady state, solute cycling is usually assessed in a model using the predicted steady-state transmural fluxes or solute flows. The residence time of a bolus provides an alternative, and perhaps more intuitive, measure of the ability of a concentrating mechanism to cycle or trap a solute. Our results indicate that, compared to NaCl, urea is much more effectively cycled within the medulla of an anti-diuretic kidney and has a substantially longer residence time (see Figs. 7, 8, 9, and 10). That cycling mechanism is disrupted when the animal becomes diuretic and renal tubular transport properties change (Fig. 11).

In addition to providing insights into the urine concentrating mechanism, a better understanding of solute distribution can also help determine the effects of drugs. For instance, it is well known that the hydration status of a patient is a key determinant on the occurrence of contrast-induced acute kidney injury, which is the most common cause of iatrogenic, drug-induced acute kidney injury in hospitals [2]. However, the mechanism by which hydration prevents contrast-induced acute kidney injury is unclear. By simulating and comparing the distribution of contrast agent within the kidney in different hydration status may provide insights into that mechanism.

In this study we have conducted simulations to track a bolus of NaCl or urea. Extending the model to track other solute boluses is straightforward: additional solutes (both background and boluses) need to be added to the model, and transport properties of the nephrons and vessels to those solutes must be specified. Following this procedure, the model can be used to determine the distribution of a contrast agent and other drugs or toxins, and to assess their impacts on kidney function.

Acknowledgements This work is the product of a workshop and short-term visits supported by the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. Support was also provided by the National Institutes of Health: National Institute of Diabetes and Digestive and Kidney Diseases and by the National Science Foundation, via grants #DK089066 and #DMS-1263995 to AT Layton.

References

1. Bankir, L., de Rouffignac, C.: Urinary concentrating ability: insights from comparative anatomy. Am. J. Physiol. (Regulatory Integrative Comp. Physiol.18) **249**, R643–R666 (1985)
2. Calvin, A.D., Misra, S., Pflueger, A.: Contrast-induced acute kidney injury and diabetic nephropathy. Nat. Rev. Nephrol. **6**, 679–688 (2010)
3. Danielson, R.A., Schmidt-Nielsen, B.: Recirculation of urea analogs from renal collecting ducts of high- and low-protein-fed rats. Am. J. Physiol. **223**(1), 130–137 (1972)
4. Eaton, D.C., Pooler, J.P.: Vander's Renal Physiology, 6th edn. McGraw-Hill Medical, New York (2004)
5. Knepper, M.A., Hoffert, J.D., Packer, R.K., Fenton, R.A.: Urine concentration and dilution. In: Rector, B.M. (ed.) Brenner & Rector's the Kidney, 8th edn., pp. 308–330. Sanders, Philadelphia (2008)
6. Kriz, W.: Der architektonische and funktionelle Aufbau der Rattenniere. Z. Zellforsch. **82**, 495–535 (1967)
7. Kriz, W., Kaissling, B.: Structural organization of the mammalian kidney. In: The Kidney: Physiology and Pathophysiology, 3rd edn., pp. 587–654. Lippincott Williams & Wilkins, Philadelphia (2000)
8. Layton, A.T.: A methodology for tracking solute distribution in mathematical models of the kidney. J. Biol. Syst. **13**(4), 1–21 (2005)
9. Layton, A.T.: A mathematical model of the urine concentrating mechanism in the rat renal medulla: I. formulation and base-case results. Am. J. Physiol. Ren. Physiol. **300**, F356–F371 (2011)
10. Layton, A.T.: A mathematical model of the urine concentrating mechanism in the rat renal medulla: II. functional implications of three-dimensional architecture. Am. J. Physiol. Ren. Physiol. **300**, F372–F394 (2011)
11. Layton, A.T., Layton, H.E.: A region-based model framework for the rat urine concentrating mechanism. Bull. Math. Biol. **65**(5), 859–901 (2003)
12. Layton, A.T., Layton, H.E.: A region-based mathematical model of the urine concentrating mechanism in the rat outer medulla: I. formulation and base-case results. Am. J. Physiol. Ren. Physiol. **289**, F1346–F1366 (2005)
13. Layton, A.T., Pannabecker, T.L., Dantzler, W.H., Layton, H.E.: Two modes for concentrating urine in the rat inner medulla. Am. J. Physiol. Ren. Physiol. **287**, F816–F839 (2004)
14. Layton, A.T., Pannabecker, T.L., Dantzler, W.H., Layton, H.E.: Functional implications of the three-dimensional architecture of the rat renal inner medulla. Am. J. Physiol. Ren. Physiol. **298**, F973–F987 (2010)
15. Layton, A.T., Dantzler, W.H., Pannabecker, T.L.: Urine concentrating mechanism: impact of vascular and tubular architecture and a proposed descending limb urea-Na^+ cotransporter. Am. J. Physiol. Ren. Physiol. **302**, F591–F605 (2012)
16. Moss, R., Layton, A.T.: Dominant factors that govern pressure natriuresis in diuresis and antidiuresis: a mathematical model. Am. J. Physiol. Ren. Physiol. **306**, F952–F969 (2014)
17. Pannabecker, T.L., Dantzler, W.H.: Three-dimensional lateral and vertical relationship of inner medullary loops of Henle and collecting duct. Am. J. Physiol. Ren. Physiol. **287**, F767–F774 (2004)

18. Pannabecker, T.L., Dantzler, W.H.: Three-dimensional architecture of inner medullary vasa recta. Am. J. Physiol. Ren. Physiol. **290**, F1355–F1366 (2006)
19. Pannabecker, T.L., Abbott, D.E., Dantzler, W.H.: Three-dimensional functional reconstruction of inner medullary thin limbs of Henle's loop. Am. J. Physiol. Ren. Physiol. **286**, F38–F45 (2004)
20. Sands, J.M., Layton, H.E.: The urine concentrating mechanism and urea transporters. In: Alpern, R.J., Hebert, S.C. (eds.) Seldin and Giebisch's the Kidney: Physiology and Pathophysiology, 4th ed., pp. 1143–1178. Elsevier, New York (2008)
21. Stephenson, J.L.: Central core model of the renal counterflow system. Kidney Int. **2**, 85–94 (1972)
22. Wexler, A.S., Kalaba, R.E., Marsh, D.J.: Three-dimensional anatomy and renal concentrating mechanism. I. modeling results. Am. J. Physiol. (Renal Fluid Electrolyte Physiol. 29) **260**, F368–F383 (1991)

Mathematical Modeling of the Effects of Nutrient Competition and Bile Acid Metabolism by the Gut Microbiota on Colonization Resistance Against *Clostridium difficile*

Arietta Fleming-Davies, Sara Jabbari, Suzanne L. Robertson, Tri Sri Noor Asih, Cristina Lanzas, Suzanne Lenhart, and Casey M. Theriot

Abstract *Clostridium difficile* is the leading cause of infectious diarrhea in hospitals and one of the most common healthcare associated infections. Antibiotics alter the normal gut microbiota and facilitate the colonization of enteric pathogens such as *C. difficile*. Our objective is to elucidate the role of bile acids and other mechanisms in providing colonization resistance against *C. difficile*. We formulated and analyzed differential equation models for microbial interactions in the gut and bile acid dynamics, as well as a combined model including both mechanisms. Our analysis indicates that bile acids do not prevent *C. difficile* colonization, but

Fleming-Davies, Jabbari, Robertson and Noor Asih contributed equally.

A. Fleming-Davies
Department of Biology, Radford University, Radford, VA, USA
e-mail: arietta.flemingdavies@gmail.com

S. Jabbari
School of Mathematics, University of Birmingham, Birmingham, UK
e-mail: s.jabbari@bham.ac.uk

S.L. Robertson
Department of Mathematics and Applied Mathematics, Virginia Commonwealth University, Richmond, VA, USA
e-mail: srobertson7@vcu.edu

T.S.N. Asih
Department of Mathematics, Semarang State University, Semarang, Indonesia
e-mail: trisrinoorasih@gmail.com

C. Lanzas • C.M. Theriot
Department of Population Health and Pathobiology, College of Veterinary Medicine, North Carolina State University, Raleigh, NC, USA
e-mail: clanzas@ncsu.edu; cmtherio@ncsu.edu

S. Lenhart (✉)
Department of Mathematics, University of Tennessee, Knoxville, TN, USA
e-mail: lenhart@math.utk.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_8

they regulate the onset of *C. difficile* colonization and growth after antibiotic perturbation. These results have implications in the development of novel ways to inhibit *C. difficile* infection.

1 Introduction

Clostridium difficile is an anaerobic, spore-forming, gram-positive bacillus first isolated in 1935 [11]. *C. difficile* is the leading cause of infectious diarrhea in hospitals. Symptoms associated with *C. difficile* infection (CDI) include diarrhea, abdominal pain, and fever. In severe cases, CDI can cause colonic perforation, peritonitis, and death. CDI is a growing public health problem; in 2011, 453,000 primary infections, 83,000 first recurrences, and 29,300 deaths were estimated in the United States alone [20].

Antibiotic therapy is a strong and independent risk factor for CDI [28, 30]. Antibiotics alter the indigenous gut microbiota decreasing colonization resistance against *C. difficile* [5, 32]. Colonization resistance is the ability of the indigenous gut microbiota to prevent colonization of enteric pathogens [43]. The gut microbiota provides colonization resistance against enteric pathogens in different ways, namely by competing for nutrients or space, producing bacteriocins or inhibitors, and stimulating the immune response [3, 31, 45]. Work from the past decade has started to shed light on how antibiotics lead to a loss of colonization resistance against *C. difficile*. Antibiotics alter the gut microbiota and metabolome; specifically, they affect the composition and concentration of bile acids, carbohydrates, and amino acids [2, 40, 41]. The gut microbiome and metabolic environment after antibiotics favors *C. difficile* spore germination and outgrowth.

C. difficile spores are resistant to denaturation and are metabolically dormant, allowing for transmission of the pathogen. Spores require specific bile acids for maximal germination into a metabolically active vegetative cell, where it requires amino acids and carbohydrates to grow to high cell density and produce toxins, which mediate disease [14, 36, 46]. The bile acid pool in the body consists of primary bile acids that are made by the host liver, which are further biotransformed into secondary bile acids by members of the gut microbiota [33]. Antibiotic alterations in the gut microbiota result in a loss of the microbial derived secondary bile acid deoxycholate (DCA), and an increase in the primary bile acid taurocholate (TCA), which enhances *C. difficile* spore germination and growth [40, 41]. Secondary bile acid DCA, which is present in the gut prior to antibiotic treatment, can inhibit the growth of *C. difficile* [36, 41]. Gut microbiota mediated secondary bile acids in the gut may play a role in colonization resistance against *C. difficile*.

Another possible mechanism of colonization resistance is competition for nutrients by members of the gut microbiota. Members of the gut microbiota have different metabolic requirements, and are able to compete for a variety of nutrients. *C. difficile* requires amino acids (cysteine, isoleucine, leucine, proline, tryptophan, and valine) and vitamins (biotin, pantothenate, and pyridoxine) for growth [6, 13].

In addition to these nutrients, *C. difficile* is also able to ferment carbohydrates including fructose, glucose, mannitol, mannose, melezitose, sorbitol, and sialic acids [25, 34]. Bacteria that overlap in metabolic requirements from the same *Clostridium* genus compete for similar nutrients in vivo, suppressing growth of pathogenic *C. difficile* [19, 24]. After antibiotic treatment there may be a decrease in bacteria that are able to compete against *C. difficile*, allowing the pathogen to grow uninhibited.

Several mechanistic models of colonization resistance have been developed and analyzed. Freter et al. [9] developed the first mathematical model of colonization resistance against *Escherichia coli* in the gastrointestinal tract. The model describes the population dynamics of two bacterial strains, the resident and invader. The two strains compete for nutrients and adhesion sites. The model is composed of four ordinary differential equations (ODEs) that track the overall resident strain population and the invader strain in three adhesion sites and uses the Monod functional form to model microbial growth. The model predicts that both strains can co-exist if the metabolically less efficient strain (invader) has specific adhesion sites for which it does not compete [9, 22]. Subsequent work expanded Freter's work to relax some of the underlying assumptions, such as perfect mixing or competition for a single nutrient [4, 8]. For example, Coleman et al. [8] developed a model with 11 ODEs that represent five carbon sources, five indigenous microbiota groups, and one enteric pathogen, *Salmonella enterica*. Overall, the earlier models of colonization resistance considered different microbial groups depending on existent knowledge of the importance and interactions of each group and focused predominantly on nutrient competition.

In recent years, the availability of metagenomic high-throughput sequencing data has stimulated the development of data-driven models that reevaluate colonization resistance, particularly for *C. difficile*. Stein et al. [37] fitted a generalized Lotka–Volterra model to the abundance of the ten most abundant genera and *C. difficile* obtained from 16S rRNA high-throughput DNA sequencing data from a mouse model studying the effect of clindamycin on *C. difficile* colonization. A subset of four genera were identified as providing protection against *C. difficile*, but the underlying mechanisms were not modeled. Steinway et al. [38] used the same data to develop a Boolean dynamic model of the interactions among genera and used genome-scale metabolic reconstruction to gain insight into the mechanisms behind the interactions. No specific metabolic pathways were identified as an important source of the interactions between the gut microbiota and *C. difficile*. Despite using the same data, the studies differed on the bacterial genera that were deemed relevant to colonization resistance against *C. difficile*. Metagenomic 16S rRNA data is high-dimensional and sparse, and is reported as proportions, which may limit inference from the data [42]. Other mathematical models of *C. difficile* focus on disease transmission (e.g., [10, 17, 27]) or toxin production (e.g., [12]), but not colonization.

As the prevalence of antibiotic resistance rises, alternative treatment methods are being sought. Given the natural ability of the (undisturbed) gut microbiota to prevent *C. difficile* colonization, an obvious avenue for investigation is to consider how to recreate this natural defense mechanism in a compromised host. To achieve

this, the factors involved must be better understood. In this paper, we use current knowledge of the mechanisms underlying interactions between key members of the gut microbiota and *C. difficile* to develop and analyze mathematical models that focus on colonization resistance. We use data from Theriot et al. [40] to model the contribution of bile acid metabolism and competition by members of the gut microbiota in both the pre-antibiotic treated gut, which is resistant to *C. difficile*, and the post-antibiotic treated gut, which is susceptible to *C. difficile*, to evaluate the role of both mechanisms. Prior to antibiotic treatment the murine gut microbiota is dominated by two bacterial phyla, the Firmicutes (Lachnospiraceae family) and Bacteroidetes (Porphyromonadaceae family). After antibiotic treatment there is a shift in the murine gut microbiota to one that is dominated by members from the Firmicutes phylum (Lactobacillaceae family), where *C. difficile* is able to colonize. Similarly, prior to antibiotic treatment, the secondary bile acid DCA, which is an inhibitor of *C. difficile* growth, is present in the murine gut. After antibiotic treatment there is a loss of DCA and an increase in primary bile acid TCA in the gut, which *C. difficile* spores can utilize for germination and outgrowth. Members of the gut microbiota are important for the biotransformation of primary bile acids into secondary bile acids, TCA to cholate (CA) to DCA.

In Sect. 2 we formulate mathematical models for microbial interactions in the gut and bile acid dynamics, as well as a combined model including both mechanisms. Parameter estimation is discussed in Sect. 3, followed by an exploration of model dynamics in Sect. 4. We conclude with a discussion on the roles played by microbial interaction and bile acids in preventing *C. difficile* from colonizing the gut and how these could be manipulated for therapeutic benefit.

2 Model Development

To investigate the contributions of the two mechanisms described above to *C. difficile* colonization resistance, we analyze three ODE models: a model describing ecological interactions between groups of microbes in the gut, a model of the role of bile acid dynamics in *C. difficile* germination and outgrowth, and a combined model containing both mechanisms (see Fig. 1). We analyze the dynamics of each of the three models separately. In all three models, we compare a pre-antibiotic gut to a post-antibiotic gut by changing the initial conditions. Antibiotics are not directly modeled; instead, initial conditions (presence of different microbial taxa) differ in the pre-antibiotic and post-antibiotic gut, depending on the susceptibility of the different microbial groups to antibiotics. As different antibiotics affect different species, we limit our assumptions to the antibiotic cefoperazone, in line with the data from Theriot et al. [40].

For simplicity, we have condensed the complex microbial community into four functional groups, based on their effects on *C. difficile* and bile acid production. The four groups modeled are:

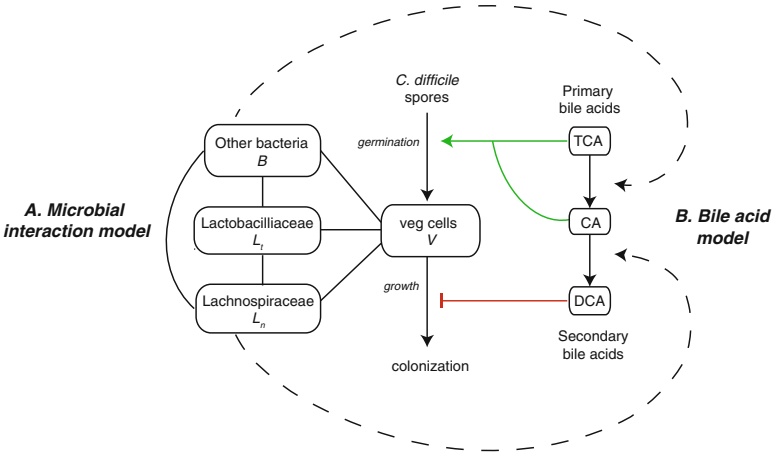


Fig. 1 Schematic of the models presented in this paper, representing the mechanisms of colonization resistance against *C. difficile* provided by the gut microbiota. *A*. Microbial interaction model between *C. difficile* vegetative cells, Lachnospiraceae, Lactobacillaceae, and other bacteria, which primarily compete with Lactobacillaceae. *B*. Bile acid model showing the bacterial conversion from primary bile acid taurocholate (TCA) to cholate (CA) to secondary bile acid deoxycholate (DCA). The *green lines* represent positive interactions with *C. difficile* germination and growth and the *red* represent negative interactions. The *dashed lines* represent the combined model. Note that some Lactobacillaceae members also have the ability to convert TCA to CA, which has been excluded from our model for simplicity

1. *C. difficile* vegetative cells, V, which are susceptible to the antibiotic cefoperazone;
2. Firmicutes phylum (primarily from the Lachnospiraceae family), L_n, which convert the primary bile acid CA to the secondary bile acid DCA, are susceptible to the antibiotic cefoperazone, and are thought to compete directly with *C. difficile*;
3. Firmicutes phylum (primarily from the Lactobacillaceae family), L_t, which do not affect bile acid metabolism, and are not susceptible to the antibiotic cefoperazone;
4. All other bacteria, B, which are dominated by the Bacteroidetes phylum (primarily from the Porphyromonadaceae family), are susceptible to the antibiotic cefoperazone, and are able to convert the bile acid TCA to CA.

See Table 1 for a summary of microbial taxa and traits. Note that there is evidence that DCA can also promote *C. difficile* germination, but at a much lower level than either TCA or CA [40, 47] and we accordingly omit this interaction term from the model.

Table 1 Summary of the four microbial groups considered in this study

Microbial group	Susceptible to cefoperazone?	Likely competitor for nutrients	Interactions with bile acids
C. difficile (V)	Yes	L_n	DCA inhibits growth; CA and TCA promote germination
Lachnospiraceae (L_n)	Yes	V	Converts CA to DCA
Lactobacillaceae (L_t)	No	B	–
Other bacteria (B)	Yes	L_t	Converts TCA to CA

2.1 Model of Microbial Interactions

We model interspecific interactions between each of the four microbial groups using a four-species Lotka–Volterra interaction model [29]:

$$\frac{dV}{dt} = r_V V(1 + \alpha_{VV}V + \alpha_{VL_n}L_n + \alpha_{VL_t}L_t + \alpha_{VB}B), \tag{1}$$

$$\frac{dL_n}{dt} = r_{L_n} L_n(1 + \alpha_{L_nV}V + \alpha_{L_nL_n}L_n + \alpha_{L_nL_t}L_t + \alpha_{L_nB}B), \tag{2}$$

$$\frac{dL_t}{dt} = r_{L_t} L_t(1 + \alpha_{L_tV}V + \alpha_{L_tL_n}L_n + \alpha_{L_tL_t}L_t + \alpha_{L_tB}B), \tag{3}$$

$$\frac{dB}{dt} = r_B B(1 + \alpha_{BV}V + \alpha_{BL_n}L_n + \alpha_{BL_t}L_t + \alpha_{BB}B). \tag{4}$$

Note that in our model, some interactions between groups are facilitative (positive) rather than competitive, and thus we refer to it as an "interaction" model rather than a competition model. Here V is the density of *C. difficile* vegetative cells, L_n is the density of Lachnospiraceae, L_t is the density of Lactobacillaceae, and B is the density of all other gut bacteria. Within-group and between-group interactions are described by the α terms, where α_{ij} gives the effect of group j on the growth rate of group i. Each α_{ij} may take a positive or negative value depending on whether group i has a net positive or negative effect on the other group (See Tables 2 and 6). Within-group competition is given by α_{ii}, the effect of group i on itself, and these values must always be negative. The intrinsic growth rate r_i is modified by the interaction of each group with all other groups, as $r_i(1 + \alpha_{iV}V + \alpha_{iL_n}L_n + \alpha_{iL_t}L_t + \alpha_{iB}B)$. See Table 3 for initial conditions in the pre-antibiotic and post-antibiotic gut and Table 6 for the default parameter set.

Table 2 Signs of interaction coefficients for the microbial interaction and full models

	V	L_n	L_t	B
V	−	−	+	−
L_n	+	−	+	−
L_t	+	−	−	−
B	+	+	+	−

The sign of each α_{ij}, indicating the effect of microbial group j on microbial group i, is found in cell i, j of the table. Negative values indicate a competitive effect, while positive values indicate a facilitative effect. Note that for some groups, α_{ij} is very close to zero (see Table 6 for exact values). These were estimated from limited data and may be subject to change

Table 3 Initial conditions for numerical simulations of the microbial interactions model in the pre-antibiotic and post-antibiotic gut

Scenario	Term	Definition	Value
Pre-antibiotics	$V(0)$	*C. difficile* vegetative cells	1×10^8
	$L_n(0)$	Lachnospiraceae	1×10^8
	$L_t(0)$	Lactobacillaceae	1×10^8
	$B(0)$	Other bacteria	1×10^8
Post-antibiotics	$V(0)$	*C. difficile* vegetative cells	100
	$L_n(0)$	Lachnospiraceae	0
	$L_t(0)$	Lactobacillaceae	100
	$B(0)$	Other bacteria	0

Note that in the pre-antibiotic gut we have unrealistically high levels of *C. difficile* and Lactobacillaceae to illustrate how a "healthy" gut could evolve purely from microbial interactions. In the post-antibiotic simulations, we assume all Lachnospiraceae and other bacteria are wiped out and a small number of *C. difficile* and Lactobacillaceae are introduced. The stability conditions of this model (discussed in Sect. 4.1) mean that the initial conditions in each case can be very broad to achieve the same steady states

2.2 Model of Bile Acid Interactions

Here we model a simplified system of the production and conversion of three key bile acids that interact with *C. difficile* vegetative cells, V, and *C. difficile* spores, S:

$$\frac{dV}{dt} = g(T + C)S + \frac{r_V}{1 + bD^2}(1 + \alpha_{VV}V)V, \tag{5}$$

$$\frac{dT}{dt} = h - \delta_T T - \frac{v_T BT}{M_T + T}, \tag{6}$$

$$\frac{dC}{dt} = \frac{v_T BT}{M_T + T} - \delta_C C - \frac{v_C L_n C}{M_C + C}, \tag{7}$$

$$\frac{dD}{dt} = \frac{v_C L_n C}{M_C + C} - \delta_D D, \tag{8}$$

$$\frac{dS}{dt} = -g(T + C)S. \tag{9}$$

T, C, and D represent the bile acids TCA, CA, and DCA, respectively. In the model of bile acid dynamics alone, vegetative *C. difficile* cells are the only one of the four microbial groups explicitly modeled (Eq. 5). Spores germinate at a rate $g(T + C)$, which is an increasing function of the bile acids TCA and CA. Growth of vegetative cells of *C. difficile* is modeled as in the microbial interactions model (Eq. 1), but including only intraspecific competition. In this model, however, the growth rate of *C. difficile* vegetative cells is a decreasing function of the concentration of the bile acid DCA.

The conversion between different bile acids (TCA to CA and CA to DCA) is modeled using the Michaelis–Menten kinetics function, where the reaction rate depends on the concentrations of the substrate and product. The parameters M_T and M_C are the half-saturation values, the substrate concentrations at which the corresponding reaction rates are half of the maximum corresponding reaction rates given by v_T and v_C, respectively. Lachnospiraceae, L_n, convert the bile acid CA to the bile acid DCA (Eq. 7), and other bacteria, B, convert the bile acid TCA to the bile acid CA (Eq. 6). In the bile acid model, L_n and B population sizes are held constant and not explicitly modeled. All bile acids are subject to natural degradation and a source term for TCA is included. See Table 4 for initial conditions in the pre-antibiotic and post-antibiotic gut.

2.3 Combined Model

In the combined model, growth of vegetative *C. difficile* cells, V, is modified by interactions with other microbial taxa as well as by the concentration of the bile acid DCA present in the gut:

$$\frac{dV}{dt} = \frac{r_V V}{1 + bD^2}\left(1 + \alpha_{VV}V + \alpha_{VL_n}L_n + \alpha_{VL_t}L_t + \alpha_{VB}B\right) + g(T + C)S, \tag{10}$$

$$\frac{dL_n}{dt} = r_{L_n}L_n(1 + \alpha_{L_nV}V + \alpha_{L_nL_n}L_n + \alpha_{L_nL_t}L_t + \alpha_{L_nB}B), \tag{11}$$

$$\frac{dL_t}{dt} = r_{L_t}L_t(1 + \alpha_{L_tV}V + \alpha_{L_tL_n}L_n + \alpha_{L_tL_t}L_t + \alpha_{L_tB}B), \tag{12}$$

$$\frac{dB}{dt} = r_B B(1 + \alpha_{BV}V + \alpha_{BL_n}L_n + \alpha_{BL_t}L_t + \alpha_{BB}B), \tag{13}$$

Table 4 Initial conditions for numerical simulations of the bile acid model in the pre-antibiotic and post-antibiotic gut

Scenario	Term	Definition	Value
Pre-antibiotics	$V(0)$	*C. difficile* vegetative cells	0
	L_n	Lachnospiraceae (assumed constant)	9.867×10^7
	B	Other bacteria (assumed constant)	1.3305×10^8
	$T(0)$	Concentration of TCA	4×10^{-4}
	$C(0)$	Concentration of CA	5×10^{-3}
	$D(0)$	Concentration of DCA	0.1
	$S(0)$	*C. difficile* spores	100
Post-antibiotics	$V(0)$	*C. difficile* vegetative cells	0
	L_n	Lachnospiraceae (assumed constant)	0
	B	Other bacteria (assumed constant)	0
	$T(0)$	Concentration of TCA	4×10^{-4}
	$C(0)$	Concentration of CA	5×10^{-3}
	$D(0)$	Concentration of DCA	0.1
	$S(0)$	*C. difficile* spores	100

In the pre-antibiotic gut the values of L_n and B are taken to be their steady states in the pre-antibiotic simulation of the microbial interaction model. The initial condition of DCA is chosen to be relevant to *C. difficile*-growth-inhibiting concentrations. The initial number of spores is chosen to match experimental work, e.g., [16]

$$\frac{dT}{dt} = h - \delta_T T - \frac{v_T B T}{M_T + T}, \tag{14}$$

$$\frac{dC}{dt} = \frac{v_T B T}{M_T + T} - \delta_C C - \frac{v_C L_n C}{M_C + C}, \tag{15}$$

$$\frac{dD}{dt} = \frac{v_C L_n C}{M_C + C} - \delta_D D, \tag{16}$$

$$\frac{dS}{dt} = -g(T + C)S. \tag{17}$$

As in the bile acid model, Lachnospiraceae bacteria, L_n, convert the bile acid CA to the bile acid DCA, and other bacteria, B, convert the bile acid TCA to the bile acid CA. However, here the populations of L_n and B are explicitly modeled (Eqs. 11 and 13, respectively), with growth rates affected by each group's interactions with other microbes as in the microbial interaction model. See Table 5 for initial conditions in the pre-antibiotic and post-antibiotic gut and Table 6 for the default parameter set.

Table 5 Initial conditions for numerical simulations of the combined model in the pre-antibiotic and post-antibiotic gut

Scenario	Term	Definition	Value
Pre-antibiotics	$V(0)$	*C. difficile* vegetative cells	0
	$L_n(0)$	Lachnospiraceae	1×10^8
	$L_t(0)$	Lactobacillaceae	1×10^8
	$B(0)$	Other bacteria	1×10^8
	$T(0)$	Concentration of TCA	4×10^{-4}
	$C(0)$	Concentration of CA	5×10^{-3}
	$D(0)$	Concentration of DCA	0.1
	$S(0)$	*C. difficile* spores	100
Post-antibiotics	$V(0)$	*C. difficile* vegetative cells	0
	$L_n(0)$	Lachnospiraceae	0
	$L_t(0)$	Lactobacillaceae	1
	$B(0)$	Other bacteria	0
	$T(0)$	Concentration of TCA	4×10^{-4}
	$C(0)$	Concentration of CA	5×10^{-3}
	$D(0)$	Concentration of DCA	0.1
	$S(0)$	*C. difficile* spores	100

In both cases we start with a small number of *C. difficile* spores. In the pre-antibiotic case we start with equal L_t, L_n, and B to see how the latter two outcompete L_t and V. In the post-antibiotic case we assume both L_n and B are wiped out and L_t is introduced to the system

3 Parameter Estimation

We currently have insufficient data to reliably estimate the relatively high number of model parameters. Furthermore, since we are collating many different microbial taxa within each group, it would be unwise to speculate on the exact value of the parameters. Instead we have used available data as follows to gauge initial estimates where possible and focus on qualitative, rather than quantitative conclusions from the model simulations. In addition, we performed sensitivity analysis for the uncertain parameters.

Growth curves for *C. difficile* in brain heart infusion (BHI) media supplemented with different concentrations of DCA [41], Lachnospiraceae in minimal media (unpublished data) and Lactobacillaceae in MRS media [26] were used to estimate the growth rates r_V, r_{L_n}, and r_{L_t} by fitting a logistic function to the data up to the points where the growth curves exhibited logistic growth. Optical density (OD) measurements were scaled by 2×10^8 to account for the conversion from OD to cell number. Data of *C. difficile* in BHI media were used to estimate growth rates under different concentrations of DCA (Fig. 2a, for example); we then fit a nonlinear function of the form $a/(1 + bD^2)$ (Fig. 2b) to create a function that represents inhibition of *C. difficile* growth by DCA. However, since Lachnospiraceae are believed to be the main competitors for nutrients with *C. difficile* and our growth rate

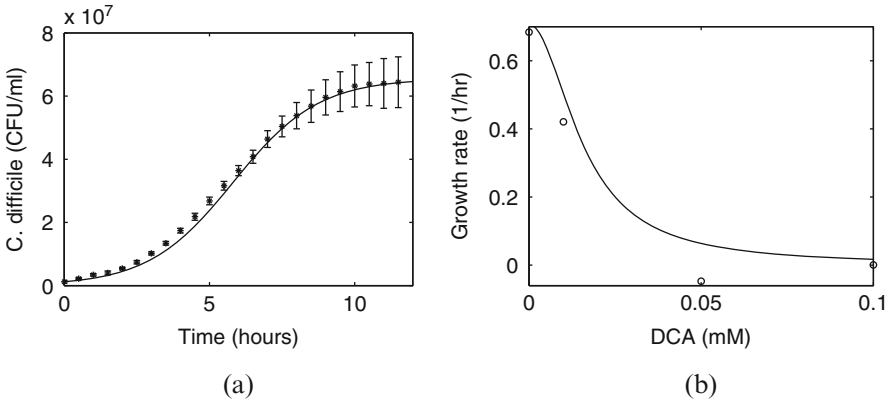


Fig. 2 (**a**) Logistic growth fit (*solid line*) to in vitro *C. difficile* growth data in BHI media with no DCA (*asterisks*) [41] (estimated growth rate for this example 0.6839 h^{-1}, and carrying capacity 6.554×10^7). Error bars signify standard deviation from the mean (from three repeats). (**b**) Estimated in vitro growth rates of *C. difficile* at different concentrations of the bile acid DCA (*circles*) were used to fit a function of the form $a/(1 + bD^2)$ describing the inhibition of *C. difficile* vegetative growth by DCA (*solid line*). Data from Theriot et al. [41]

for the Lachnospiraceae was estimated from data extracted from minimal media, this function was scaled so that in the absence of DCA it would equal the growth rate of *C. difficile* estimated from another data set for *C. difficile* measured in minimal media. Since the other bacterial group represents a collection of bacteria, we simply used the same growth rate as that for the Lactobacillaceae as they use similar resources (Table 1).

All microbial interaction coefficients α_{ij} were summarized from data from Stein et al. [37]. In order to simplify microbial interactions into just four groups, we summed the terms α_{ij} from Stein et al. across all taxa within each of our four groups. Interaction coefficients α_{ij} ranged from negative (competition) to positive (facilitation). We chose not to specify signs of the interaction coefficients in the model equations, as these depend on parameter values and are not based on known mechanisms of ecological interactions between groups. The signs of the interaction coefficients (and thus the direction of the effects) are summarized in Table 2.

In order for each group of bacteria to be able to achieve steady states of the correct order of magnitude (based on the growth curves mentioned above) when simulated in isolation from the other species, the α_{ij} were each scaled by 2×10^{-8} to maintain relative sizes and then rounded to one decimal place (notice that carrying capacities are not included explicitly in the model). For the model to reproduce what is seen in experimental work [40] (B and L_n dominating the pre-antibiotic gut and L_t dominating initially in a post-antibiotic gut), three of the α_{ij} were subsequently modified ($\alpha_{LnB}, \alpha_{LtV}$ and α_{LtB}), but kept within the range of the other α_{ij} and their positive or negative effects maintained.

Enzyme kinetics parameters for CA and DCA production were chosen to produce reasonable concentrations of TCA, CA, and DCA at steady state [33].

The default parameter set is used throughout the paper unless otherwise stated and can be found in Table 6.

Table 6 Parameter values for the combined model, including both the microbial interaction model (top section) and the bile acid model (lower section)

Parameter	Definition	Units	Value
r_V	Growth rate of *C. difficile* vegetative cells	h^{-1}	0.332
r_{L_n}	Growth rate of Lachnos	h^{-1}	0.711
r_{L_t}	Growth rate of Lactos	h^{-1}	0.665
r_B	Growth rate of other bacteria	h^{-1}	0.665
α_{VV}	Effect of V on V	$cells^{-1}\ h^{-1}$	-3×10^{-9}
α_{VL_n}	Effect of L_n on V	$cells^{-1}\ h^{-1}$	-6×10^{-9}
α_{VL_t}	Effect of L_t on V	$cells^{-1}\ h^{-1}$	1×10^{-9}
α_{VB}	Effect of B on V	$cells^{-1}\ h^{-1}$	-6×10^{-9}
α_{L_nV}	Effect of V on L_n	$cells^{-1}\ h^{-1}$	8×10^{-9}
$\alpha_{L_nL_n}$	Effect of L_n on L_n	$cells^{-1}\ h^{-1}$	-1×10^{-8}
$\alpha_{L_nL_t}$	Effect of L_t on L_n	$cells^{-1}\ h^{-1}$	8×10^{-10}
α_{L_nB}	Effect of B on L_n	$cells^{-1}\ h^{-1}$	-1×10^{-10}
α_{L_tV}	Effect of V on L_t	$cells^{-1}\ h^{-1}$	2×10^{-10}
$\alpha_{L_tL_n}$	Effect of L_n on L_t	$cells^{-1}\ h^{-1}$	-8×10^{-10}
$\alpha_{L_tL_t}$	Effect of L_t on L_t	$cells^{-1}\ h^{-1}$	-5×10^{-9}
α_{L_tB}	Effect of B on L_t	$cells^{-1}\ h^{-1}$	-1×10^{-8}
α_{BV}	Effect of V on B	$cells^{-1}\ h^{-1}$	8×10^{-9}
α_{BL_n}	Effect of L_n on B	$cells^{-1}\ h^{-1}$	2×10^{-9}
α_{BL_t}	Effect of L_t on B	$cells^{-1}\ h^{-1}$	9×10^{-10}
α_{BB}	Effect of B on B	$cells^{-1}\ h^{-1}$	-9×10^{-9}
b	Inhibition of *C. difficile* growth by DCA	mM^{-2}	4037
h	Production of TCA	$mM\ h^{-1}$	0.01
g	Germination rate of *C. difficile* spores	$h^{-1}\ mM^{-1}$	10
v_T	Production of CA from TCA	$mM\ cells^{-1}\ h^{-1}$	4×10^{-8}
v_C	Production of DCA from CA	$mM\ cells^{-1}\ h^{-1}$	4×10^{-8}
M_T	Half saturation of CA production	mM	0.6325
M_C	Half saturation of DCA production	mM	0.6325
δ_T	TCA decay rate	h^{-1}	0.1
δ_C	CA decay rate	h^{-1}	0.1
δ_D	DCA decay rate	h^{-1}	0.1

All interaction coefficients α_{ij} are summarized from data from Stein et al. [37]. Growth rates were fit to in vitro experimental data from Theriot et al. [41]. Enzyme kinetics parameters for CA and DCA production were chosen to produce reasonable concentrations of TCA, CA, and DCA at the steady state [33]

4 Model Dynamics

4.1 Microbial Interaction Model Dynamics

Figure 3a depicts the pre-antibiotic gut simulation for the microbial interaction model (Eqs. 1–4) and shows how the interactions between the different groups of microbes can determine the make-up of the gut. *C. difficile* and Lactobacillaceae cells are quickly forced to zero steady states by the growth of Lachnospiraceae and the other bacteria, in agreement with [5, 41], where the dominant bacteria in a gut that has not been exposed to antibiotics fall largely into these two groups.

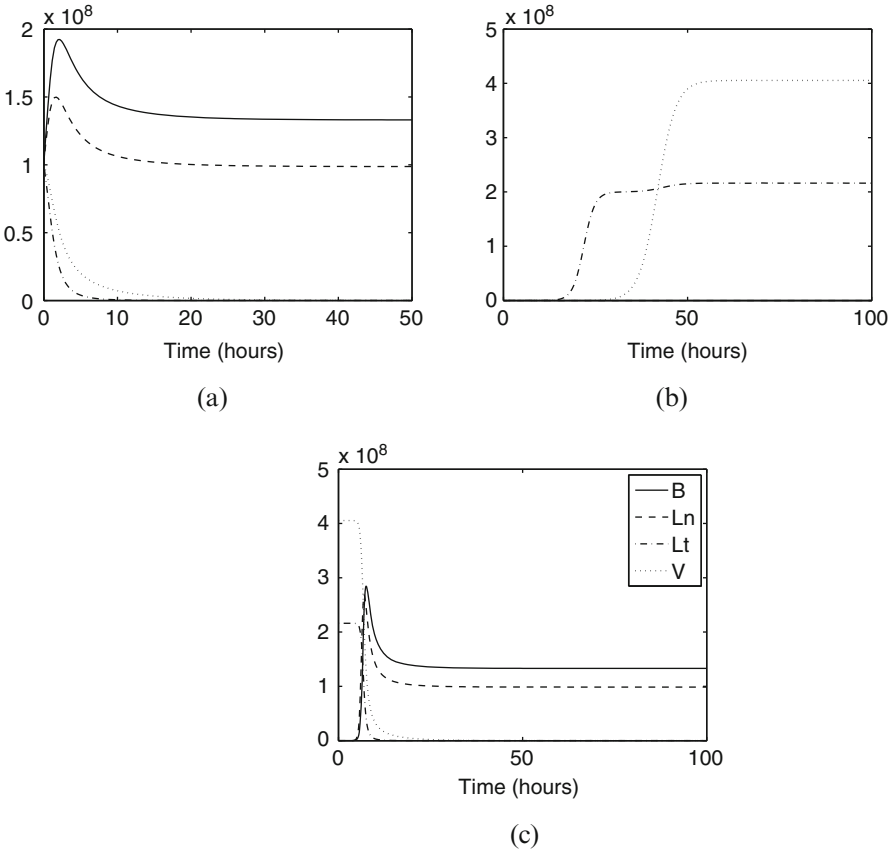


Fig. 3 Numerical solutions to the microbial interaction model (Eqs. 1–4). In (**a**) all four groups begin at equal concentrations and the interactions between the groups force V and L_t into zero steady states. Removing L_n and B in (**b**) enables *C. difficile* to establish colonization. Using the steady states of V and L_t from (**b**) as initial conditions in (**c**), and re-introducing L_n and B through $L_n(0) = B(0) = 1$ results in restoration of the pre-antibiotic gut microbiota

It can be shown mathematically that this steady state (where $V = L_t = 0$ but L_n and B co-exist) is locally stable under our parameter regime. To simulate the post-antibiotic gut, where antibiotics are assumed to kill the Lachnospiraceae and other bacteria, we must remove these two groups from the system entirely to avoid returning to the pre-antibiotic steady state. Setting $B(0) = L_n(0) = 0$ and assuming small numbers of V and L_t are present (Fig. 3b) enables the Lactobacillaceae and *C. difficile* to co-exist. Lactobacillaceae dominates initially, but *C. difficile* eventually grows to higher numbers, reaching levels where it could likely establish colonization in the gut. Other bacteria inhibit the three remaining groups (see Table 2). The Lachnospiraceae also inhibit all groups apart from other bacteria, which they promote, hence reinforcing the dominance of other bacteria when present. Therefore, removing these two groups essentially gives *C. difficile* and the Lactobacillaceae the freedom to grow. The Lactobacillaceae have a higher growth rate, so they emerge initially. Both *C. difficile* and the Lactobacillaceae promote each other's growth, but the Lactobacillaceae do this more strongly in our parameter set ($\alpha_{VL_t} > \alpha_{L_tV}$), hence *C. difficile* will eventually take over and dominate the gut, i.e., though the Lactobacillaceae have a higher growth rate than *C. difficile*, the *C. difficile* bacteria can be considered to have a higher relative fitness under our parameter set. This replicates the microbiome dynamics detected in [40] well.

In the absence of L_n and B, the model of microbial interactions is reduced to a two-dimensional system with equations for V and L_t. For the V, L_t system, the co-existence equilibrium is locally stable [29] if

$$\frac{\alpha_{VL_t}\alpha_{L_tV}}{\alpha_{VV}\alpha_{L_tL_t}} < 1.$$

The condition holds when intragroup competition dominates over the V, L_t cooperative interaction; in this simple system with our parameter values, in the numerator α_{VL_t} and α_{L_tV} are both positive while in the denominator, α_{VV} and $\alpha_{L_tL_t}$ are both negative. When considering the full microbial interactions model with four equations, it is interesting to note that the equilibrium with $L_n = B = 0$ and V and L_t positive is unstable. Thus re-introducing B and L_n into the model following *C. difficile* colonization enables fast restoration of the gut to the pre-antibiotic stable steady state (Fig. 3c). We note that in reality this would be unlikely to occur, at least on this timescale, due to the host of post-colonization mechanisms employed by *C. difficile* to establish infection that are not included in this model. Interestingly, re-introducing either B or L_n without the other is not sufficient to restore the gut to eradicate the *C. difficile* cells (Fig. 4). We note that these results are highly dependent on the values of the interaction parameters and require further investigation once more data is available for parameterization of the model.

Using our parameters, we investigated all the feasible equilibria (non-negative components) for this system. Besides the two cases mentioned above, note that there are also three other feasible, unstable equilibria with two components being zero: $L_n = L_t = 0, B = V = 0, B = L_t = 0$, i.e., other bacteria and *C. difficile* cells can co-exist, as can Lachnospiraceae and Lactobacillaceae or *C. difficile* and

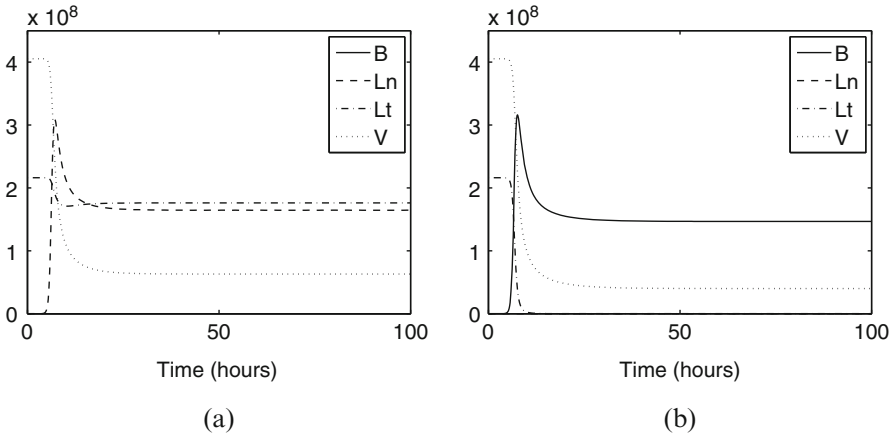


Fig. 4 The microbial interaction model with $V(0)$ and $L_t(0)$ taken to be the steady states of Fig. 3b and (**a**) $L_n(0) = 1, B(0) = 0$ (**b**) $L_n(0) = 0, B(0) = 1$. In either of these scenarios co-existence with *C. difficile* can occur

Lachnospiraceae. There is one feasible equilibrium with only one zero component: when $B = 0$, there is an unstable equilibrium with the other three taxa able to survive together. Equilibria with three zero components are not relevant here since they only tell us that each of the microbial taxa can survive in isolation. Interestingly, there is not a feasible equilibrium with four positive components: under our parameter choice at least one microbial group will not be able to survive in the presence of the others, matching what is seen in mouse models of *C. difficile* infection, e.g., [16, 40]. The full stability analysis therefore tells us that the only feasible and stable steady state is that where other bacteria and Lachnospiraceae co-exist while suppressing *C. difficile* and the Lactobacillaceae.

4.2 Bile Acid Model Dynamics

Considering the effect of metabolites on *C. difficile* in isolation enables us to track the effect of bile acids on the germination of spores and outgrowth of vegetative *C. difficile* cells. Solving Eqs. (5)–(9) numerically in Fig. 5 allows us to compare the pre-antibiotic gut microbiota (where B and L_n are present and can catalyze the conversion of TCA to CA and CA to DCA, respectively) to the post-antibiotic gut microbiota metabolism where neither B nor L_n are present. Spore germination, triggered largely by TCA, occurs much faster in the post-antibiotic model but outgrowth of vegetative cells does occur in both cases. Indeed, V reaches the same steady state under both conditions, but in the pre-antibiotic case (Fig. 5, solid line) this level is attained over the unrealistic timescale of 2000 h (simulation not illustrated over that timescale). Over the 5 days simulated, while vegetative *C.*

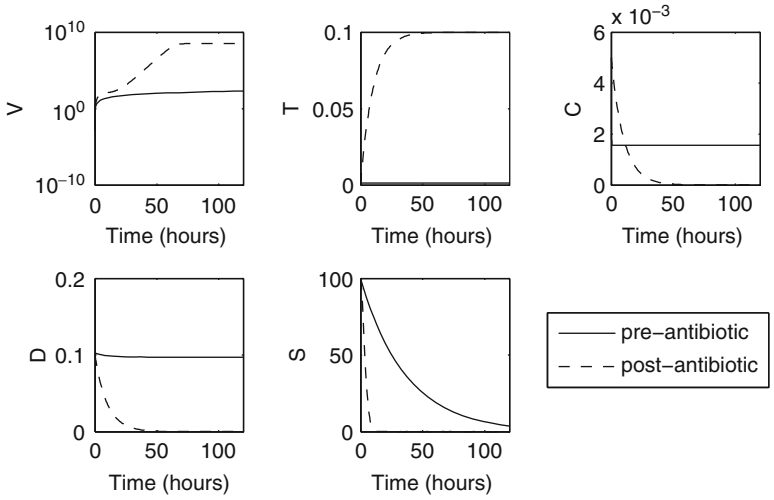


Fig. 5 Numerical simulations to the bile acid model (Eqs. 5–9) in the pre-antibiotic regime (B and L_n taken to be the steady states of the pre-antibiotic microbial interaction model, *solid line*) and post-antibiotic ($B = L_n = 0$, *dashed line*). Note that V is plotted on a log-scale for clarity. Initial conditions are taken from Table 4

difficile cells are present in the pre-antibiotic model, they don't reach what might be deemed significant levels for colonization. Conversely, in the post-antibiotic model (Fig. 5, dashed line) these levels are reached within 4 days (which is a feasible timeframe). Bile acids alone, therefore, cannot represent the full long-term mechanisms of colonization resistance in the pre-antibiotic state. Another factor, whether that is microbial interactions or the immune response, say, must also play a part. However, given that experimental evidence does suggest that bile acids are involved in protecting the gut from colonization (e.g., through inhibition of *C. difficile* growth [36, 41]), we can explore the model further to elucidate this role.

The sensitivity analysis in Sect. 4.2.1 below suggests further investigation into the degradation rate of DCA, δ_D. At sufficiently high levels, DCA inhibits growth of *C. difficile* (Fig. 2b). Therefore, the rate at which it degrades out of the system once antibiotics have cleared the gut of Lachnospiraceae bacteria, preventing more DCA from being produced, will be instrumental in determining vegetative *C. difficile* dynamics. Since DCA will always eventually degrade out of the system in the absence of Lachnospiraceae, δ_D does not affect final *C. difficile* numbers; however, the timing of the onset of colonization is affected, occurring earlier with faster degradation of DCA (see Fig. 6). This model therefore suggests that bile acids may be more important in delaying colonization than in ultimately preventing it. Of course, we must consider that there are likely to be additional factors besides those included in this model that could also affect DCA levels over time.

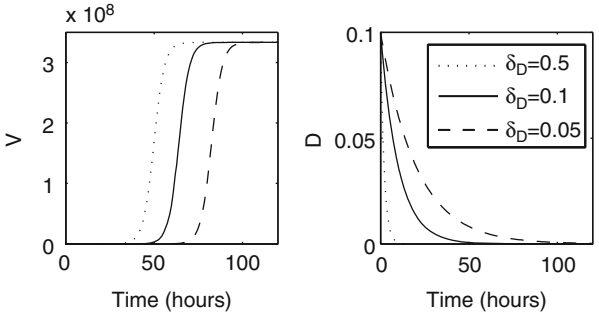


Fig. 6 Numerical solutions to the bile acid model under post-antibiotic conditions ($B = L_n = 0$) and varying the degradation rate of DCA, δ_D (default value is $\delta_D = 0.1$). Lowering the value of δ_D delays the onset of colonization

4.2.1 Sensitivity Analysis on Bile Acid Model

To assess the effect of specific bile acid model parameters on *C. difficile* dynamics, we performed a sensitivity analysis on the model following [21] with a model output of maximum *C. difficile* vegetative cells after either 6, 48, or 120 h. Model parameters were varied uniformly over the ranges shown in Table 7 using Latin Hypercube Sampling. Methods and code from [15] and [21] were used to calculate partial rank correlation coefficients (PRCCs). PRCC values and their corresponding p-values are reported in Table 7.

After 6 h, maximum *C. difficile* levels are most sensitive to the decay rate of DCA (δ_D, see Fig. 6), the intrinsic growth rate of *C. difficile* (r_V), and the coefficient b, which determines the strength of inhibition of the growth rate of *C. difficile* by DCA. Although the PRCCs for g, h, and α_{VV} are also statistically significant ($p < 0.0001$), their values are low indicating they do not have a large effect on maximum *C. difficile* levels.

As time since introduction of spores increases, maximum *C. difficile* levels become more sensitive to α_{VV} and less sensitive to r_V and δ_D. By 120 h, *C. difficile* cells have either reached or grown close to carrying capacity. At this time maximum *C. difficile* levels are most sensitive to α_{VV} and r_V, parameters that determine the value of this carrying capacity. The PRCC for δ_D is still significant but decreases from 0.8858 to 0.1138. This is not surprising as the decay rate of DCA has a large effect on the rate of growth of *C. difficile*, affecting when carrying capacity is reached but not the value of the carrying capacity itself (Fig. 6).

4.3 Combined Model Dynamics

Figure 7 depicts the numerical solution to the combined microbial interaction and bile acid model, i.e., Eqs. (10)–(17). We see the results of the microbial interaction model reproduced in the sense that for the pre-antibiotic case Lachnospiraceae and

Table 7 Partial Rank Correlation Coefficient (PRCC) and p-values from a global sensitivity analysis of the bile acid model with $N = 5000$ (number of model runs)

		After 6 h		After 48 h		After 120 h	
Parameter	Range	PRCC	p-value	PRCC	p-value	PRCC	p-value
δ_D	$(10^{-4}, 1)$	**0.8858**	<0.0001	**0.3326**	<0.0001	**0.1138**	<0.0001
r_V	$(0.1, 1)$	**0.8407**	<0.0001	**0.7576**	<0.0001	**0.4986**	<0.0001
b	$(10^3, 10^4)$	**−0.4280**	<0.0001	−0.0244	0.0854	−0.0141	0.3185
g	$(0.1, 100)$	**0.1031**	<0.0001	0.0034	0.8077	−0.0180	0.2035
h	$(10^{-4}, 1)$	**0.0853**	<0.0001	0.0080	0.5708	−0.0196	0.1656
α_{VV}	$(-10^{-8}, -10^{-10})$	**0.0305**	<0.0001	**0.5127**	<0.0001	**0.7979**	<0.0001
v_C	$(10^{-10}, 0.1)$	0.0189	0.1812	0.0070	0.6216	−0.0108	0.4471
δ_T	$(10^{-4}, 1)$	−0.0169	0.2316	−0.0172	0.2236	0.0149	0.2913
Ln	$(0, 10^{10})$	−0.0163	0.2485	−0.0089	0.5288	−0.0049	0.7300
M_T	$(0.01, 1)$	0.0157	0.2678	0.0010	0.9421	−0.0129	0.3626
B	$(0, 10^{10})$	−.0142	0.3159	0.0168	0.2370	0.0241	0.0885
v_T	$(10^{-10}, 0.1)$	−0.0125	0.3786	−0.0007	0.9609	0.0099	0.4837
M_C	$(0.01, 1)$	0.0109	0.4405	0.0152	0.2828	0.0017	0.9070
δ_C	$(10^{-4}, 1)$	0.0104	0.4644	−0.0003	0.9834	0.0125	0.3770

Output is maximum *C. difficile* vegetative cells after 6, 48, and 120 h. Statistically significant PRCCs ($p < 0.0001$) are bolded

other bacteria co-exist, with the situation reversed in the post-antibiotic model. The system achieves near identical steady states to the microbial interaction model and here we are not seeing the role for bile acids in the loss of colonization resistance in the post-antibiotic simulation. However, drawing on our investigations in Sect. 4.2, we simulate the combined model with varying values of δ_D, the degradation rate of DCA (Fig. 8), and again see that this secondary bile acid may play a crucial role in the onset of colonization in the post-antibiotic case. Thus it is possible that each of the two colonization-resistance mechanisms we have modeled have their own purpose: following disruption of the microbiome by antibiotics, bile acids could effectively delay the onset of colonization, providing a window of opportunity for microbial competition to restore the flora to a *C. difficile*-resistant state. We demonstrate this in Fig. 9. Note that, as in the microbial interaction model, restoration would occur even if B and L_n are re-introduced when V is at colonization-indicative levels because we have not included any post-colonization mechanisms in the current model formulation that could prevent this occurring. Nevertheless, we include this simulation for illustrative purposes. As in the microbial interaction model, *both* L_n and B must be introduced to restore the gut, i.e., neither one can achieve this in isolation.

A possible therapy to prevent *C. difficile* infection therefore could be to manipulate DCA production by maintaining it in a sufficiently high state to delay the onset of colonization for long enough for the pre-antibiotic flora to restore itself.

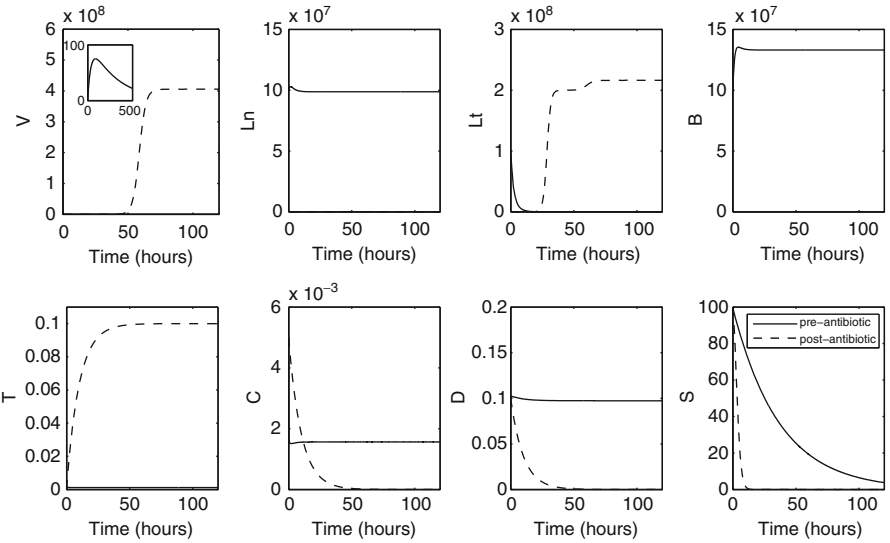


Fig. 7 Numerical solution to the combined model (Eqs. 10–17) for the pre- (*solid line*) and post-antibiotic (*dashed line*) conditions. In the first panel we use an inset to illustrate that spores do germinate to cause a rise in *C. difficile* vegetative cells, but microbial competition quickly suppresses these so that this is not visible on the outer figure. In the post-antibiotic simulation, both L_n and B are zero throughout the timecourse

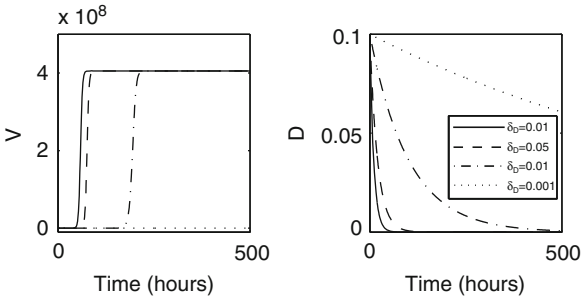


Fig. 8 Numerical solution to the combined model under a range of values of δ_D and the post-antibiotic conditions. The slower DCA degrades, the longer colonization can be prevented

4.3.1 Sensitivity Analysis on Combined Model

To assess the effect of the parameters in the combined model, we again performed a global sensitivity analysis with a model output of maximum *C. difficile* vegetative cells. We looked at this output at 6, 48, and 120 h after the introduction of *C. difficile* spores into a post-antibiotic gut, using the initial conditions in Table 5. Parameter ranges, PRCCs, and corresponding p-values can be found in Table 8. The interaction coefficients α_{ij} were held constant at their values in Table 6 and were not included in the sensitivity analysis.

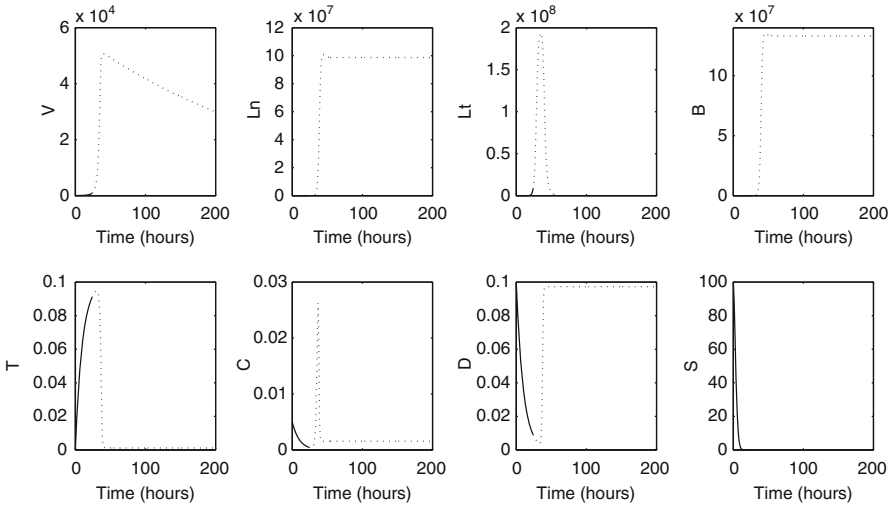


Fig. 9 Numerical simulation of the combined model under post-antibiotic conditions (i.e., $B(0) = L_n(0) = 0$) but where B and L_n are re-introduced to the gut at time $t = 24$ (i.e., $B(24) = L_n(24) = 1000$—values chosen for illustrative purposes, lower numbers would achieve equivalent results in the long term). The addition of microbial interactions enables the gut to inhibit *C. difficile* vegetative cell growth sufficiently to prevent colonization and eradicate the pathogen from the gut (over a longer timescale than shown here). *Solid lines* are used to illustrate $t \leq 24$ and *dotted* for $t > 24$

We find that after 6 h, maximum *C. difficile* vegetative cells are most sensitive to the same five parameters as in the bile acid model (δ_D, r_V, b, g, and h), with PRCCs of similar magnitude. After 48 h, the growth rate of *C. difficile* vegetative cells (r_V) still has the highest PRCC. However, the next highest PRCC is the growth rate of Lactobacillaceae (r_{Lt}). Increasing r_{Lt} allows Lactobacillaceae to grow to high levels more quickly, where they are able to have a positive impact on the growth of *C. difficile* vegetative cells (since $\alpha_{VLt} > 0$). Sensitivity to the growth rate of Lactobacillaceae increases with time since introduction of *C. difficile* spores, and at 120 h r_{LT} has the highest PRCC.

5 Discussion

Earlier models of colonization resistance explicitly addressed two mechanisms: competition for a single nutrient and competition for adhesion sites [9]. Recent research has identified numerous mechanisms by which resident gut microbiota can inhibit pathogens including: (1) competition for nutrients, (2) indirect inhibition mediated by the host immune system, (3) metabolic exclusion by short chain fatty acids, (4) direct inhibition by bacteriocins, and (5) inhibition by bile acids [18, 39].

Table 8 Partial Rank Correlation Coefficient (PRCC) and *p*-values from a global sensitivity analysis of combined model under post-antibiotic conditions with N = 5000 (number of model runs)

		After 6 h		After 48 h		After 120 h	
Parameter	Range	PRCC	*p*-value	PRCC	*p*-value	PRCC	*p*-value
δ_D	$(10^{-4}, 1)$	**0.9252**	< 0.0001	**0.4701**	< 0.0001	**0.1882**	< 0.0001
r_V	$(0.1, 0.8)$	**0.8253**	< 0.0001	**0.8065**	< 0.0001	**0.6467**	< 0.0001
b	$(10^3, 10^4)$	**−0.4944**	< 0.0001	**−0.0675**	< 0.0001	−0.0207	0.1439
g	$(0.1, 100)$	**0.0832**	< 0.0001	0.0026	0.8570	−0.0201	0.1552
h	$(10^{-4}, 1)$	**0.0766**	< 0.0001	0.0183	0.1967	0.0200	0.1581
δ_T	$(10^{-4}, 1)$	−0.0277	0.0502	−0.0154	0.2762	−0.0274	0.0534
M_C	$(0.01, 1)$	−0.0127	0.3704	−0.0055	0.6987	0.0162	0.2539
δ_C	$(10^{-4}, 1)$	−0.0108	0.4457	−0.0016	0.9109	−0.0015	0.9158
r_B	$(0.1, 0.8)$	−0.0067	0.6349	0.0116	0.4110	0.0052	0.7142
υ_C	$(10^{-10}, 0.1)$	−0.0047	0.7425	−0.0163	0.2498	0.0039	0.7835
r_{Lt}	$(0.1, 0.8)$	0.0038	0.7864	**0.4869**	< 0.0001	**0.7003**	< 0.0001
M_T	$(0.01, 1)$	0.0032	0.8208	−0.0003	0.9815	0.0037	0.7955
r_{Ln}	$(0.1, 0.8)$	−0.0022	0.8751	−0.0003	0.9847	0.0209	0.1391
υ_T	$(10^{-10}, 0.1)$	−0.0007	0.9629	−0.0150	0.2909	−0.0321	0.0235

Output is maximum *C. difficile* vegetative cells after 6, 48, and 120 h. Statistically significant PRCCs ($p < 0.0001$) are bolded

In particular, bile acids interact with *C. difficile* in a complex way. In our paper, we model the interactions between bile acids and *C. difficile* explicitly, whereas we model the remaining mechanisms in an aggregate way using Lotka–Volterra equations as our main objective is to elucidate how bile acids influence colonization resistance against *C. difficile*.

The microbial interaction model suggests that *C. difficile* cannot colonize the gut in the presence of resident microbiota represented in the model by the Lachnospiraceae (L_n) and other bacteria (B). The steady state in which *C. difficile* vegetative cells (V) and Lactobacillaceae (L_t) are absent and L_n and B co-exist is shown to be stable. For the post-antibiotic scenario, both groups (B and L_n) are necessary to suppress *C. difficile* colonization. Re-introducing either B or L_n alone is not sufficient to eliminate *C. difficile*. Therefore, perturbations that extensively reduce one or both bacterial populations can compromise colonization resistance.

These results align with the findings in both mouse and human studies looking at the interaction between the gut microbiota, antibiotics, and *C. difficile* infection (CDI). Prior to antibiotics the indigenous gut microbiota provides colonization resistance against *C. difficile* and it is not until perturbation by antibiotics that the microbial community allows for *C. difficile* colonization [32, 40]. The loss of gut bacterial diversity and members from the Lachnospiraceae and Ruminococcaceae families is associated with antibiotic use and *C. difficile* colonization [41]. Restoration of both is associated with resistance against *C. difficile* [40, 41]. Members of the Lachnospiraceae family include many anaerobic Clostridia, which overlap in

metabolic niche with *C. difficile*, potentially providing competition for nutrients in the gut after antibiotics. Mono-colonization with Lachnospiraceae strains in a germfree mouse ameliorated disease from CDI but did not eliminate *C. difficile* [16]. More recently, *C. scindens* and non-toxigenic *C. difficile*, both Clostridia, were able to decrease disease in an antibiotic treated CDI mouse model, although the precise mechanism is still unknown [7, 23]. There is also evidence of this in human studies, where a loss in bacterial diversity and members from the Lachnospiraceae and Ruminococcaceae families are associated with CDI [1, 35].

The analysis of the bile acid and combined models indicates that bile acids do not prevent *C. difficile* colonization, but they regulate the onset of *C. difficile* colonization and growth after antibiotic perturbation. This effect on the timing of colonization was particularly sensitive to DCA decay rate (δ_D) and the inhibition rate of *C. difficile* growth by DCA (b). In our models, the degradation of DCA was assumed to be first order kinetics, depending only on DCA concentration and δ_D was assumed constant and identical in both the pre- and post-antibiotic scenarios. It is plausible that DCA degradation dynamics are more complex. Model outcomes were also sensitive to the growth rate of *C. difficile* vegetative cells (r_V). Nevertheless, this opens up the possibility of somehow promoting DCA production to delay the onset of colonization and buy time for the pre-antibiotic microflora to be restored. We can test this experimentally in the future by treating antibiotic treated mice that are susceptible to *C. difficile* colonization with DCA orally. We can then challenge mice with *C. difficile* and define the level of colonization with and without the addition of DCA. We can also define the gut microbiota in these mice to measure the restoration of the gut microbiota after antibiotic treatment with and without the addition of DCA.

Similar to the combined model, it is evident that the production of secondary bile acids, such as DCA, by the gut microbiota is important for colonization resistance against *C. difficile*, reviewed here [47]. Secondary bile acids like DCA inhibit *C. difficile* growth in vitro and are also associated with resistance against *C. difficile* in mouse models [7, 36, 40]. More recently, patients with recurrent *C. difficile* that receive a fecal microbial transplant (FMT) show restoration of their fecal secondary bile acids, suggesting that microbial derived secondary bile acids could play an important role in the process of clearing CDI in humans as well [44]. However, the complex interactions between the microbiota, DCA, and other bile acids in the gut need further investigation including studying kinetics, flux, and degradation dynamics over time in a gut that is healthy, antibiotic treated, and *C. difficile* colonized. Our model provides an early step in achieving this goal.

Our model can be expanded and refined in several ways. First, our model focuses on *C. difficile* colonization. However, once *C. difficile* reaches a high density in the gut, it produces toxins A and B, which cause gut damage. To address infection, the model should include the effects of toxins and the associated host response. Second, the effects of antibiotics on the gut microbiota were addressed in the model by modifying the initial conditions. Antibiotic perturbation in the gut microbiota could be included explicitly. Finally, microbial interactions were modeled using the Lotka–Volterra equations, but nutrients and other interactions could be addressed in the model more explicitly by incorporating nutrient and other intermediate metabolites.

Acknowledgements The work was partially funded through the Research Collaborative Workshop for Women in the Mathematical Biology at the National Institute for Mathematical and Biological Synthesis, sponsored by the National Science Foundation through NSF Award DBI-1300426, with additional support from The University of Tennessee, Knoxville. Jabbari thanks the BBSRC and MRC for their support in the form of the grant awards BB/M021386/1 and G1002093. The work of Lanzas and Lenhart was partially supported by the joint NSF/NIGMS Mathematical Biology Program through NIH award R01GM113239. Theriot is funded by Career Development Award in Metabolomics grant K01GM109236 and R35GM119438 from the NIH NIGMS.

References

1. Antharam, V.C., Li, E.C., Ishmael, A., Sharma, A., Mai, V., Rand, K.H., Wang, G.P.: Intestinal dysbiosis and depletion of butyrogenic bacteria in *Clostridium difficile* infection and nosocomial diarrhea. J. Clin. Microbiol. **51**(9), 2884–2892 (2013)
2. Antunes, L.C.M., Han, J., Ferreira, R.B.R., Lolić, P., Borchers, C.H., Finlay, B.B.: Effect of antibiotic treatment on the intestinal metabolome. Antimicrob. Agents Chemother. **55**(4), 1494–1503 (2011)
3. Artis, D.: Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. Nat. Rev. Immunol. **8**(6), 411–420 (2008)
4. Ballyk, M., Smith, H.: A model of microbial growth in a plug flow reactor with wall attachment. Math. Biosci. **158**(2), 95–126 (1999)
5. Bassis, C.M., Theriot, C.M., Young, V.B.: Alteration of the murine gastrointestinal microbiota by tigecycline leads to increased susceptibility to *Clostridium difficile* infection. Antimicrob. Agents Chemother. **58**(5), 2767–2774 (2014)
6. Bouillaut, L., Self, W.T., Sonenshein, A.L.: Proline-dependent regulation of *Clostridium difficile* Stickland metabolism. J. Bacteriol. **195**(4), 844–854 (2013)
7. Buffie, C.G., Bucci, V., Stein, R.R., McKenney, P.T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., et al.: Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. Nature **517**(7533), 205–208 (2015)
8. Coleman, M.E., Dreesen, D.W., Wiegert, R.G.: A simulation of microbial competition in the human colonic ecosystem. Appl. Environ. Microbiol. **62**(10), 3632–3639 (1996)
9. Freter, R., Brickner, H., Fekete, J., Vickerman, M.M., Carey, K.E.: Survival and implantation of *Escherichia coli* in the intestinal tract. Infect. Immun. **39**(2), 686–703 (1983)
10. Grima, D.T., Webb, G.F., D'Agata, E.M.C.: Mathematical model of the impact of a nonantibiotic treatment for *Clostridium difficile* on the endemic prevalence of vancomycin-resistant enterococci in a hospital setting. Comp. Math. Methods Med. **2012**, 1–8 (2012)
11. Hall, I., O'Toole, E.: Intestinal flora in new-born infants: with a description of a new pathogenic anaerobe, *Bacillus difficilis*. Am. J. Dis. Child **49**(2), 390–402 (1935)
12. Jabbari, S., Cartman, S.T., King, J.R.: Mathematical modelling reveals properties of TcdC required for it to be a negative regulator of toxin production in *Clostridium difficile.* J. Math. Biol. **70**, 773–804 (2015)
13. Karasawa, T., Ikoma, S., Yamakawa, K., Nakamura, S.: A defined growth medium for *Clostridium difficile*. Microbiology **141**(2), 371–375 (1995)
14. Karlsson, S., Lindberg, A., Norin, E., Burman, L.G., Åkerlund, T.: Toxins, butyric acid, and other short-chain fatty acids are coordinately expressed and down-regulated by cysteine in *Clostridium difficile*. Infect. Immun. **68**(10), 5881–5888 (2000)
15. Kirschner Lab - Publications: Uncertainty and Sensitivity Analysis. http://malthus.micro.med.umich.edu/lab/usanalysis.html (2015)
16. Koenigsknecht, M.J., Theriot, C.M., Bergin, I.L., Schumacher, C.A., Schloss, P.D., Young, V.B.: Dynamics and establishment of *Clostridium difficile* infection in the murine gastrointestinal tract. Infect. Immun. **83**(3), 934–941 (2015)

17. Lanzas, C., Dubberke, E.R., Lu, Z., Reske, K.A., Gröhn, Y.T.: Epidemiological model for *Clostridium difficile* transmission in healthcare settings. Infect. Control Hosp. Epidemiol. **32**(6), 553–561 (2011)
18. Lawley, T.D., Walker, A.W.: Intestinal colonization resistance. Immunology **138**(1), 1–11 (2013)
19. Leatham, M.P., Banerjee, S., Autieri, S.M., Mercado-Lubo, R., Conway, T., Cohen, P.S.: Precolonized human commensal *Escherichia coli* strains serve as a barrier to *E. coli* O157: H7 growth in the streptomycin-treated mouse intestine. Infect. Immun. **77**(7), 2876–2886 (2009)
20. Lessa, F.C., Mu, Y., Bamberg, W.M., Beldavs, Z.G., Dumyati, G.K., Dunn, J.R., Farley, M.M., Holzbauer, S.M., Meek, J.I., Phipps, E.C., et al.: Burden of *Clostridium difficile* infection in the United States. New Engl. J. Med. **372**(9), 825–834 (2015)
21. Marino, S., Hogue, I.B., Ray, C.J., Kirschner, D.E.: A methodology for performing global uncertainty and sensitivity analysis in systems biology. J. Theor. Biol. **254**(1), 178–196 (2008)
22. McKay, I.C., Speekenbrink, A.: Implications of Freter's model of bacterial colonization. Infect. Immun. **44**(1), 199–203 (1984)
23. Merrigan, M.M., Sambol, S.P., Johnson, S., Gerding, D.N.: New approach to the management of *Clostridium difficile* infection: colonisation with non-toxigenic *C. difficile* during daily ampicillin or ceftriaxone administration. Int. J. Antimicrob. Agents **33**, S46–S50 (2009)
24. Nagaro, K.J., Phillips, S.T., Cheknis, A.K., Sambol, S.P., Zukowski, W.E., Johnson, S., Gerding, D.N.: Nontoxigenic *Clostridium difficile* protects hamsters against challenge with historic and epidemic strains of toxigenic BI/NAP1/027 *C. difficile*. Antimicrob. Agents Chemother. **57**(11), 5266–5270 (2013)
25. Nakamura, S., Nakashio, S., Yamakawa, K., Tanabe, N., Nishida, S.: Carbohydrate fermentation by *Clostridium difficile*. Microbiol. Immunol. **26**(2), 107–111, 1982.
26. O'Flaherty, S.J., Klaenhammer, T.R.: Functional and phenotypic characterization of a protein from *Lactobacillus acidophilus* involved in cell morphology, stress tolerance and adherence to intestinal cells. Microbiology **156**(11), 3360–3367 (2010)
27. Otete, E.H., Ahankari, A.S., Jones, H., Bolton, K.J., Jordan, C.W., Boswell, T.C., Wilcox, M.H., Ferguson, N.M., Beck, C.R., Puleston, R.L.: Parameters for the mathematical modelling of *Clostridium difficile* acquisition and transmission: a systematic review. PLoS One **8**, e84224 (2013)
28. Owens, R.C., Donskey, C.J., Gaynes, R.P., Loo, V.G., Muto, C.A.: Antimicrobial-associated risk factors for *Clostridium difficile* infection. Clin. Infect. Dis. **46**(Supplement 1), S19–S31 (2008)
29. Pastor, J.: Mathematical Ecology of Populations and Ecosystems. Wiley-Blackwell, Singapore (2011)
30. Pépin, J., Saheb, N., Coulombe, M.-A., Alary, M.-E., Corriveau, M.-P., Authier, S., Leblanc, M., Rivard, G., Bettez, M., Primeau, V., et al.: Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*–associated diarrhea: a cohort study during an epidemic in Quebec. Clin. Infect. Dis. **41**(9), 1254–1260 (2005)
31. Rea, M.C., Sit, C.S., Clayton, E., O'Connor, P.M., Whittal, R.M., Zheng, J., Vederas, J.C., Ross, R.P., Hill, C.: Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. Proc. Natl. Acad. Sci. U. S. A. **107**(20): 9352–9357, 2010.
32. Reeves, A.E., Theriot, C.M., Bergin, I.L., Huffnagle, G.B., Schloss, P.D., Young, V.B.: The interplay between microbiome dynamics and pathogen dynamics in a murine model of *Clostridium difficile* infection. Gut Microbes **2**(3), 145–158 (2011)
33. Ridlon, J.M., Kang, D.-J., Hylemon, P.B.: Bile salt biotransformations by human intestinal bacteria. J. Lipid Res. **47**(2), 241–259 (2006)
34. Scaria, J., Chen, J.-W., Useh, N., He, H., McDonough, S.P., Mao, C., Sobral, B., Chang, Y.-F.: Comparative nutritional and chemical phenome of *Clostridium difficile* isolates determined using phenotype microarrays. Int. J. Infect. Dis. **27**, 20–25 (2014)
35. Schubert, A.M., Rogers, M.A.M., Ring, C., Mogle, J., Petrosino, J.P., Young, V.B., Aronoff, D.M., Schloss, P.D.: Microbiome data distinguish patients with *Clostridium difficile* infection and non-*C. difficile*-associated diarrhea from healthy controls. MBio **5**(3), e01021–14 (2014)

36. Sorg, J.A., Sonenshein, A.L.: Bile salts and glycine as cogerminants for *Clostridium difficile* spores. J. Bacteriol. **190**(7), 2505–2512 (2008)
37. Stein, R.R., Bucci, V., Toussaint, N.C., Buffie, C.G., Ratsch, G., Pamer, E.G., Sander, C., Xavier, J.B.: Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Comput. Biol. **9**, e1003388 (2013)
38. Steinway, S.N., Biggs, M.B., Loughran Jr, T.P., Papin, J.A., Albert, R.: Inference of network dynamics and metabolic interactions in the gut microbiome. PLoS Comput. Biol. **11**(6), e1004338 (2015)
39. Theriot, C.M., Young, V.B.: Interactions between the gastrointestinal microbiome and *Clostridium difficile*. Annu. Rev. Microbiol. **69**, 445–461 (2015)
40. Theriot, C.M., Koenigsknecht, M.J., Carlson Jr, P.E., Hatton, G.E., Nelson, A.M., Li, B., Huffnagle, G.B., Li, J.Z., Young, V.B.: Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. Nat. Commun. **5** (2014)
41. Theriot, C.M., Bowman, A.A., Young, V.B.: Antibiotic-induced alterations of the gut microbiota alter secondary bile acid production and allow for *Clostridium difficile* spore germination and outgrowth in the large intestine. mSphere **1**(1), e00045–15 (2016)
42. Tsilimigras, M.C.B., Fodor, A.A.: Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Ann. Epidemiol. **26**(6), 330–335 (2016)
43. Vollaard, E.J., Clasener, H.A.: Colonization resistance. Antimicrob. Agents Chemother. **38**(3), 409 (1994)
44. Weingarden, A.R., Chen, C., Bobr, A., Yao, D., Lu, Y., Nelson, V.M., Sadowsky, M.J., Khoruts, A.: Microbiota transplantation restores normal fecal bile acid composition in recurrent *Clostridium difficile* infection. Am. J. Physiol. Gastrointest. Liver Physiol. **306**(4), G310–G319 (2014)
45. Wilson, K.H., Perini, F.: Role of competition for nutrients in suppression of *Clostridium difficile* by the colonic microflora. Infect. Immun. **56**(10), 2610–2614 (1988)
46. Wilson, K.H., Kennedy, M.J., Fekety, F.R.: Use of sodium taurocholate to enhance spore recovery on a medium selective for *Clostridium difficile*. J. Clin. Microbiol. **15**(3), 443–446 (1982)
47. Winston, J.A., Theriot, C.M.: Impact of microbial derived secondary bile acids on colonization resistance against *Clostridium difficile* in the gastrointestinal tract. Anaerobe (2016). doi:10.1016/j.anaerobe.2016.05.003

Revisiting the Physics of Spider Ballooning

Kimberly S. Sheldon, Longhua Zhao, Angela Chuang, Iordanka N. Panayotova, Laura A. Miller, and Lydia Bourouiba

Abstract Spiders use a unique type of aerial dispersal called "ballooning" to move from one location to another. In order to balloon, a spider must first release one or more flexible, elastic, silk draglines from its spinnerets. Once enough force is generated on the dragline(s), the spider becomes airborne. This "take-off" stage of ballooning is followed by the "flight" stage and finally the "settling" stage when spiders land in a new location. Though the ecology of spider ballooning is well understood, little is known about the physical mechanisms. This is in part due to the significant challenge of describing the relevant physics for spiders that are ballooning across large distances. One difficulty, for example, is that properties of both the spider, such as body size and shape, and the silk dragline(s) can vary among species and individuals. In addition, the relevant physics may differ among the three stages of ballooning. Finally, models must take into account the interaction between the flexible dragline and air, and resolving this multi–scale, fluid–structure interaction can be particularly difficult. Here, we review the literature on spider ballooning, including the relevant physics, meteorological conditions that

K.S. Sheldon (✉) • A. Chuang
Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA
e-mail: ksheldon@utk.edu; angelachuang@utk.edu

L. Zhao
Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
e-mail: longhua.zhao@case.edu

I.N. Panayotova
Department of Mathematics, Christopher Newport University, Newport News, VA 23606, USA
e-mail: iordanka.panayotova@cnu.edu

L.A. Miller
Departments of Mathematics and Biology, University of North Carolina, Chapel Hill, NC 27599, USA
e-mail: lam9@email.unc.edu

L. Bourouiba
The Fluid Dynamics of Disease Transmission Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02130, USA
e-mail: lbouro@mit.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_9

favor ballooning, and previous mechanical models used to describe ballooning. We then highlight challenges and open questions relevant to future modeling of spider ballooning.

1 Ecology of Spider Dispersal

Dispersal, or the movement of individuals from their birth location to a new location [10], is a critical stage in the life cycle of many species. The benefits of dispersal include reduced competition for resources (e.g., reproductive mates and food) and inbreeding avoidance [11, 30, 48], which may ultimately lead to greater survival and reproduction.

Given the importance of dispersal for reproductive success, species have evolved a variety of mechanisms to move to new locations. One such mechanism that can be observed in spiders (Araneae) is a type of aerial dispersal called "ballooning" [5, 16]. To become airborne via ballooning, a spider first climbs to a high point and initiates a "tiptoe" posture in which it stands on the ends of its tarsi and elevates its abdomen (Fig. 1). The spider then releases one or more silk lines, or draglines, from its spinnerets [38]. With enough force on the dragline(s), the spider is lifted and becomes airborne [36, 38]. This "take-off" stage is followed by the "flight" stage in which individuals can travel as high as 5 km above ground [15] and as far as 3200 km in distance [23]. Finally, a spider reaches the "settling" stage when it lands in a new location.

Because adult spiders can have hundreds of spiderlings—or young spiders—from a single reproductive event, ballooning allows spiderlings to avoid competition and cannibalism from siblings and conspecific adults [44]. Though some species of spider are known to balloon at any time of year and any phenological stage [45], other species balloon only during specific time periods and clearly defined stages [13]. Regardless of when they disperse, ballooning spiders rarely have a mass greater than 100 mg [38, 46] (but see [34] for exception), which suggests constraints on body size for aerial dispersal. In addition, wind speeds for take-off appear to always be below 3 m/s [44].

Despite all we know about ballooning and its importance in the ecology and evolution of spiders, the physical mechanisms related to this unique type of dispersal are still poorly understood. Generating a model for spider ballooning seems simple in practice, but it is a complex problem given the variety of parameters involved. Ballooning is used by a host of species that come in different shapes and sizes. Both inertial and viscous forces in the surrounding air are important. Models must also consider dragline properties, including number of threads, length, elasticity, electrostatics, and compliance as well as constraints on wind speed [44] and spider mass [38, 46]. After incorporating all of the various parameters, one is left with a host of possible regimes, and different regimes could dominate during the three stages of spider ballooning (i.e., take-off, flight, and settling).

Fig. 1 Prior to ballooning, spiders display a tiptoe behavior which involves standing on the ends of the tarsi, angling the abdomen upward, and releasing one or more silk draglines from spinnerets. © Sarah Rose

Here, we review the existing literature on the physics of this fascinating mode of ecological dispersal and identify areas for future research. We first examine the relevant physics related to ballooning. We then discuss previous analytic models that researchers have used to parse out the dominant regimes. Finally, we highlight the remaining questions surrounding spider ballooning and how future research might address these questions.

2 Relevant Physics

2.1 The Physical Parameter Space

The parameters relevant to models of spider ballooning can be broadly characterized as those relating to individual morphology, dragline characteristics, and ambient environmental conditions. Individual morphology can be a constraining factor in flight [27], where mass determines the amount of vertical drag force required to lift off and sustain flight. Due to upper limits in dragline tensile strength and ambient air velocity, larger masses may completely preclude take-off for individuals [38, 42]. This can partially explain why ballooning as a dispersal mechanism is typically observed in younger instars and smaller taxa [3].

Body length and shape influence the amount of drag generated to sustain flight. Despite the large variation in body shapes among spider taxa that balloon (Fig. 2), all

Fig. 2 Spiders may display unique body shapes that cannot be realistically modeled as a sphere, including spiders in the genera (*clockwise from the top left*) *Cyrtarachne, Poltys, Gasteracantha,* and *Ariamnes*. Photo credit to Yung Yi Tsai, Fisher Chen, Robert Tsai, Sun Jong Liu

previous work on spider ballooning has used a spherical shape to model the spider [27, 31, 32]. Post-release posturing of the legs may have additional implications for terminal velocities [38], but the effects of a spider's eight legs on drag and flight speed have thus far been ignored.

The physical characteristics of the dragline act together with the body to create surfaces for aerodynamic drag to be generated. Dragline mass is generally considered negligible in models since it is minuscule compared to the spider's body mass. This is true even in cases where an individual may release two or more dragline threads, as has been observed across multiple ballooning taxa [14, 34]. Unlike mass, dragline length has been shown to be important in take-off dynamics and in decreasing terminal velocity [2, 37], but its influence during flight and on dispersal distance has been debated [31]. Because silk can flex and extend [31], distortion and bending during flight can decrease the effective length of the dragline. Interestingly, however, observations of entangled draglines during ballooning have not been reported.

Dragline flexibility influences the spider's movement through space. Since the dragline bends and adapts to flow structures in the air, it subjects the attached spider to variations in flow profiles. Thus, factors that influence dragline flexibility, including elasticity, stiffness, bending rigidity, and flexural rigidity, should be included in models.

Another important characteristic to consider when modeling ballooning is the electrostatic properties of silk [28]. Electrically charged silk may explain how a long, single dragline can avoid becoming entangled during flight. In cases where dozens of silken threads are released together, electrostatic repulsion in individual strands may similarly prevent mass entanglement [34]. Electrostatic forces are speculated to act independently from adjective and aerodynamic forces on the spider, thus providing additional lift [19].

Lastly, environmental conditions play a substantial role in spider take-off, flight, and landing. Meteorological conditions seem to affect the initiation of tiptoe behavior; multiple field observations suggest that spiders balloon during daylight hours, under sunny and clear skies, and at wind speeds less than 3 m/s [13, 22, 27, 38, 44]. Wind speed has strong implications for take-off velocities and distance traveled. Horizontal wind velocity is directly related to the horizontal velocity that ballooning spiders attain [27, 37], whereas vertical wind velocity and the resulting drag generated counteracts the falling speed of the body. Given the small spatial scales in which ballooning spiders operate, minor variations in meteorological conditions can subject individuals to frequent changes in velocity and direction [38].

2.2 Relevant Dimensionless Parameters

Several dimensionless parameters are relevant to models of spider ballooning and can be obtained through scale analysis, which describes fluid flow behavior. Scaling estimates the magnitudes of various forces acting on a body and governing fluid motion. Thus, scale analysis can be used to understand the meteorological state of the atmosphere that favors the initiation and maintenance of spider ballooning.

Initiation of the spider ballooning, or the take-off, could be aided by the presence of thermals. Thermals are parcels of hot air that develop as a result of convection when cold air sinks and hot air rises as shown in Fig. 3. The atmospheric conditions

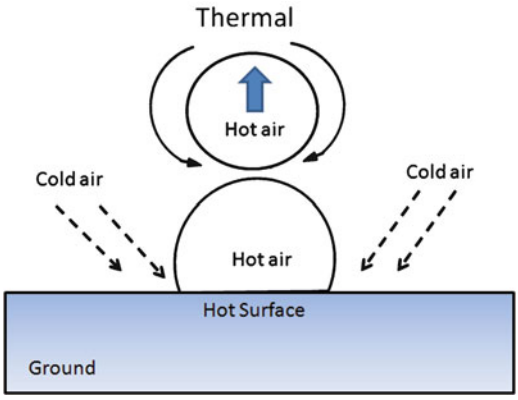


Fig. 3 The development of a thermal, with the rise of hot, lighter air initiating convection

that favor thermal formation can be established by examining the Richardson number, which characterizes the relationship between buoyancy and shear forces and is defined by

$$Ri = \frac{N^2}{\left(\frac{d\bar{U}}{dz}\right)^2} = -\left(\frac{g}{\rho_0}\right)\frac{\frac{d\bar{\rho}}{dz}}{\left(\frac{d\bar{U}}{dz}\right)^2}, \tag{1}$$

where g is the gravitational constant, $\bar{\rho}$ is the mean pressure, $\bar{U}$ is the mean horizontal wind velocity, z denotes the vertical coordinate direction, and N is the Brunt–Vaisala frequency or buoyancy frequency. The Richardson number quantifies the ratio of buoyancy-generated to shear-generated instability in the flow field [40].

Observational studies of aerial dispersal of spiders show a strong correlation between the number of spiders caught during dispersal and the Richardson number [42, 43]. The number of spiders caught in traps was significantly larger when the Richardson number was negative, which characterizes strong, unstable stratification and/or weak horizontal wind shear typical of calm, sunny days [21, 43]. This number could be used to describe the overall general conditions of take-off, though it does not relate to the properties of take-off or flight of the spider itself.

Another important dimensionless number relevant to spider ballooning is the Reynolds number. As an object moves through the air, aerodynamic forces between the object and the air are generated. The magnitude of forces depends on the shape, speed, and mass of the object, as well as properties of the fluid, specifically the viscosity, or stickiness, and the density. The Reynolds number relates the inertial forces to the viscous forces and is defined as

$$Re = \frac{\rho UL}{\mu}, \tag{2}$$

where ρ and μ are the density and the dynamics viscosity of the surrounding media (air), respectively, U is the characteristic velocity, and L is the characteristic length, which could be either the spider's body length or the length of the dragline [29]. The Reynolds number is used to characterize different flow regimes. $Re < 2000$ is typical for laminar flows, and $Re > 4000$ characterizes turbulent flows [29]. For spider ballooning, values of $Re_D < 10$ (where D is the spider's body length) indicate that the stimulus prompting the ballooning activity is too weak. In contrast, for values of $Re_D > 200$, it seems that the motion of the wind is too vigorous for the spiders to safely conduct aerial dispersal [27].

A third important dimensionless parameter used to analyze aerodynamic forces is the Froude number. Defined as the ratio of the inertial force to the gravitational force, it is used to quantify the effect of gravity on the moving object. The Froude number is described as $Fr = \frac{U}{\sqrt{gL}}$, where U is the wind velocity or the velocity of free fall, L is the characteristic length of the moving object, and, in the case of spiders, L is the diameter of the body. The Froude number has been used in studies of both terrestrial and aquatic locomotion to account for the effect of gravity on

movement of arthropods and mammals [4]. It has not, however, been used in relation to spider ballooning. One of the goals of our future research is to use this parameter to evaluate the significance of gravity on the dynamics of spider ballooning.

The last dimensionless number that is important for models of spider ballooning is the Strouhal number (St), which is defined as $St = \frac{fL}{U}$, where f is the frequency of vortex shedding (found numerically or experimentally), L is the length of the dragline, and U is the velocity of air relative to the velocity of the dragline (found as a sum of free fall velocity and wind speed). The Strouhal number is frequently used to describe the tail or wing kinematics of swimming or flying animals because it is known to govern a well-defined series of vortex growth and shedding regimes for airfoils undergoing pitching and heaving motions [25, 39]. Propulsion efficiency is high under a narrow range of Strouhal numbers and usually peaks within the interval $0.2 < St < 0.4$ [39]. The Strouhal number has not been used in analyses of spider ballooning dynamics. One of our future goals is to evaluate the Strouhal number for the passive locomotion of spider ballooning both numerically and experimentally.

3 Meteorological Conditions Favoring Spider Ballooning

Small, negative Richardson numbers are found on calm, sunny days when unstable stratification and/or weak horizontal wind shears dominate. Such meteorological conditions are known to be favorable for spider ballooning and occur close to the Earth's surface in the early morning or late afternoon when temperature convection, or temperature inversion, happens naturally in the atmosphere [1]. The vertical movement of warmer air up and cooler air down results in static instabilities in the atmosphere and leads to the formation of a vertical, turbulent layer filled with vortices, called thermals, that can be used as a lifting force by ballooning spiders.

Greenstone [22] focused on defining meteorological variables related to the production and maintenance of thermals and used them to predict the number of ballooning spiders. The author showed that clear sky meteorological conditions, which are necessary for thermal production, were present in only 82% of spider ballooning observations. This suggests that another source of uplift for ballooning that can generate vertical velocity may be present. A continuously stratified atmosphere can destabilize if the Richardson number is less than 0.25 (i.e., $0 < Ri < 0.25$). Kelvin–Helmholtz instability can take place as shown in Fig. 4 and vortices that grow in amplitude can lead to turbulence and vertical mixing as well [12]. However, a recent study showed that the timing of spider ballooning is not purely dominated by convection as some horizontal shear is also present [32].

The Kelvin–Helmholtz instability is only one particular type of instability that can affect spider ballooning. A zone with a wind shear, or a sudden change in the wind's speed or direction, is a well-known source of turbulence. The shearing, which can be horizontal, vertical, or both, creates forces that produce eddies along the mixing zone. The formed eddies may range in diameter from a couple of meters to several hundred meters [1]. Shear layers are a major cause of turbulence and, if they are formed high in the atmosphere, can often produce clear-air turbulence.

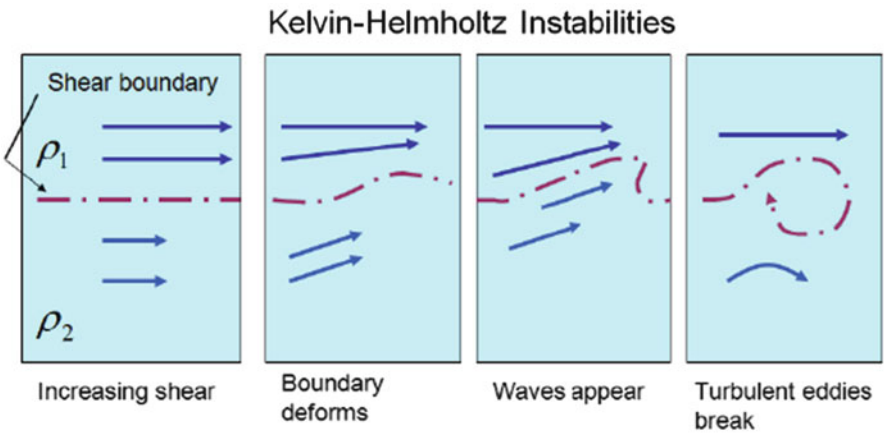


Fig. 4 Kelvin-Helmholtz shear instability resulting in the formation of turbulent eddies and mixing

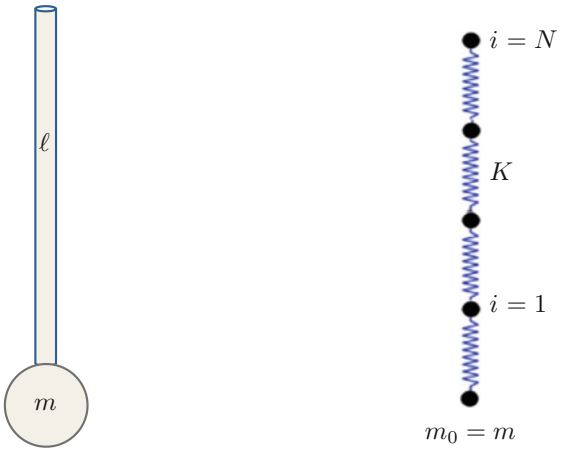
Shear layers can also be found near the Earth's surface, where turbulent whirling eddies are formed due to the roughness of the ground. As wind blows over a landscape full of obstacles such as trees and mountains, it breaks into irregular air motions known as wind gusts or turbulent eddies. The size and shape of these eddies depend on the dimensions of obstacles and the wind speed and can influence the air flow for hundreds of meters above the surface [1].

Another source of eddies are orographic perturbations, which occur when strong winds blow perpendicular to mountain ranges. Air flowing over the top of a mountain produces perturbations when it reaches the other side. The associated updrafts and downdrafts formed on the leeward side of the mountain may extend to heights from 2 to 20 times the height of the mountain peaks [1]. All of the meteorological scenarios summarized above may produce turbulence and/or eddies and may serve as a driving force for spider lift and subsequent dispersal.

4 Previous Mechanical Models for Ballooning

4.1 Humphrey's Take-Off Model

The first analytical model for spider ballooning was built by Humphrey [27] who studied the phenomenon with a simple fluid mechanical model (see Fig. 5a). In Humphrey's model, the spider was represented as a solid sphere. Attached to the solid sphere was a rigid, inextensible, cylindrical rod that was used to approximate a silk dragline, thus producing a "lollipop" appearance. The rod was considered to be massless relative to the spider.



(A) Humphrey's model (B) Reynolds et al.'s model

Fig. 5 Schematic of previous mechanical models of spider ballooning. In (**a**) Humphrey represented the spider as a solid sphere (m) attached to a rigid, inextensible, cylindrical, slender rod (ℓ) that approximates a dragline. The rod is considered to be massless relative to the sphere. In (**b**) Reynolds et al. represented the spider as a mass point m_0 attached to a flexible dragline made of a chain of N springs with spring constant K. The mass of the dragline is distributed on the nodes along the chain

For the lollipop model, the governing equation of the model was a balance of inertia, buoyancy, frictional and pressure forces:

$$\rho_s V_s \frac{d\mathbf{U}}{dt} = -\rho\chi V_s \frac{d(\mathbf{U}-\mathbf{V})}{dt} - \alpha|\mathbf{U}-\mathbf{V}|(\mathbf{U}-\mathbf{V}) + (\rho-\rho_s)V_s\mathbf{g} + \rho V_s \frac{d\mathbf{V}}{dt} + B(t), \tag{3}$$

where ρ_s and V_s are the sphere's density and volume, respectively, α and $\mathbf{U}$ are the sphere-rod drag coefficient and velocity, respectively, $\mathbf{V}$ is the fluid (air) velocity, ρ is the density of fluid (i.e., air), χ is the coefficient of added mass (0.5 for a sphere), $\mathbf{g}$ is the gravity vector, and $B(t)$ is the Basset term. Both added fluid mass and Basset force are due to the unsteadiness of the problem and the author ended up neglecting the Basset force in the simulation. Equation (3) was modified from previous research [6, 26] in which α was not specified for a sphere-rod system. A range of Reynolds numbers was defined based on wind velocity, spider length, and mass. For the drag coefficients for silk, slender bodies in low Reynolds number regimes were taken from Happel and Brenner [24]. The equations for the drag coefficient for the sphere were empirical correlations reported by Clift et al. [6]. However, the drag expression used to obtain the vertical take-off velocity of the spider was that of the turbulent regime:

$$mg = C_d\, d\, \ell\, \frac{1}{2}\,(V_z)^2_{\min}\,, \tag{4}$$

where m is the mass of the spider, g is the acceleration due to gravity, C_d is the filament drag coefficient, d is the filament diameter, ℓ is the filament length, and $(V_z)_{\min}$ is the minimum vertical component of air velocity required to initiate ballooning. In this model, drag determines environmental conditions necessary for take-off and is highly dependent on the dragline length. Numerical simulations were used to examine the spider-filament free fall in a quiescent environment and in sinusoidally oscillating vertical wind with average background wind vertical velocity of $\bar{V}_z = 0$. The minimum vertical wind required to initiate ballooning and the distance travelled over time were reported as relations of spider mass and dragline length.

Humphrey's results were supported by empirical investigations of drag on spiders and their silk by Suter [36–38]. However, the physical properties and dimensions in Humphrey's model were not validated. In particular, the numerical simulation of the system of Ordinary Differential Equations (Eq. (1.3) in Humphries 1987) for the simple lollipop model suggested that spiders with short silk draglines in strong winds travel faster and farther than the same sized spider with a long dragline in weaker winds. This contradicts the observation that spiders usually balloon in <3 m/s winds. Finally, Humphrey's lollipop model suggested that a group of small spiders on a silk dragline would travel faster and farther than a single spider of equivalent mass on the same dragline and in the same wind conditions.

4.2 Reynolds et al.'s Passive Dispersal Model

Instead of the rigid rod used in Humphrey's model [27], Reynolds et al. [31, 32] modeled the silk dragline as a chain of springs and spheres that resist stretching but not bending (Fig. 5b). Passive dispersal of the flexible silk filament was then modeled in turbulent flow that approximated the atmospheric boundary layer. Similar to Humphrey's work [27], take-off dynamics were not described. The Reynolds et al. model [31] described the force $\mathbf{F}_i$ acting on each node $i = 0, N$ as follows:

$$\mathbf{F}_i = m_i \tau_s^{-1}(\mathbf{u}_i - \mathbf{v}_i) + K(\mathbf{p}_{i,i-1} s_{i,i-1} + \mathbf{p}_{i,i+1} s_{i,i+1}) + m_0 \mathbf{g} \delta_{i,0}, \tag{5}$$

where $\mathbf{u}$ is the local air velocity, m_0 is the spider mass, $\mathbf{v}_0$ is the velocity located at node $i = 0$, m_i and $\mathbf{v}_i$ are mass and velocities at nodes $i = 1, \cdots, N$ along the chain, and K is the spring constant. $s_{i,j}$ are differences between the spring lengths to the fixed rest lengths from node i to node j. $\mathbf{p}_{i,i+1}$ are unit vectors orientated along the segment joining the i node and the $i + 1$ node. τ_s is the aerodynamic response time of the dragline, and $\delta_{i,0}$ is the Kronecker delta. This model reflects the assumption of massless spring and that gravity only acts on the node $i = 0$ (i.e., the body of the spider).

In a subsequent Reynolds et al. paper [32], the simplistic model of turbulence used in the first Reynolds et al. simulation [31] was extended to a Lagrangian

stochastic model. Lagrangian stochastic models for the velocity (**u**) and position (**x**) are in the following form

$$\frac{du_i}{dt} = a_i(\mathbf{x}, \mathbf{u}, t) + \sqrt{C_0 \epsilon}\frac{d\xi_i}{dt}, \qquad (6)$$
$$\frac{d\mathbf{x}}{dt} = \mathbf{u},$$

where the indices $i = 1, 2, 3$ denote Cartesian directions, $C_0 = 5$ is the Kolmogorov's Lagrangian velocity structure constant, ϵ is the mean rate of dissipation of turbulent kinetic energy divided by the density of air, ξ is the Wiener white noise, and a_i is a solution of the Fokker–Plank equation. This approach considered a simple model of turbulence.

Contrary to Humphrey's results [27], Reynolds found that the length of a dragline did not affect dispersal distance [31]. With the flexible dragline model, the condition for maximum dispersal distance was found to be a gradient Richardson number of approximately −3.2 [32]. This model showed that dispersal over several hundred kilometers was possible. However, based on the flexible dragline model, silk could become entangled (Fig. 6). Interestingly, experimental studies report that the silk dragline does not become entangled or clump in this manner once the spider is in flight [34].

4.3 Thomas et al.'s Diffusion Model

Instead of using a turbulence model that approximates the atmospheric boundary layer [32], Thomas et al. [41] built a one-dimensional diffusion model,

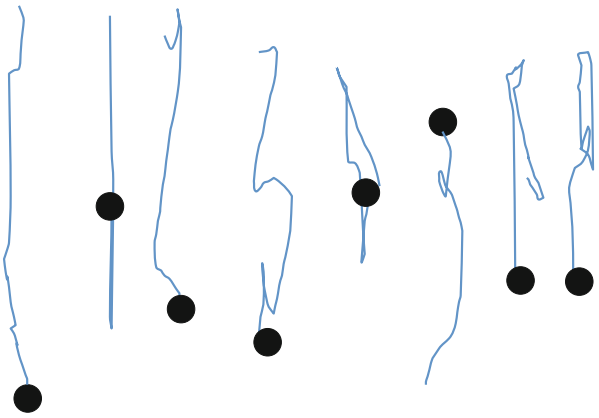


Fig. 6 The entanglement of the silk dragline predicted by Reynolds et al. model [31]. This figure was redrawn from [31] with permission

$$\frac{\partial n}{\partial t} = D_s \frac{\partial^2 n}{\partial x^2}, \tag{7}$$

to capture evolution of the spatial and temporal number of ballooning spiders $n(x, t)$. The solution to the diffusion Eq. (7) is [7]

$$n(x, t) = \frac{n_0}{2} \text{Erfc}\left(\frac{x}{2\sqrt{D_{s(t)}t}}\right), \tag{8}$$

where n_0 is the initial number of spiders at position x_0, n is the number of spiders at position x at time t, $D_{s(t)}$ is the diffusion coefficient of the species s at time t, and Erfc is the complementary error function. Based on the distribution of spiders in space and time, Thomas and colleagues estimated considerable horizontal distance traveled by the ballooning spider on the order of up to 90 km in 8h [42].

4.4 Models Incorporating Electrostatics

Electrical charging of the spider's dragline is believed to create electrostatic repulsion both within and among the silk filaments [28], allowing filaments to avoid the type of entanglement shown in Fig. 6. Researchers have suggested that, combined with aerodynamic force, the negative surface charge density of the Earth may play a role in spider ballooning. Though meteorological conditions are often believed to be the dominant driver in the take-off stage of spider ballooning, Gorham [19] conjectured that the electrostatic field is a necessary condition for take-off. In Gorham [19], an exponential approximate model for the atmospheric electric field is assumed to be

$$E(h) = E_0 e^{-\alpha h} \, (Vm^{-1}),$$

where h is height, α is the exponential rate for the exponential fit to the electric field (i.e., $\alpha = 3.0 \times 10^{-4}\,\text{m}^{-1}$), and E_0 is the electric field at h = 0 (i.e., $E_0 = -120\,\text{Vm}^{-1}$). Assuming a constant acceleration for a single dragline in a pure electrostatic field, the charge required for take-off is computed as

$$Q_{accel} = m(a_{net} + g)/E_0 \approx 100\,(nC), \tag{9}$$

where m is the mass of the spider and an initial net vertical acceleration is in the range of $a_{net} = 3 \sim 6\,\text{ms}^{-2}$. The author concluded that spiders use electrostatic forces only to lift them into the air [19]. Electrostatics may account for some unexplained ballooning phenomena, such as high velocity ballooning in conditions of little or no wind or the observation made by Charles Darwin that spiders were ballooning from his ship, The Beagle, in horizontal movements [8, 19]. However, the

role of electrostatics may not be significant since empirical research has shown that spiders are using rising thermals on hot days without wind [34] and, thus, thermal currents could provide all the necessary lift for ballooning.

5 Challenges, Open Questions, and Needs

Describing the relevant physics for models of spiders that are ballooning across large distances in turbulent flows presents significant challenges. The material properties of spider silk are complex and can vary among species, individuals, and even within the same organism [17, 20]. The interaction between the flexible dragline and air must be taken into consideration, and resolving this fluid–structure interaction can be particularly difficult.

Spider ballooning is also an inherently multiscale problem. When zooming into the flow around an individual spider, turbulence can be handled explicitly using direct numerical simulation. When considering dozens of organisms within large-scale geophysical flows, it is necessary to use turbulence models such as Reynolds averaged Navier–Stokes equations (RANS) [18, 47] or large eddy simulations [9, 33, 35]. Both methods average out either the entire or the sub-grid flow structure for computational efficiency. However, the details of the flow near the spider and the dragline are critical to resolving the silk's reconfiguration. The development of new multiscale methods, such as hybrid direct numerical simulation / RANS models, may be required to resolve both the viscous flow near the dragline and the atmospheric turbulent flows.

In addition to resolving the flow near the elastic dragline, there are challenges to modeling the atmospheric processes relevant to spider ballooning. The vertical ascent and descent of the spider is likely sensitive to short-term turbulent variability. Mathematical models that average velocities across relatively large distances and over long time intervals may not be appropriate since these simplifications can effectively smooth strong nonlinearities relevant to take-off and flight. On the other hand, resolving such details can quickly make numerical simulations that describe long distance ballooning intractable. As an example, for some ballooning events, it may be desirable to simulate a meteorological pressure system for which the relevant length scale of the pressure system would be on the order of tens to hundreds of kilometers. In contrast, the relevant length scale of the smallest vortices or eddies may be on the order of one millimeter and the relevant length scale of the spider itself may be on the order of hundreds of microns.

In terms of temporal resolution, ballooning events may take place over the time scale of days, while transient flows that affect ballooning may occur on the scale of milliseconds. Brute force simulation of such a problem would require on the order of 10^{24} spatial grid points and 10^{8} time steps. Highly efficient, adaptive, and parallelized algorithms that can refine spatial and temporal scales are required to handle this challenge. Scaling laws evaluating the relevant and dominant features required for such modeling would also be critical.

There are a variety of ways in which detailed fluid–structure interaction modeling could help us understand the dynamics of the silk dragline and the ballooning behavior of spiders. Predictive multiscale and multiphysics models could allow us to determine if spiders prefer to take off in certain conditions such as updrafts or within rolling eddies. These models may also help us understand other aspects of ballooning such as turbulence sensing during the tiptoeing behavior and how flow structures may be utilized to enable very long distance dispersal.

In addition to improved numerical methods and mathematical models describing spider ballooning, additional experimental data are needed to guide and validate such models. Wind tunnel studies could be used to examine not only the desired wind speed for takeoff but also the desired wind direction and/or turbulent profile. Spatial and temporal resolution of dispersal patterns could be used to validate models that couple the elastic dynamics of the dragline with complicated flow structures. Visualization of the dynamics of the silk dragline in the air could aid in the selection of the appropriate dragline model.

By combining modeling and experimental approaches, a host of possible questions could be explored, including:

- Is elasticity of the dragline (both in terms of resistance to bending and stretching) important for take-off and flight in turbulent regime?
- Is there an optimal dragline length or wind flow profile for take-off, flight, and settling?
- Why do we see such a narrow range of body sizes in ballooning spiders?
- Why is there a maximum wind speed for take-off?

Addressing such questions would help shed light on an interesting ecological mystery surrounding dispersal that has implications for species distributions, invasion biology, and impacts of climate change.

Acknowledgements This work was conducted as part of the Aerodynamics of Spider Ballooning Working Group at the National Institute for Mathematical and Biological Synthesis (NIMBioS), sponsored by the National Science Foundation (NSF) award DBI-1300426 with additional support from The University of Tennessee, Knoxville. We are grateful to Dr. Anita Layton for organizing the NIMBioS workshop. Additional funding was provided by NSF to KSS (Postdoctoral Research Fellowship 1306883), LAM (CBET 1511427), and AC (Graduate Research Fellowship 201315897), and by the Reed and Edgerton Funds at MIT to LB.

References

1. Ahrens, D.: Meteorology Today. Brooks/Cole Press, Pacific Grove, CA (2003)
2. Barth, F.G., Komarek, S., Humprhey, J.A.C., Treidler, B.: Drop and swing dispersal behavior of a tropical wandering spider – experiments and numerical-model. J. Comp. Physiol. A Sens. Neural Behav. Physiol. **169**(3), 313–322 (1991)
3. Bishop, L.: Meteorological aspects of spider ballooning. Environ. Entomol. **19**(5), 1381–1387 (1990)
4. Bush, J.W.M., Hu, D.L.: Walking on water: biolocomotion at the interface. Annu. Rev. Fluid Mech. **38**(1), 339–369 (2006)

5. Cantabrigian, M.L.: Some observations concerning the odd turn of some shell snails and the darting of spiders. Philos. Trans. R. Soc. Lond. **4** 1011–1016 (1669)
6. Clift, R., Grace, J.R., Weber, M.E.: Bubbles, Drops, and Particles. Academic, London (1978)
7. Crank, J.: The Mathematics of Diffusion. Oxford University Press, Oxford (1975)
8. Darwin, C.R.: Journal of Researches into the Natural History and Geology of the Countries Visited During the Voyage of H.M.S. Beagle Round the World, Under the Command of Capt. Fitz Roy, R.N., 2nd edn. John Murray, London (1845)
9. Deardorff, J.W.: A numerical study of three-dimensional turbulent channel flow at large reynolds numbers. J. Fluid Mech. **41**, 453–480 (1970)
10. Dingle, H.: Migration: The Biology of Life on the Move: The Biology of Life on the Move. Oxford University Press, Oxford (1996)
11. Dobson, S.F.: Competition for mates and predominant juvenile male dispersal in mammals. Anim. Behav. **30**(4), 1183–1192 (1982)
12. Drazin, P., Reid, W.: Hydrodynamic Stability. Cambridge University Press, Cambridge (1981)
13. Duffey, E.: Aerial dispersal in a known spider population. J. Anim. Ecol. **25**(1), 85–111 (1956)
14. Eberhard, W.G.: How spiders initiate airborne lines. J. Arachnol. **15**, 1–9 (1987)
15. Foelix, R.F.: Biology of Spiders. Harvard University Press, Cambridge, MA (1982)
16. Foelix, R.F.: Biology of Spiders, 3rd edn. Oxford University Press, Oxford (2011)
17. Gatesy, J., Hayashi, C., Motriuk, D., Woods, J., Lewis, R.: Extreme diversity, conservation, and convergence of spider silk fibroin sequences. Science **291**(5513), 2603–2605 (2001)
18. Gatski, T.B., Bonnet, J.-P.: Compressibility, Turbulence and High Speed Flow. Elsevier, Amsterdam (2009)
19. Gorham, P.W.: Ballooning Spiders: The Case for Electrostatic Flight. ArXiv:1309.4731v1 [physics.bio–ph] (2013)
20. Gosline, J.M., Guerette, P.A., Ortlepp, C.S., Savage, K.N.: The mechanical design of spider silks: from fibroin sequence to mechanical function. J. Exp. Biol. **202**(23), 3295–3303 (1999)
21. Greenstone, M.H.: Ballooning frequency and habitat predictability in two wolf spider species (lycosidae: Pardosa). Fla. Entomol. **65**(1), 83–89 (1982)
22. Greenstone, M.H.: Meteorological determinants of spider ballooning: the roles of thermals vs. the vertical windspeed gradient in becoming airborne. Oecologia **84**(2), 164–168 (1990)
23. Gressitt, J.L.: Biogeography and ecology of land arthropods of Antarctica. In: Van Miegham, J., Van Oye, P. (eds.) Biogeography and Ecology of Antarctica. Monographiae Biologicae, vol. 15, pp. 431–490. Springer, Dordrecht (1965)
24. Happel, J., Brenner, H.: Low Reynolds Number Hydrodynamics: with Special Applications to Particulate Media. M. Nijhoff, The Hague/Boston/Hingham (1983)
25. Hedenström, A.: A general law for animal locomotion? Trends Ecol. Evol. **19**(5), 217–219 (2004)
26. Houghton, G.: The behaviour of particles in a sinusoidal velocity field. Proc. R. Soc. A: Math. Phys. Eng. Sci. **272**(1348), 33–43 (1963)
27. Humphrey, J.A.C.: Fluid mechanic constraints on spider ballooning. Oecologia **73**, 469–477 (1987)
28. Kronenberger, K., Vollrath, F.: Spiders spinning electrically charged nano-fibres. Biol. Lett. **11**(1), 20140813 (2015)
29. Oertel, H.: Prandtl's Essentials of Fluid Mechanics. Springer, New York (2004)
30. Pusey, A.E.: Sex-biased dispersal and inbreeding avoidance in birds and mammals. Trends Ecol. Evol. **2**(10), 295–299 (1987)
31. Reynolds, A.M., Bohan, D.A., Bell, J.R.: Ballooning dispersal in arthropod taxa with convergent behaviours: dynamic properties of ballooning silk in turbulent flows. Biol. Lett. **2**(3), 371–373 (2006)
32. Reynolds, A.M., Bohan, D.A., Bell, J.R.: Ballooning dispersal in arthropod taxa: conditions at take-off. Biol. Lett. **3**(3), 237–240 (2007)
33. Sagaut, P.: Large Eddy Simulation for Incompressible Flows, 3rd edn. Springer, London (2006)
34. Schneider, J.M., Roos, J., Lubin, Y., Henschel, J.R.: Dispersal of Stegodyphus dumicola (Araneae, Eresidae): they do balloon after all! J. Arachnol. **29**(1), 114–116 (2001)

35. Smagorinsky, J.: General circulation experiments with the primitive equations. Mon. Weather Rev. **91**(3), 99–164 (1963)
36. Suter, R.B.: Ballooning in spiders: results of wind tunnel experiments. Ethol. Ecol. Evol. **3**(1), 13–25 (1991)
37. Suter, R.B.: Ballooning: data from spiders in freefall indicate the importance of posture. J. Arachnol. **20**, 107–113 (1992)
38. Suter, R.B.: An aerial lottery: the physics of ballooning in a chaotic atmosphere. J. Arachnol. **27**, 281–293 (1999)
39. Taylor, G.K., Nudds, R.L., Thomas, A.L.R.: Flying and swimming animals cruise at a Strouhal number tuned for high power efficiency. Nature **425**(6959), 707–711 (2003)
40. Tennekes, H., Lumley, J.L.: A First Course in Turbulence. MIT Press, Cambridge (1972)
41. Thomas, C.F.G., Hol, E.H.A., Everts, J.W.: Modelling the diffusion component of dispersal during recovery of a population of linyphiid spiders from exposure to an insecticide. Funct. Ecol. **4**(3), 357–368 (1990)
42. Thomas, C.F.G., Brain, P., Jepson, P.C.: Aerial activity of linyphiid spiders: modelling dispersal distances from meteorology and behaviour. J. Appl. Ecol. **40**(5), 912–927 (2003)
43. Vugts, H.F., Van Wingerden, W.K.R.E.: Meteorological aspects of aeronautic behaviour of spiders. Oikos **27**(3), 433–444 (1976)
44. Weyman, G.S.: A review of the possible causative factors and significance of ballooning in spiders. Ethol. Ecol. Evol. **5**, 279–291 (1993)
45. Weyman, G.S.: Laboratory studies of the factors stimulating ballooning behavior by linyphiid spiders (Araneae, Linyphiidae). J. Arachnol. **23**(25), 75–84 (1995)
46. Wickler, W., Seibt, U.: Aerial dispersal by ballooning in adult Stegodyphus mimosarum. Naturwissenschaften **73**(10), 628–629 (1986)
47. Wilcox, D.C.: Turbulence Modeling for CFD. DCW Industries, La Canada, CA (2006)
48. Wolff, J.O.: What is the role of adults in mammalian juvenile dispersal? Oikos **68**(1), 173–176 (1993)

Flying Spiders: Simulating and Modeling the Dynamics of Ballooning

Longhua Zhao, Iordanka N. Panayotova, Angela Chuang, Kimberly S. Sheldon, Lydia Bourouiba, and Laura A. Miller

Abstract Spiders use a type of aerial dispersal called "ballooning" to move from one location to another. In order to balloon, a spider releases a silk dragline from its spinnerets and when the movement of air relative to the dragline generates enough force, the spider takes flight. We have developed and implemented a model for spider ballooning to identify the crucial physical phenomena driving this unique mode of dispersal. Mathematically, the model is described as a fully coupled fluid–structure interaction problem of a flexible dragline moving through a viscous, incompressible fluid. The immersed boundary method has been used to solve this complex multi-scale problem. Specifically, we used an adaptive and distributed-memory parallel implementation of immersed boundary method (IBAMR). Based on the nondimensional numbers characterizing the surrounding flow, we represent the spider as a point mass attached to a massless, flexible dragline. In this paper, we explored three critical stages for ballooning, takeoff, flight, and settling in two dimensions. To explore flight and settling, we numerically simulate the spider in free fall in a quiescent flow. To model takeoff, we initially tether the spider-dragline

L. Zhao (✉)
Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
e-mail: longhua.zhao@case.edu

I.N. Panayotova
Department of Mathematics, Christopher Newport University, Newport News, VA 23606, USA
e-mail: iordanka.panayotova@cnu.edu

A. Chuang • K.S. Sheldon
Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA
e-mail: angelachuang@utk.edu; ksheldon@utk.edu

L. Bourouiba
The Fluid Dynamics of Disease Transmission Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02130, USA
e-mail: lbouro@mit.edu

L.A. Miller
Departments of Mathematics and Biology, University of North Carolina, Chapel Hill, NC 27599, USA
e-mail: lam9@email.unc.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_10

system and then release it in two types of flows. Based on our simulations, we can conclude that the dynamics of ballooning is significantly influenced by the spider mass and the length of the dragline. Dragline properties such as the bending modulus also play important roles. While the spider-dragline is in flight, the instability of the atmosphere allows the spider to remain airborne for long periods of time. In other words, large dispersal distances are possible with appropriate wind conditions.

1 Introduction

Dispersal is the nonreturning movement of organisms away from their birth sites [25], often triggered by density and habitat-dependent factors [5, 9]. These factors play a role in the initiation and frequency of activities such as foraging, choosing nest sites, searching for mates, and avoiding predation, competition, and inbreeding [6]. Dispersal traits and mechanisms are thus wide and varied across taxa, even for organisms that disperse passively through air or water currents [19].

Spiders (Arachnida: Araneae) represent one taxon that undergoes a specialized form of passive dispersal. Besides walking from site to site, most spiders also engage in a type of aerial dispersal known as ballooning [2]. This begins with a distinctive "tiptoe" behavior where an individual straightens its legs, balancing on the tips of its tarsi. After tiptoeing, the spider raises its abdomen, releasing a silken dragline in the air (Fig. 1). Wind then allows for drag-induced lift of the whole body. Once airborne, individuals have little control over the direction and distance of displacement; rather, they join other floating life forms collectively known as "aerial plankton" [10], which are subject to air currents.

Spiders have long been observed to balloon to distances as far as 3200 km [12] and heights of up to 5 km [10]. The extreme heights and distances achieved from a seemingly simple mechanism have generated interest in the flight physics of these arachnid aeronauts. This intriguing behavior is apparently constrained by body mass (<100 mg) and wind speed (<3 m/s). The complex interactions of the physical characteristics of the spider's morphology, silk dragline properties, and meteorological conditions have also motivated the identification of the dominant regimes during takeoff, flight, and settling. Since ballooning spiders are very small and cannot be easily tracked, conventional measures of dispersal are difficult. This has motivated theoretical work in determining the physics and resulting distributions of ballooning spiders.

Early models of spider ballooning primarily focused on the factors that are important for flight as well as the distances that can be achieved (see [27] for a review of previous models). Humphrey [15] developed the first simple force balance model where the physical properties of the spider and its attached dragline were simplified to a massive sphere (spider) and a massless, rigid rod (dragline). Described as a "lollipop system," this model evaluates the possible relationships between spider mass, dragline length, and dispersal distance during initial takeoff. The results were used to define a region of physical parameters that mechanically support ballooning based on wind velocity and spider mass, although dragline length also played a role in travel speed and distance [15].

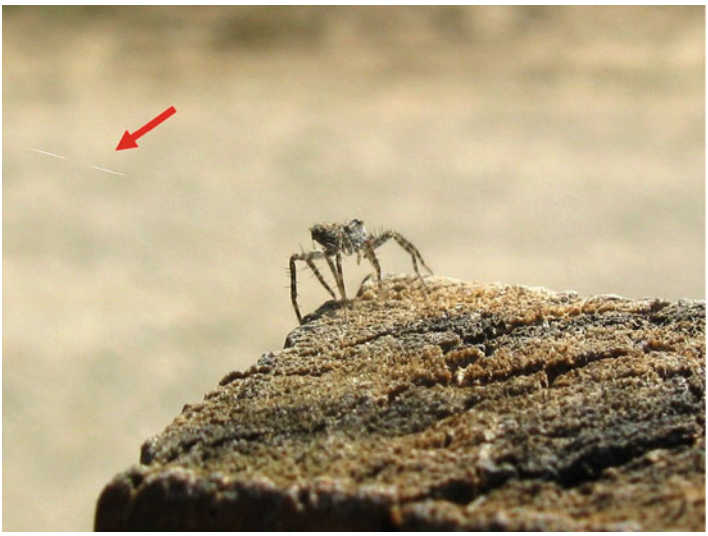


Fig. 1 Tiptoe behavior in which a spider stands on tarsi, raises the abdomen, and releases a dragline (*indicated by arrow*) in order to initiate ballooning. Photo copyright belongs to Sarah Rose

Subsequent models were built on this system by considering the dragline as a series of spheres and springs. This approach allowed more realistic properties to be included in the spider-dragline system such as flexibility and extensibility of the dragline [22]. Using a set of Lagrangian stochastic models to capture the turbulence in the air, simulations of ballooning were able to predict reasonable dispersal distances in the presence and absence of wind shear conditions [23].

Statistical approaches, empirical measurements, and simulations have also furthered our understanding of ballooning dynamics. Suter [30] measured spiders in free fall in statistical models that related the body mass and dragline lengths to their terminal velocities. The potential importance of body posturing was also noted, as it could account for deviations from the expected values of terminal velocity. In other words, spiders can possibly posture their legs and body in a way that impacts their fallout [31]. However, it is unlikely that their decision to balloon is based on accurate meteorological predictions, as shown in models that relate their probability of dispersal with mass, silk length, and local wind velocity variation. Thomas et al. used numerical simulations to understand the temporal and spatial dynamics through diffusion models [32]. These were subsequently used to understand feasible dispersal distances under a simple atmospheric model [33].

These earlier models illustrate the various methods that have been utilized to understand different aspects of the spider ballooning process but many simplifying assumptions are made regarding the fluid–structure interaction. In this study, we investigate the dominant physical regimes of passive aerial dispersal in spiders, with a particular focus on the fluid dynamics of their flight. We consider the physical

parameter space that influences all stages of ballooning, including takeoff, transport, and landing. We use a numerical approach to model the complex interaction of the coupled spider-dragline system and its movement through various air-flow conditions. Like earlier studies, this model includes the spider body mass and a flexible dragline. We go beyond previous work by resolving the full aeroelasticity problem of a flexible dragline moving through a viscous fluid. We also directly simulate a variety of background air-flow profiles.

In the next section, we describe the numerical method for solving the fully coupled fluid–structure interaction problem of a flexible dragline immersed in a viscous fluid. In Sect. 2.2, we discuss our model of the spider–flow interaction and then consider the validity of our model in various scenarios: free fall in quiescent flow and nonquiescent flows. We then discuss the dynamics of ballooning in Sect. 3 in various flow conditions. Lastly, we summarize our findings in Sect. 4.

2 Methods

2.1 Immersed Boundary Method

Our goal is to mathematically model a flexible dragline that is both deformed by the air and also moves the air. In other words, we wish to consider the fully coupled fluid–structure interaction. We used the immersed boundary method to model this fully coupled fluid–structure interaction problem [18, 20, 21]. After over 30 years of application to problems in biological fluid dynamics, the immersed boundary (IB) method represents a relatively straightforward and standard approach for studying problems in animal locomotion including insect flight [17], lamprey swimming [34], and jellyfish swimming [14].

The basic idea behind the immersed boundary method is that the equations of fluid motion are solved on a (typically Cartesian) grid using an Eulerian frame of reference. The equations describing the immersed elastic boundary are solved on a curvilinear mesh defined using a Lagrangian frame of reference. The collection of Lagrangian nodes on which the equations describing the immersed elastic boundary are solved move independently of the fluid grid. The immersed boundary is moved at the local fluid velocity, and the elastic forces are spread to the fluid through regularized discrete delta functions.

The following equations describing the immersed boundary method are given in two dimensions, but the extension to three dimensions is mathematically straightforward, though efficient implementation in three dimensions is challenging. More

details may be found in Peskin [21]. The Navier–Stokes equations are used to describe a viscous incompressible fluid (such as air at low *Re*) as follows:

$$\rho(\mathbf{u}_t(\mathbf{x},t) + \mathbf{u}(\mathbf{x},t)\cdot\nabla\mathbf{u}(\mathbf{x},t)) = \nabla p(\mathbf{x},t) + \mu\nabla^2\mathbf{u}(\mathbf{x},t) + \mathbf{F}(\mathbf{x},t), \tag{1}$$

$$\nabla\cdot\mathbf{u}(\mathbf{x},t) = 0, \tag{2}$$

where $\mathbf{u}(\mathbf{x},t)$ is the fluid velocity, $p(\mathbf{x},t)$ is the pressure, $\mathbf{F}(\mathbf{x},t)$ is the force per unit area applied to the fluid by the immersed boundary, ρ is the density of the fluid, and μ is the dynamic viscosity of the fluid. The independent variables are the time t and the position $\mathbf{x}$.

The interaction equations between the fluid and the boundary are given by the following integral transforms with delta function kernels:

$$\mathbf{F}(\mathbf{x},t) = \int \mathbf{f}(r,t)\delta\left(\mathbf{x} - \mathbf{X}(r,t)\right) dr, \tag{3}$$

$$\mathbf{X}_t(r,t) = \mathbf{U}(\mathbf{X}(r,t)) = \int \mathbf{u}(\mathbf{x},t)\delta\left(\mathbf{x} - \mathbf{X}(r,t)\right) d\mathbf{x}, \tag{4}$$

where $\mathbf{f}(r,t)$ is the force per unit length applied by the boundary to the fluid as a function of Lagrangian position and time, $\delta(\mathbf{x})$ is a delta function, $\mathbf{X}(r,t)$ gives the Cartesian coordinates at time t of the material point labeled by the Lagrangian parameter r. Equation (3) applies force from the boundary to the fluid grid, and Eq. (4) evaluates the local fluid velocity at the boundary.

In order to tether the boundary points to a fixed location, a penalty force is applied that is proportional to the distance between the boundary and the desired location of target points. This force is given by:

$$\mathbf{f}(r,t) = \kappa_{targ}\left(\mathbf{Y}(r,t) - \mathbf{X}(r,t)\right), \tag{5}$$

where $\mathbf{f}(r,t)$ is the force per unit length, κ_{targ} is a stiffness coefficient, and $\mathbf{Y}(r,t)$ is the prescribed position of the target boundary. The deviations from the target position can be controlled by the parameter κ_{targ}.

The flexible dragline used in the following simulations resists stretching and bending. To model the resistance to stretching, we insert elastic links connecting adjacent boundary points that act as linear springs. Let boundary points m and n have the corresponding position coordinates $\mathbf{X}_m$ and $\mathbf{X}_n$, and let these points be connected by elastic link w. The stretching energy function for this link is then given by:

$$E_S(\mathbf{X}_m,\mathbf{X}_n) = \frac{1}{2}\kappa_s(||\mathbf{X}_m - \mathbf{X}_n|| - l_w)^2, \tag{6}$$

where l_w is the resting length of the spring and κ_s is its stiffness coefficient. Note that E_S is equal to zero when the distance between the points equals the resting length.

The dragline also has a small resistance to bending. We assume zero preferred curvature (the dragline wants to be straight). The bending energy is then given by:

$$E_b = \frac{1}{2}\kappa_b\int\left|\frac{\partial\mathbf{X}(s,t)}{\partial s}\right|^2 ds, \tag{7}$$

where κ_b is the bending stiffness. We discretize the bending energy with zero preferred curvature as follows:

$$E_b = \frac{1}{2}\kappa_b \sum_i |D_s D_s \mathbf{X}|^2 \Delta s = \frac{1}{2}\kappa_b \sum_{i=2}^{N-1} \frac{|\mathbf{X}_{i+1} - 2\mathbf{X}_i + \mathbf{X}_{i-1}|^2}{\Delta s^2} \Delta s, \tag{8}$$

The total elastic energy is calculated as the sum of the stretching and bending energies for each immersed boundary point. For example, a dragline is made up of a string of N immersed boundary points arranged in the order so that each pair of consecutive points is joined by a linear spring that resists stretching and each consecutive triplet resists bending. This results in the following equation for the total elastic energy:

$$E(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_N, t) = \sum_{i=1}^{N-1} E_S(\mathbf{X}_i, \mathbf{X}_{i+1}) + \sum_{i=2}^{N-1} E_B(\mathbf{X}_{i-1}, \mathbf{X}_i, \mathbf{X}_{i+1}). \tag{9}$$

The elastic force at point m is then calculated using the derivatives of the elastic energy as follows:

$$\mathbf{F}_m(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_N, t) = -\frac{\partial \mathbf{E}(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_N, t)}{\partial \mathbf{X}_m}. \tag{10}$$

Values of the constants κ_s and κ_b must be chosen to specify reasonable energies and forces associated with the dragline and are selected to be within the range of what is observed for spiders. Mass was added to the spider using the penalty immersed boundary method [16]. The boundary points that are assigned a mass are anchored with linear springs to "ghost" massive particles. The linear springs have zero resting lengths, and the spring stiffness coefficients are chosen such that the boundary point moves with the massive particle within some tolerance. The massive particles do not interact with the fluid (the boundary points that they are connected to do) and simply move according to Newton's laws. With a stiffness spring connecting the mass point and the boundary point, an energetic penalty is imposed when the position of the Lagrangian immersed boundary point deviates from that of the mass. Similar to Eq. (5), the energetic penalty is introduced into the system by a large value of the penalty stiffness κ_s between the point of spider and the point of dragline it is attached to.

To perform direct numerical simulations, we used an adaptive and parallelized version of the immersed boundary method, IBAMR [13]. IBAMR is a C++ framework that provides discretization and solver infrastructure for PDEs on block-structured locally refined Eulerian grids [3, 4] and on Lagrangian (structural) meshes, as well as infrastructure for coupling Eulerian and Lagrangian representations. The adaptive method used four grid levels to discretize the Eulerian equations with a refinement ratio of four between levels. Regions of fluid that contained the immersed boundary or vorticity magnitude above $0.125\,\mathrm{s}^{-1}$ were discretized at the highest refinement. The effective resolution of the finest level of the grid corresponded to that of a uniform 512^2 discretization.

2.2 Spider Model

In previous mechanical models [15, 22], the spider body was modeled as a sphere (see [27] for a review of previous models). However, the detailed aerodynamics of the viscous fluid interacting with the spider-dragline system were not resolved. Based on an analysis of the relevant dimensionless numbers, which are outlined below in Sect. 2.2.2, we neglect the drag acting on the spider itself and focus on the dragline. We do consider the mass of the spider which is represented as a point mass tethered to the dragline. The dragline is modeled as a massless beam that resists bending and stretching. The governing equations are similar to Eqs. (1)–(2) as following:

$$\rho(\mathbf{u}_t(\mathbf{x},t) + \mathbf{u}(\mathbf{x},t)\cdot\nabla\mathbf{u}(\mathbf{x},t)) = \nabla p(\mathbf{x},t) + \mu\nabla^2\mathbf{u}(\mathbf{x},t) + mg + \mathbf{F}(\mathbf{x},t), \quad (11)$$

$$\nabla\cdot\mathbf{u}(\mathbf{x},t) = 0, \quad (12)$$

where mg is the gravity force due to the point mass of the spider and $\mathbf{F}$ is the force that the dragline applies to the fluid.

For the numerical discretization of the elastic dragline, the dragline is represented as discrete Lagrangian points connected by springs that resist bending and stretching with stiffnesses κ_s and κ_b, respectively. Note that the relevant elasticity Eqs. (5)–(7) represent a very different system from the chain of springs in Reynolds et al.'s model [22]. In Reynolds et al.'s model, the dragline is defined by spring modulus K only, i.e., κ_s in our model. Their dragline can freely bend in any direction, which may result in unrealistic entanglement. In our model, the bending modulus limits the bending of the dragline. Our model may still, however, result in some entanglements due to the fluid–structure interaction. Note that we do not include electrostatic forces in our model which may further limit the degree of entanglement.

The spider-dragline system is then immersed in air with appropriate boundary conditions for different scenarios (e.g., no slip for settling in a quiescent fluid, Dirichlet for prescribed background flow, and mixed for cavity flow). In this fluid–structure interaction system, the flow field is obtained by numerically solving the full Navier–Stokes Eqs. (11)–(12). The spider-dragline is moved at the local fluid velocity (4).

2.2.1 Numerical and Physical Parameters Used for Simulation

Due to the computational challenges associated with immersed boundary simulations in three dimensions, we consider only a two-dimensional representation of the spider-dragline system in this initial study. Note that in two dimensions, the spider is actually a sheet, and the point mass representing the spider is with units of mass per length (M/L) converted from the three-dimensional mass. As a rough approximation of the relationship between the actual mass of a real spider and the two-dimensional idealization, one could divide the mass of a spider by its diameter to obtain the mass per unit length used in the simulations.

Parameters used for the simulation are summarized in Table 1.

Table 1 Parameters used in the numerical simulations

Physical parameters	Values in literature	Values in simulation
Elasticity (Spring modulus)	n/a	20 (N/m)
Bending modulus	n/a	$10^{-5} \sim 5 \times 10^{-4}$ (N·m)
Dragline length [2, 22]	$0 \sim 2.3$ (m)	$0.05 \sim 0.2$ (m)
Dragline diameter [26]	20–100 (nm)	Line
Dragline density [29]	$1.1 \sim 1.4$ g/cm^3	Massless
Spider diameter [8, 24]	$1 \sim 5$ (mm)	Pointwise
Spider mass [8, 30]	$0.09 \sim 84.70$ (mg)	$2 \sim 800$ (mg/m)
Air mass density [1]	1.165 (kg/m^3)	1.177 (kg/m^3)
Air dynamic viscosity [1]	1.86×10^{-5} (N·s/m^2)	1.846×10^{-5} (N·s/m^2)

Note the difference in the units of stiffness and mass since the simulations are in two dimensions rather than three dimensions for actual spiders. The physical properties of the air are at temperature 30 °C [1]

2.2.2 Dimensionless Parameters

Dimensionless parameters are important to characterize the properties of the fluid and its interaction with the organism. The first dimensionless parameter we consider is the Reynolds number (Re), which is computed as the ratio of inertial forces over viscous forces. Re is given as $\frac{\rho LU}{\mu}$, where ρ is the density of the fluid, μ is the dynamic viscosity of the fluid, L is a characteristic length that is chosen based on the application, and U is a characteristic velocity. Re is often used to characterize different flow regimes. When Re is low ($Re << 10^3$), the flow is in the laminar regime. When $Re << 1$, viscous forces are dominant, the flow is reversible, and the fluid motion is smooth. For $Re >> 1$, the flow is dominated by inertial forces. Flows at $Re > 2300$ (for the case of pipe flow) are typically, but not necessarily, turbulent and tend to produce chaotic eddies, vortices, and other flow instabilities.

There are several ways that one can choose the characteristic length for the calculation of Re. In Humphrey's model [15], the Reynolds number is defined as $Re_D = \frac{\rho D|\mathbf{V}|}{\mu}$, where D is the diameter of spider and $|\mathbf{V}|$ is the modulus of the wind velocity. In our simulations, we choose U as the velocity of the spider relative to the air. The characteristic length L could be chosen as the dragline length ℓ (Re_ℓ), the radius of the dragline d (Re_d), or the spider body diameter D (Re_D), respectively. Keeping the same characteristic velocity, Re varies with ratios from 1 to 10^4 for different choices of characteristic length L, using the radius of the dragline d and the length of dragline ℓ.

Another important dimensionless parameter is the Richardson number Ri. It is defined as the ratio of density gradient over the flow gradient. Ri is used as the threshold parameter for convective instability, which is an environment factor that may be important in the decision to balloon. Thomas et al. [33] reported that the number of airborne spiders was significantly correlated with Richardson numbers. In our study, we explore the dynamics of airborne spiders and neglect the influence of temperature. Winds are specified as the boundary and initial conditions. Besides

temperature, we also neglect the effect of electrostatics in the model. Because electrostatic forces can prevent sticking, coiling, and entanglement of the dragline, and because Gorham [11] reported that the effects of electrostatic forces could be substantial for distances traveled, we plan to include electrostatic forces in our future work.

The last dimensionless number considered is the Strouhal number (**St**), defined as:

$$\mathbf{St} = \frac{L}{U\tau}.$$

Here, τ is the relaxation characteristic time scale or the inverse of disturbance frequency f. **St** represents a measure that relates oscillation frequency to fluid velocity. For the case of spider ballooning, the oscillations in fluid velocity are due to alternate vortex shedding from the end of the dragline. Note that for $\mathbf{St} << 1$, oscillations of the fluid have a minimal impact on the dynamics. At intermediate Strouhal numbers $0.1 < \mathbf{St} < 1$, oscillation is characterized by the buildup and rapidly subsequent shedding of vortices [28]. Such vortex shedding could be important to spider ballooning since large forces are generated during vortex separation. Such peaks in force may impact takeoff and flight trajectories.

2.2.3 Boundary and Flow Conditions

The background flows are driven in the simulations using Dirichlet boundary conditions. In the quiescent fluid simulations, where we study the free fall of spiders, zero initial and boundary velocities are used. For various background flows, we specify the wind velocity on the boundary of the domain. The velocities are initially zero everywhere, and the flow velocities at the boundaries are increased until the target background velocity is reached. For the cases of cavity flow, the bottom and sides of the domain are fixed at zero velocity. The top boundary condition is continuous functions in time with zero initial value. Details about the boundary conditions are provided in Sect. 3 with the results.

3 Results

To identify the crucial physical properties for spider ballooning, we solve the fully coupled fluid–structure interaction problem using the immersed boundary method. The numerical simulations are performed with IBAMR (revision 3803) [13] for a single massive spider attached to a flexible, massless dragline. We first consider the spider-dragline system free-falling in a quiescent fluid. We then numerically simulate the free movement of the spider-dragline in uniform background flow and in cavity flow (to approximate the conditions of an eddy). To reveal the dynamics of

takeoff, we tether the spider-dragline system in both uniform and cavity flows and release it after a certain time period. In the quiescent fluid simulations, the bottom boundary of the computational domain is modeled as ground without penetration. In the simulations with uniform background flow, the boundary conditions are set to the prescribed target velocity. In the cavity flow simulations, the bottom and sides of the domain have zero velocity boundary conditions, and the top is set to a uniform velocity.

3.1 Free Fall in a Quiescent Fluid

The spider-dragline system is immersed in quiescent air. Due to gravity, the spider-dragline system free-falls and generates air flow around it. The vorticity

$$\omega = \nabla \times \mathbf{u}$$

of a two-dimensional flow is always perpendicular to the two-dimensional plane and describes the local rotating motion. Therefore, we consider it a scalar field and visualize the flow by its vorticity. Figure 2 shows four snapshots of vorticity during the free fall in a quiescent fluid. Except for the mass of the spider, all other parameters and initial and boundary conditions are set to the same values for these figures. The mass per unit length is set to $M = 2 \times 10^{-6}$, 2×10^{-5}, 4×10^{-5}, and 2×10^{-4} kg/m, respectively. The other key parameters are dragline length $\ell = 0.1$ m, beam bending stiffness constant $\kappa_b = 5.0 \times 10^{-15}$ N·m, spring stiffness coefficient $\kappa_s = 20$ N/m, and the initial position of the dragline's middle point $(x_0, y_0) = (0, 0.15)$.

As the mass of the spider increases, the spider-dragline system falls faster to the ground. The spider-dragline system falls slowly with the smallest mass ($M = 2 \times 10^{-6}$ kg/m), and the vorticity is plotted at time $t = 8$ s in Fig. 2a. For the larger masses, the vorticity is plotted before the spider-dragline system reaches the ground (Fig. 2b-c). Note that red indicates clockwise vorticity and blue indicates counterclockwise vorticity. In the case of the smallest mass, we observe smooth, streaming flow. For the larger masses, $M = 4 \times 10^{-5}$ and 2×10^{-4} kg/m, vortices are alternately shed from the end of the dragline. For the intermediate case, $M = 2 \times 10^{-5}$ kg/m, the vorticity generated by the dragline induces oscillations of the dragline. These phenomena are consistent with the Reynolds number computed using the average settling velocity as the characteristic velocity and the length of the dragline as the characteristic length. For these four simulations, the Reynolds numbers Re_ℓ are about 146, 960, 1450, and 3320, respectively.

To reveal more of the dynamics during the spider-dragline free fall in a quiescent air, Fig. 3 shows the vertical velocity dy/dt of the bottom point of the dragline (the location of the point mass) vs. time. Figure 3a compares the four simulations in which the masses per unit length are varied. For $M = 2 \times 10^{-6}$ kg/m (Fig. 3b), the system falls slowly and continues to accelerate during the entire length of

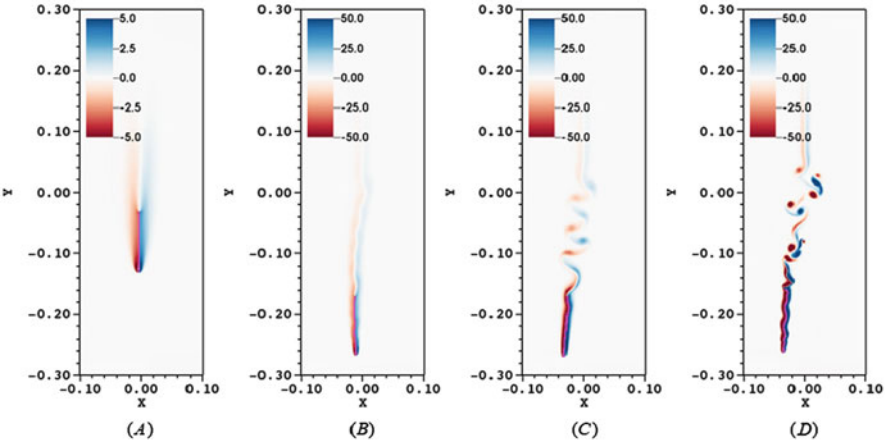


Fig. 2 Vorticity (s^{-1}) of the flow generated by the spider-dragline system during free fall in a quiescent fluid with the spider's mass per unit length set to $M = 2 \times 10^{-6}$, 2×10^{-5}, 4×10^{-5}, and 2×10^{-4} kg/m, respectively. The other parameters are held constant for this set of simulations: string length $\ell = 0.1$ m, beam bending stiffness $\kappa_b = 5 \times 10^{-15}$ N·m, spring stiffness coefficient $\kappa_s = 20$ N/m, and initial position of the middle point $(x_0, y_0) = (0, 0.15)$. Vorticity plots in this paper are generated by VisIt [7]. (**a**) $M = 2 \times 10^{-6}$ kg/m. (**b**) $M = 2 \times 10^{-5}$ kg/m. (**c**) $M = 4 \times 10^{-5}$ kg/m. (**d**) $M = 2 \times 10^{-4}$ kg/m

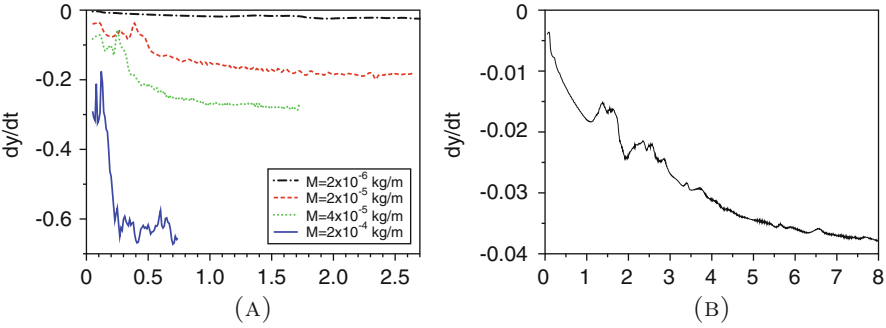


Fig. 3 Vertical velocity (m/s) of the bottom point of the dragline where the spider mass is located. Results are shown for (**a**) the comparison for spiders with masses per unit length of $M = 2 \times 10^{-6}$, 2×10^{-5}, 4×10^{-5}, and 2×10^{-4} kg/m, and (**b**) longer period for spiders with $M = 2 \times 10^{-6}$ kg/m. Except for $M = 2 \times 10^{-6}$ kg/m, the curves end when the spider-dragline reaches the ground

the simulation ($t \leq 8$ s). For the other three masses, vortices develop behind the dragline, the terminal velocities are quickly reached, and the spiders reach the ground before $t = 3$ s. After the spider approaches the ground, the vertical velocity of the dragline is almost zero, except when it waves back and forth horizontally.

Since the dragline velocity sets the effective *Re* of the system, we report the average terminal velocities (or settling speeds) for the different cases as illustrated in Figs. 4 and 5. The average settling speed is computed as the average speed of the middle of the dragline before the spider-dragline system reaches the ground.

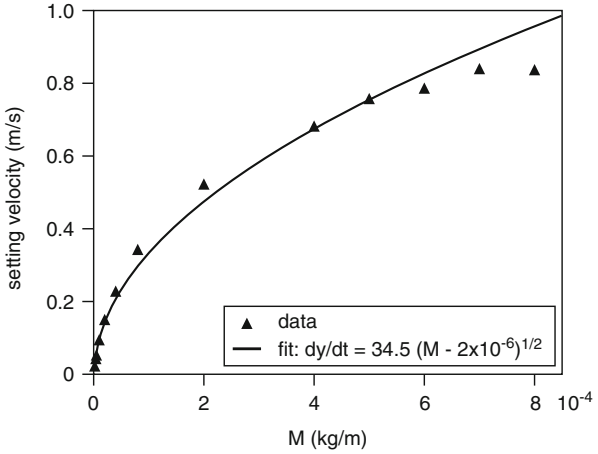


Fig. 4 The average settling velocity (m/s) vs. the spider mass per unit length (kg/m) with the masses set to $M = 2 \times 10^{-6} \sim 8 \times 10^{-4}$ kg/m. The spider-dragline system free-falls in the quiescent air with the dragline length fixed at $\ell = 0.1$ m, spring stiffness coefficient set to $\kappa_s = 20$ N/m, bending modulus fixed at $\kappa_b = 5 \times 10^{-15}$ N·m, and initial position of the middle point set to $(x_0, y_0) = (0, 0.15)$

Figure 4 shows the average settling speed for different masses per unit length with a fixed dragline length $\ell = 0.1$ m. From these results, we see that settling velocity monotonically increases as the mass of the spider increases. Figure 5 shows the average settling speed for different dragline lengths with a fixed spider mass per unit length of $M = 2 \times 10^{-4}$ kg/m. The shorter the dragline, the larger the average settling velocity of the spider-dragline. With a linear least square fit, the relation between the settling velocity vs. the dragline length is $\frac{dy}{dt} = 0.682 - 1.536\ell$. The settling velocity as a function of the spider mass is nonlinear. Using a power fit, we find that $\frac{dy}{dt} = 34.5\sqrt{m - 2 \times 10^{-6}}$.

Recall that in Humphrey's model [15], the dragline is rigid. In the study by Reynolds el al. [22, 23], the silk dragline is described as a line of springs joined at nodes. Those springs themselves are stretchable. At the nodes, the dragline can freely bend in any direction. To more accurately model the dragline, we introduced resistance to bending that is proportional to the bending stiffness modulus, κ_b, as given in Eq. (7). Figure 6 shows the horizontal drift that results from only changing the bending modulus κ_b. We observe that there is no pattern between the direction and magnitude of the horizontal shift and the bending moduli. The direction of the shift depends upon the side on which the first vortex separates from the dragline, highlighting the complicated interaction of the elastic dragline and the fluid.

For all subsequent simulations in Sects. 3.2–3.3, we keep the dragline length fixed at $\ell = 0.1$ m, the bending stiffness set to $\kappa_b = 5 \times 10^{-15}$ N·m, and the spring stiffness coefficient set to $\kappa_s = 20$ N/m. To directly compare the results between different scenarios, we set all other physical parameters to the values used in Fig. 2.

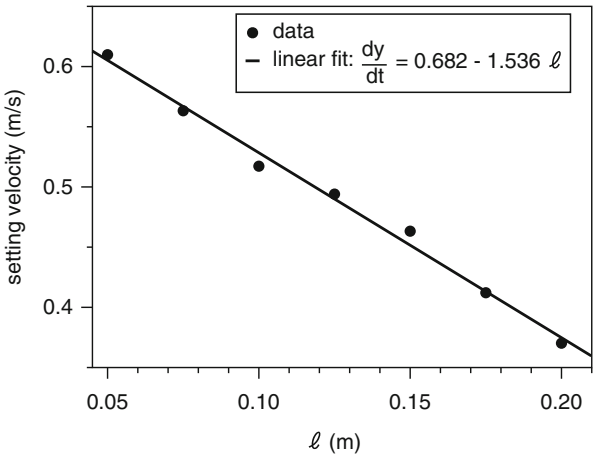


Fig. 5 The average settling velocity vs. the dragline length ℓ (m) varied from $0.075 \sim 0.2$ with the spider mass per unit length set to $M = 2 \times 10^{-4}$ kg/m. The spider-dragline free-falls in the quiescent air with other parameters fixed to the same values as shown in Fig. 4

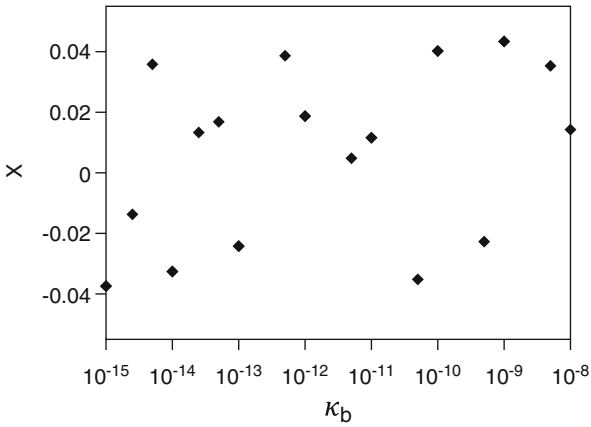


Fig. 6 Horizontal shift x (m) as a function of the beam bending stiffness modulus κ_b (N·m) when the spider-dragline free-falls in quiescent fluid. Spider mass $M = 2 \times 10^{-4}$ kg/m, dragline length $\ell = 0.1$ m, and spring stiffness coefficient $\kappa_s = 20$ N/m

3.2 Free Fall with Background Flows

As spider ballooning is greatly influenced by local meteorological conditions, we simulate spider free fall with two different types of background flows. The first is a uniform background wind and the second is a cavity flow driven by a horizontal velocity at the top of the domain. Note that cavity flow is used to approximate the behavior of a spider ballooning in an eddy. Recall from the free fall results in quiescent air, a spider with a shorter dragline falls faster. In this section, we keep the dragline fixed at $\ell = 0.1$ m. Extrapolating from the study in quiescent air, we can predict that spiders with longer draglines in background flow will also fall more slowly and stay suspended in air for longer periods of time.

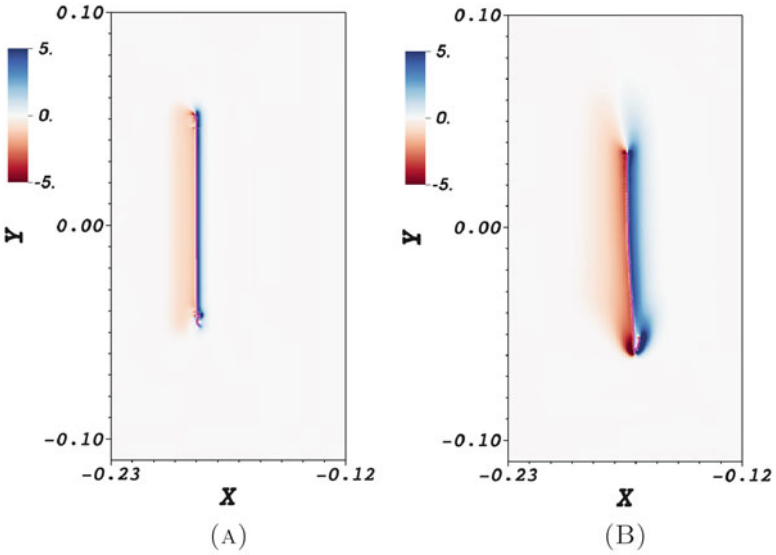


Fig. 7 Vorticity generated by the spider-dragline system free-falling in a uniform background wind with $U = V = 0.008$ m/s, where the prescribed boundary conditions are set to $\mathbf{u}(x,t) = [U,V]^T$. The mass per unit length of the spider is $M = 2\times10^{-6}$ kg/m. (**a**) t = 1.25 s. (**b**) t = 3.25 s

3.2.1 Uniform Background Flows

The vorticity fields in Fig. 7 show the case when the spider-dragline system free-falls in a 45° uniform wind with a constant velocity $[U,V] = [0.008, 0.008]^T$ m/s, where the boundary condition for the simulation was set to $\mathbf{u}(x,t) = [U,V]^T$. The mass per unit length of the spider was set to $M = 2 \times 10^{-6}$ kg/m. Compared to falling in quiescent air, these vorticity plots show slight asymmetry due to the background wind. The vortex developed on the left (upwind direction) side of the dragline has a larger area than on the left side of the dragline as Fig. 7a, but the bottom of the dragline is continually deforming as vorticity grows near the curved tip (seen Fig. 8b). For this set of parameters, the spider-dragline mostly moves with the air.

The profiles in Fig. 8 show the positions of the dragline when the spider-dragline system free-falls in the background winds, which are in the same direction (45°) but with different strengths. The time increment dt between each dragline is 0.25. The velocities of the uniform wind are $[U,V]^T = [0.01, 0.01]^T$, $[0.008, 0.008]^T$, and $[0.005, 0.005]^T$ m/s, respectively. Note that for these simulations, the spider-dragline system has a spider mass per unit length set to $M = 2\times10^{-6}$ kg/m, a dragline length fixed at $\ell = 0.1$ m, beam bending stiffness constant set to $\kappa_b = 5 \times 10^{-15}$ N·m, and spring stiffness coefficient set to $\kappa_s = 20$ N/m. The initial position of the dragline is the dotted line in the figures. With a stronger background wind, advection dominates. The spider-dragline system goes with the flow with little deformation.

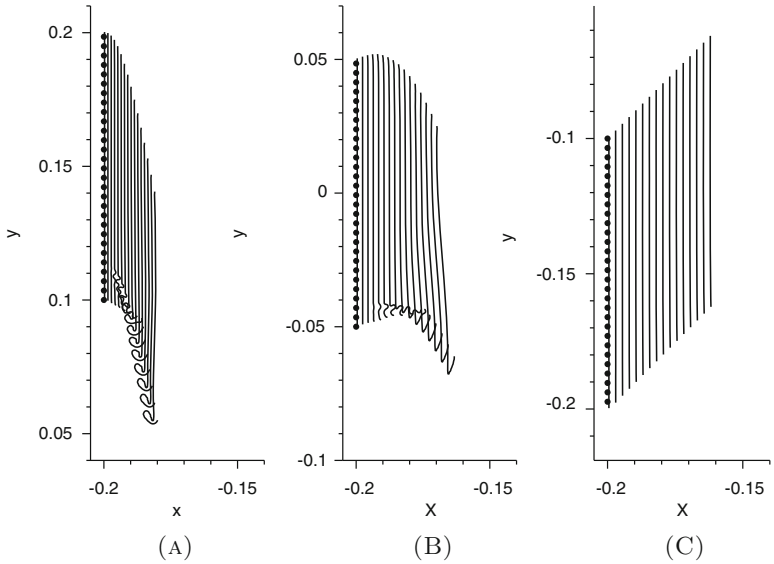


Fig. 8 Positions of the dragline at different snapshots in time with the time increment $dt = 0.25$ s between each dragline. The spider-dragline falls in the background wind (U, V) with different strengths. The *black dotted line* is the initial position. (**a**) $U = V = 0.005$ m/s, (**b**) $U = V = 0.008$ m/s, and (**c**) $U = V = 0.01$ m/s. The spider mass per unit length is set to $M = 2\times10^{-6}$ kg/m

In particular, we observe no entanglement as was reported in Reynolds et al. [22]. Compared to the wind speeds observed for tiptoeing behavior, for example, 1.7–2.6 m/s for *Pardose purbckenisis* [24], the background wind in our study is much weaker. For stronger winds, the spider-dragline system would advect out of the computational domain in a very short period with a similar profile as Fig. 8c. When the background wind is weaker, small deformations appear at the tip of the dragline where the spider is attached, likely due to shearing and the formation of vorticity. Note that the bending modulus is very small relative to the strength of the wind, and the dragline behaves as an extremely flexible line.

Changing the angle of the wind relative to the horizontal and keeping its magnitude constant, we demonstrate the sequence positions of the dragline in Fig. 9. The time increment between each dragline is the same as in Fig. 8, i.e., $dt = 0.25$ s. The spider mass per unit length is set to $M = 2 \times 10^{-6}$ kg/m, and the initial position of the middle point is fixed at $(x_0, y_0) = (-0.2, -0.15)$. The direction of the wind has a significant effect on the trajectory of the ballooning spider. The horizontal component determines the distance it travels along the landscape, and the vertical component combined with the mass of the spider determines whether the spider will land or fly up. With a vertical wind (Fig. 9d), horizontal movement is negligible. Figure 9a presents the situation in which the spider-dragline systems free-fall in a weak horizontal breeze.

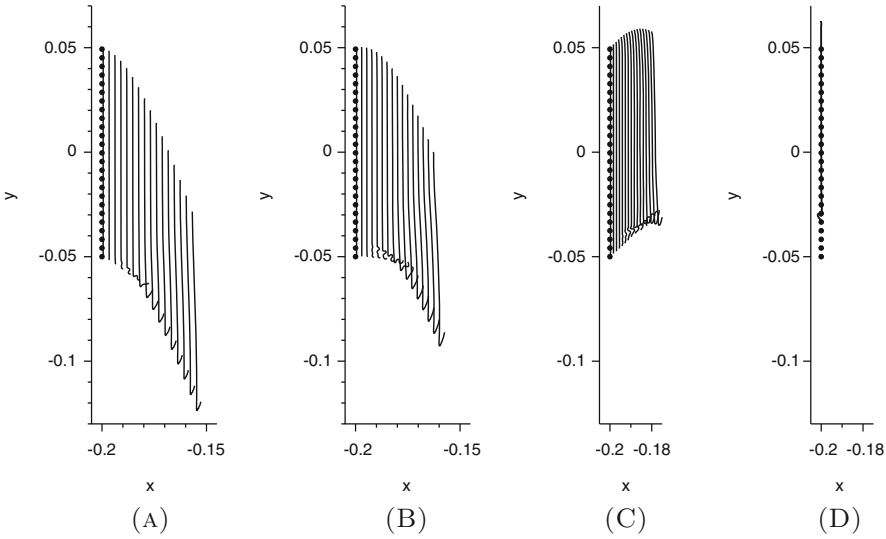


Fig. 9 Profiles of draglines in wind with fixed magnitudes of velocity but different directions: (**a**) horizontal wind, (**b**) 30°, (**c**) 60°, and (**d**) vertical wind. The *black dotted line* indicates the initial position. For the vertical wind case (**d**), only the initial and the end positions in that time period were plotted as there is almost no horizontal movement. Wind strength is fixed at $|\mathbf{U}| = 0.008\sqrt{2}$ m/s. Note that Fig. 8c shows a 45° wind of the same magnitude

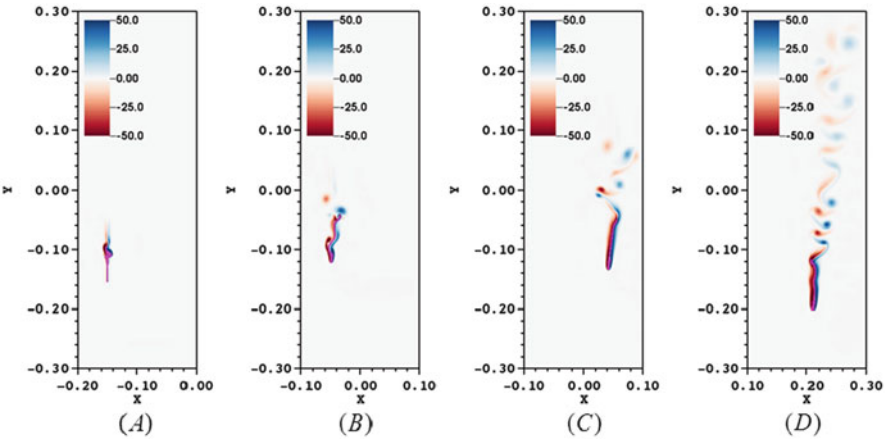


Fig. 10 Vorticity (s^{-1}) snapshots of the flow generated by the spider-dragline system with a 45° background wind, $(U, V) = (0.2, 0.2)$ m/s, at different times. The spider mass $M = 4\times10^{-5}$ kg/m. (**a**) $t = 0.25$ s. (**b**) $t = 0.75$ s. (**c**) $t = 1.25$ s. (**d**) $t = 2.25$ s

With a different spider mass $M = 4 \times 10^{-5}$ kg/m, Figs. 10 and 11 provide more details for the flow field and the dragline when the spider-dragline system free-falls in the 45° uniform wind, $(U, V) = (0.2, 0.2)$ m/s. Other parameters are matched

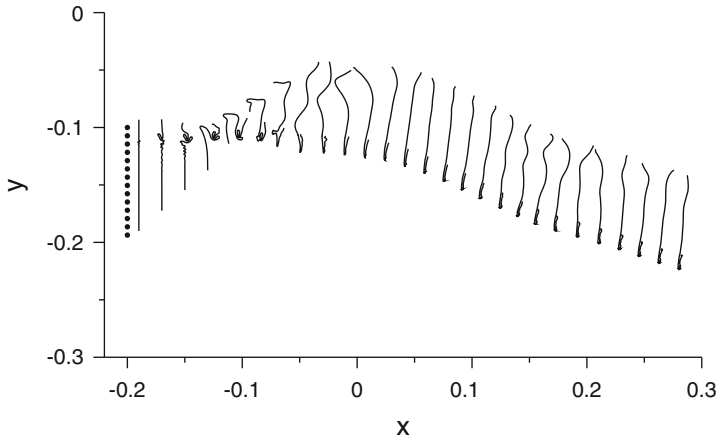


Fig. 11 Positions of the dragline at different times during the simulation. The time increment *dt* between draglines is 0.1 s. The *dotted line indicates* the initial position. The spider-dragline is flying from the left to the right with the flow. The spider mass is set to $M = 4 \times 10^{-5}$ kg/m and the uniform background wind blows 45° with respect to the horizontal and with $(U, V) = (0.2, 0.2)$ m/s

to those reported in Figs. 7, 8 and 9: the dragline length is set to $\ell = 0.1$ m, the bending stiffness is set to $\kappa_b = 5 \times 10^{-15}$ N·m, the spring stiffness is $\kappa_s = 20$ N/m, and the initial position of the middle point is given by $(x_0, y_0) = (-0.2, -0.15)$. As seen in Fig. 2c, the mass per unit length of the spider is sufficient to strongly shear the fluid, resulting in vortex shedding from the tip of the dragline. The magnitude of the resulting vorticity is stronger than for the case of a uniform background wind. Figure 11 shows the profiles of the dragline during the flight. The time increment *dt* between draglines is 0.1. The spider-dragline system moves up due to the background wind and then moves down as the effect of the gravity decelerates the system and produces negative settling velocities. Deformation of the dragline initially occurs toward the top of the dragline as vortices are shed. Eventually, the whole dragline is twisted and later the top of the dragline straightens while the bottom is curved.

In Fig. 12, the spider's mass per unit length is set to $M = 2 \times 10^{-5}$ kg/m. Two snapshots of the vorticity field are shown for free fall in a 45° background wind with $|\mathbf{U}| = 0.2$ m/s. Initially the flow relative to the dragline system is smooth, and eventually vortices develop and are shed from the tip. These vortices induce deformations in the dragline. After some time, the dragline and spider get entangled. These dynamics are distinct from the case of free fall in a quiescent fluid.

Figure 13 shows the profiles of the dragline with winds in three different directions, i.e., 30°, 45°, and 60°. The strength of the wind is fixed at $|\mathbf{U}| = 0.2$ s. The green line is the initial position for all three cases. The time increment *dt* between draglines is 0.15 s. After the spiders are advected about 0.1–0.2 m, the spider and dragline become entangled, resulting in a stable configuration. It is

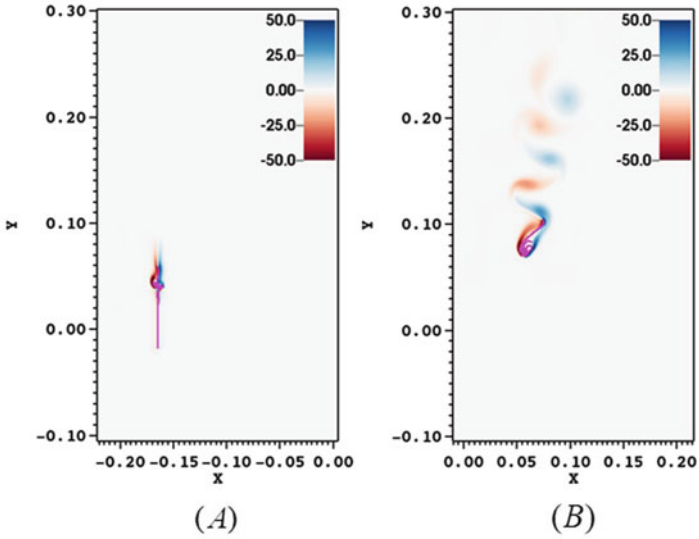


Fig. 12 Snapshots of vorticity (s^{-1}) of the flow during free fall of the spider-dragline with background wind set to $|\mathbf{U}| = 0.2$ m/s at 45° from horizontal. The spider's mass per unit length is set to $M = 2 \times 10^{-5}$ kg/m. The dragline-spider system is shown in pink. (**a**) $t = 0.25$ s. (**b**) $t = 2$ s

possible that this entanglement effectively generates a large surface area that also acts to generate sufficient drag to keep the spiders afloat. By varying the direction of the wind, we change the horizontal and vertical velocity components but keep the same strength of the wind. With a large vertical component (60° wind, blue profiles in Fig. 13), the spider-dragline system keeps rising in the air, while its horizontal shift is decreased compared with the other directions (red and black profiles). In the intermediate case (45° shown in red), the spider-dragline system advects beyond the computational domain. With the smallest vertical component (30° in Fig. 13), the spider-dragline system descends and will drop to the ground eventually, unless the wind direction or speed is changed.

Figure 14 shows the vertical velocity profile for the dragline's tip, where the spider is attached, when the spider-dragline free-falls in uniform winds of the same strength ($|\mathbf{U}| = 0.2$ m/s) but different directions. The spider mass per unit length was fixed at $M = 2 \times 10^{-5}$ kg/m. Initially, the spider moves with the background flow. Due to gravity, the spider-dragline system begins to decelerate. The movement of the spider against the background flow causes shearing, vortex formation, and eventual oscillations in the vertical velocity. The black curve for the 30° wind ends earlier than the other two curves since the spider-dragline system has left the $[-0.3,\ 0.3] \times [-0.3, 0.3]$ computational domain.

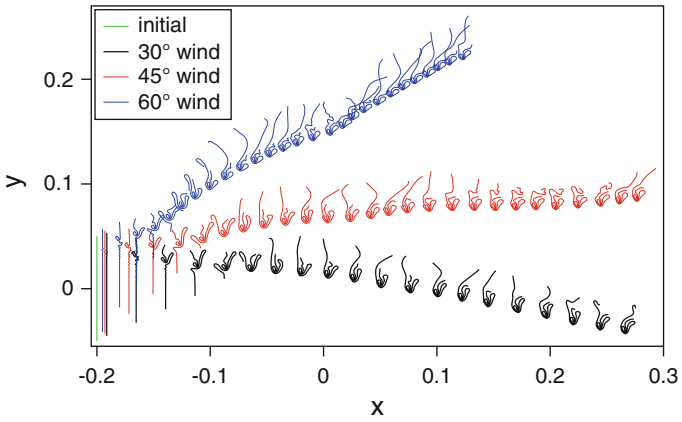


Fig. 13 Profiles of draglines in wind of the same magnitude, $|\mathbf{U}| = 0.2$ m/s, but different directions from the horizontal, 30°, 45°, and 60°, respectively. The *green line* is the initial position. *Blue* is used for 60° wind, *red* for 45° wind, and *black* for 30° wind. The time increment dt between draglines is 0.15 s. The spider's mass per unit length is $M = 2 \times 10^{-5}$ kg/m, and the initial position of the middle point is $(x_0, y_0) = (-0.2, 0)$

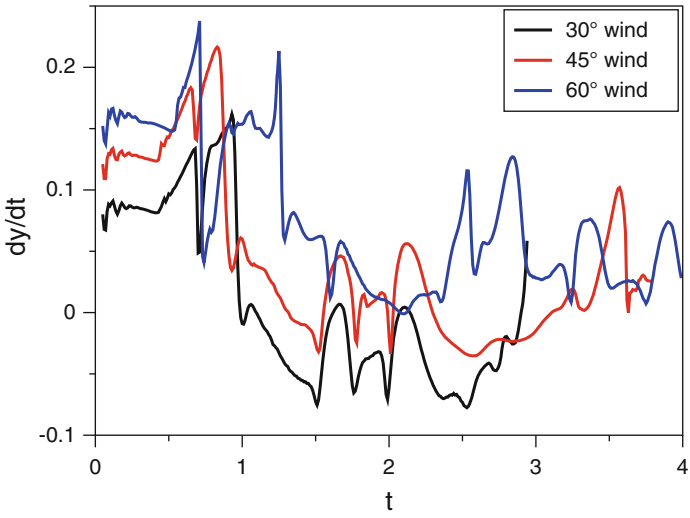


Fig. 14 Vertical velocity vs. time of the spider-dragline system in winds of the same magnitude ($|\mathbf{U}| = 0.2$ m/s) but from different directions

3.2.2 Free Fall in a Cavity Flow

In Sect. 3.2.1, we considered spider ballooning with a uniform background wind. The relevant meteorological conditions for ballooning are not, however, always as simple as uniform flow. To explore the spider ballooning in nonuniform flow, we

simulate the spider-dragline system free fall in a "lid-driven" square cavity flow. Such flows roughly approximate the conditions of ballooning with an eddy.

The no-slip velocity boundary condition ($U = V = 0$) was applied on the bottom and sides of the domain. On the top of the domain, the velocity was set to $U \neq 0$ and $V = 0$. This models a moving lid in a box, which forms an eddy. To avoid the discontinuity for the initial condition, we set the velocity at the top boundary of the domain to a hyperbolic tangent function given by $U = 3\tanh(100t)$ m/s. With this boundary condition, results for the two different spider masses per unit length are considered, $M = 2 \times 10^{-5}$ kg/m and 4×10^{-5} kg/m.

Figure 15 shows vorticity snapshots of the spider-dragline system in a cavity flow. The time values for these vorticity plots are t = 0.6, 0.65, 1, 1.5, 2.2, 2.5, 3, 3.5, and 4 s, respectively. The flow velocity in the domain is initially set to zero. The

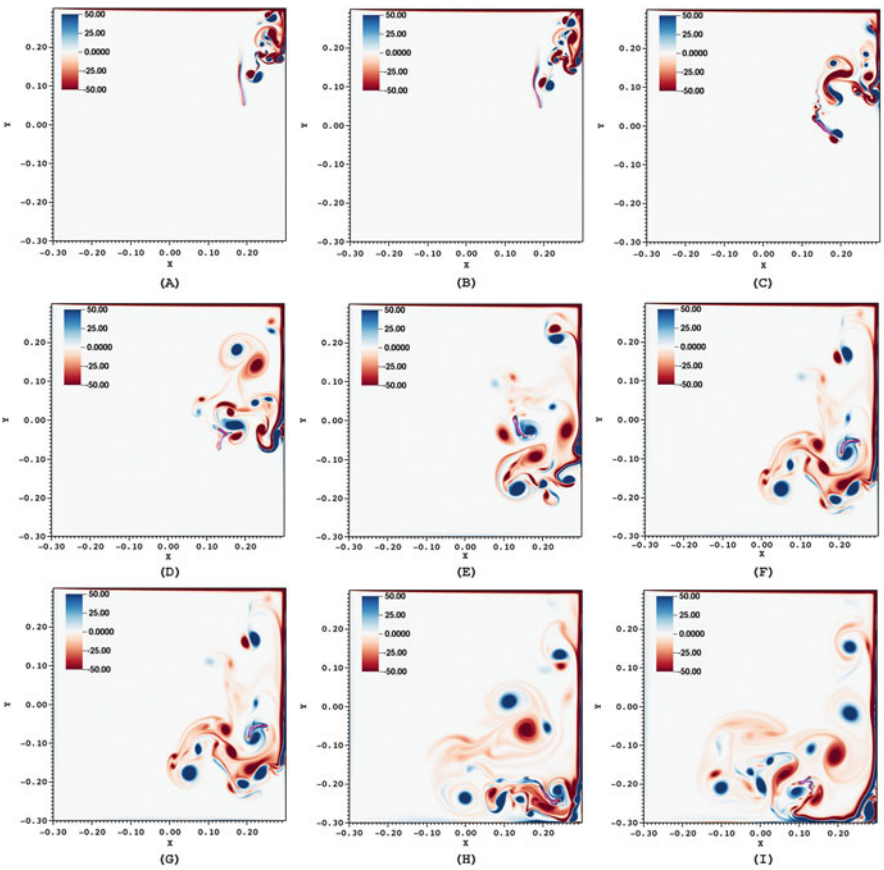


Fig. 15 Vorticity (s^{-1}) snapshots of the spider-dragline system in a cavity flow with background velocity $U = 3\tanh(100t)$ m/s at the top of the domain and the spider mass per unit length is set to $M = 2 \times 10^{-5}$ kg/m. (**a**) t = 0.6 s. (**b**) t = 0.65 s. (**c**) t = 1 s. (**d**) t = 1.5 s. (**e**) t = 2.2 s. (**f**) t = 2.5 s. (**g**) t = 3 s. (**h**) t = 3.5 s. (**i**) t = 4 s

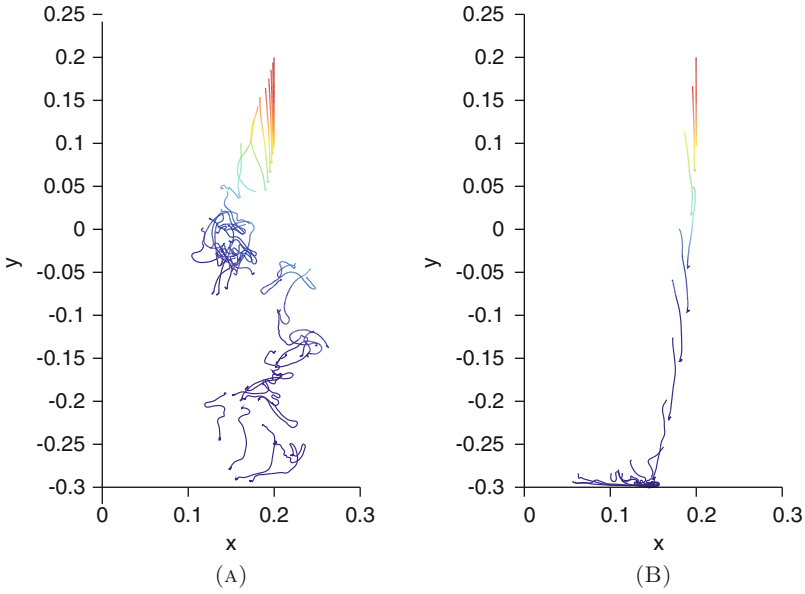


Fig. 16 Dragline vs. time in the cavity flow $U = 3\tanh(100t)$ m/s. The colormap is of the dragline changes from red to blue during the time. (**a**) The spider mass per unit length is $M = 2\times10^{-5}$ kg/m and the time increment dt between draglines is 0.1 s; (**b**) the spider mass per unit length is $M = 4 \times 10^{-5}$ kg/m and the time increment dt is 0.25 s

spider begins to fall as the flow develops. Notice that vortices develop in the upper right corner of the domain due to the velocity at the top. As the cavity flow develops, the spider-dragline system interacts with these vortices in complicated ways.

Figure 16 shows temporal snapshots of the dragline for spiders with masses per unit length of $M = 2 \times 10^{-5}$ kg/m and $M = 4 \times 10^{-5}$ kg/m. The initial positions are the same for both simulations. The time increment dt between draglines is 0.1 s for Fig. 16a; while $dt = 0.25$ s for Fig. 16b. More frames are plotted to show the dynamics in Fig. 16a. The lighter spider, Fig. 16a, settles slowly and eventually interacts with the vortices developed in the cavity flow. During this interaction, the spider-dragline system becomes entangled. For the heavier spider with mass per length set to $M = 4 \times 10^{-5}$ kg/m, the situation is simple: The spider settles before the cavity flow develops. Note that more frames are shown in the plot for $M = 2 \times 10^{-5}$ kg/m than for $M = 4 \times 10^{-5}$ kg/m in Fig. 16b since the snapshots end when the spider hits the ground.

Figure 17 shows the vertical velocity of the bottom point of the dragline vs. time, which corresponds to the dragline profiles shown in Fig. 16. The heavy spider with $M = 4\times10^{-5}$ kg/m, shown as the blue curve, has a large downward vertical velocity until it hits the ground around $t = 2$ s. Recall that it falls to the ground before the cavity flow develops, and no upward motion is observed. For the lighter spider with mass per unit length set to $M = 2 \times 10^{-5}$ kg/m, the vertical velocity oscillates from positive to negative. Positive velocity can be attributed to the interaction of the spider-dragline with the background vortices due to cavity flow.

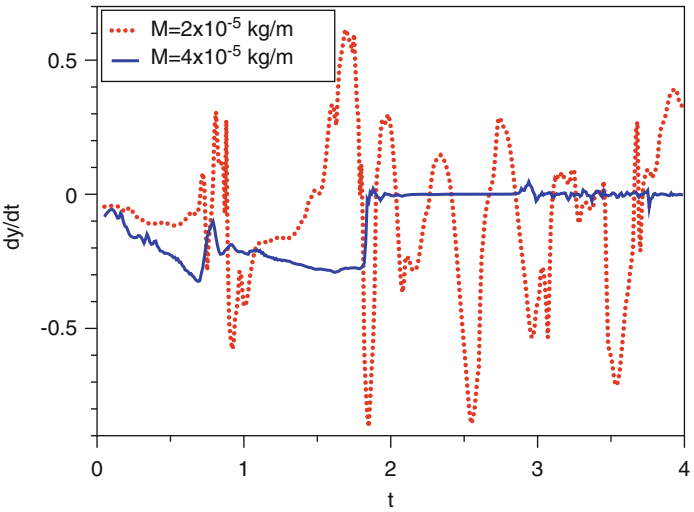


Fig. 17 Vertical velocity (m/s) of the end of the dragline attached to the spider vs. time (s) for free fall in the cavity flow. Spiders with masses per unit length of $M = 2 \times 10^{-5}$ kg/m and $M = 4 \times 10^{-5}$ kg/m are shown. (**a**) t = 0.3 s. (**b**) t = 0.6 s (before release). (**c**) t = 1.04 s (after release). (**d**) t = 1.2 s

By comparing the dragline dynamics with different background winds, we have found that the details of the air movement are important for determining the amount of time the spider spends in the air and the distances traveled. Not only the strength but also the direction and local dynamics of the wind are critical. However, when a spider initiates the climb to a tiptoe position, what are the important signals available to control the subsequent takeoff? To explore this question, we simulate the spider-dragline tethered in the flow to simulate tiptoeing. We then release the spider to examine the dynamics of takeoff.

3.3 Dynamics of Takeoff

Herein, we identify the mechanical factors of takeoff associated with spider ballooning by simulating the spider-dragline system tethered in the flow and released. Beside the flow field and the dynamics of the dragline, the force acting on the tether is analyzed. Note that the spider mass per unit length is fixed at $M = 2 \times 10^{-5}$ kg/m for the subsequent simulations.

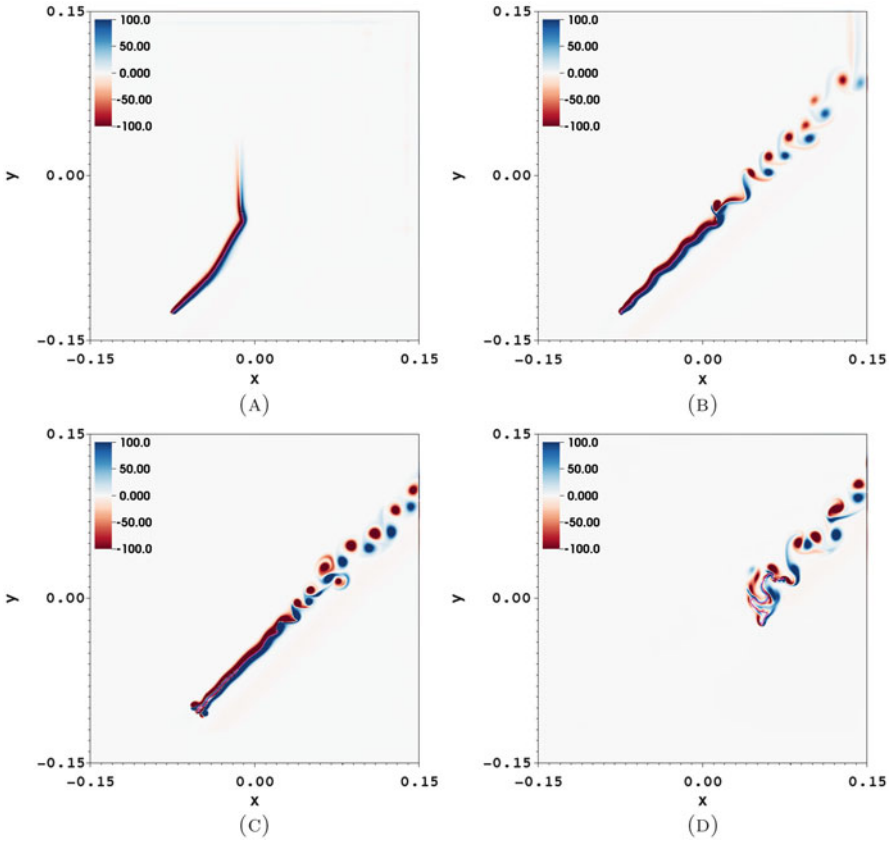


Fig. 18 Vorticity (s^{-1}) snapshots showing the dynamics of a tethered spider-dragline that is released in a 45° wind with strength $|\mathbf{U}| = 1$ m/s. The spider is released at $t = 1$ s. Here, $M = 2 \times 10^{-5}$ kg/m. The *pink curve* is the dragline

3.3.1 Takeoff in Uniform Winds

Four snapshots in time of the vorticity of the flow are shown in Fig. 18. At the earliest time, $t = 0.3$ s (in Fig. 18a), the dragline gradually tilts and aligns with the background wind profile. Subsequently, flapping and shedding of alternately spinning vortices begins as shown in Fig. 18b at t = 0.6 s. Once release occurs (as shown in Fig. 18c-d at t = 1.04–1.2 s), the spider and dragline entangle and move with the background wind. More deformations are created by the vortex in the surrounding air as seen in Fig. 18d at t = 1.2 s.

Figure 19 shows the force per unit length (N/m) acting on the spider when it is tethered ($t < 1$) with different wind directions. The strength of the wind is fixed at $|\mathbf{U}| = 1$ m/s. As the dragline is massless in our model, the comparison confirmed that the tether force is of a similar magnitude. At the beginning of simulation, the

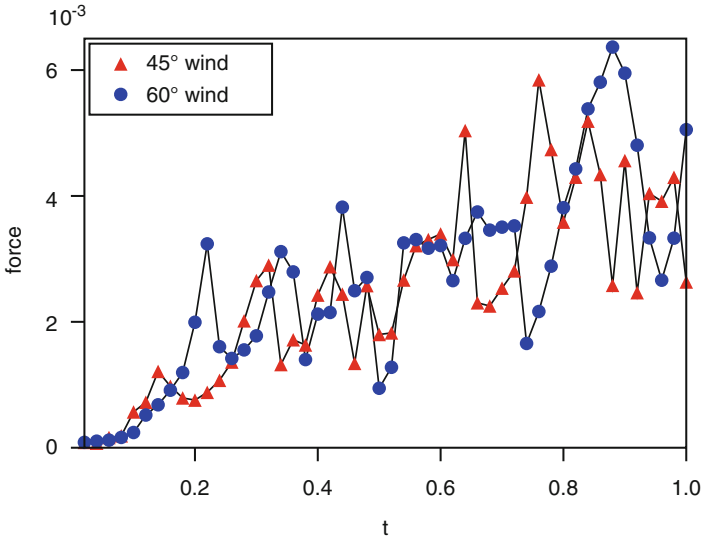


Fig. 19 Force per unit length (N/m) acting on the tether vs. time (s) in a uniform wind moving in two directions (45° and 60°)

forces per unit length increase as the draglines align with the uniform background and are slightly stretched. Around $t = 0.15 - 0.2$ s, vortex shedding begins and the forces per unit length begin to oscillate. This is interesting since laboratory studies show that the length of time spent attempting to takeoff is a factor for whether or not to balloon [35]. This may be correlated to the alignment of the dragline with the wind and the dynamical forces experienced by the spider.

In order to visualize the spider-dragline system in the flow, Fig. 20 shows successive positions of the dragline at selective time points representative of the typical stages of tethering, release, and free flight. The blue line shows the dragline in a uniform wind that is directed 45° from the horizontal, and the red lines show the draglines in a wind directed 60° from the horizontal. The green line shows the initial position for the spider-dragline system. Dashed or dotted lines show the profiles of the dragline while the spider is tethered at position $(-0.75, -0.75)$ and for $t \leq 1$ s. During the tether ($t \leq 1$), the time increment between draglines is $dt = 0.15$ s for both background winds. After the release ($t > 1$ s), the draglines are plotted at $t =$ 2.25, 2.8, 2.9, 3, 3.1, 3.2, and 3.3 s in the 45° wind, and $t = 2.15$, 2.85, 3.05, 3.2, and 3.35 s in the 60° wind. These selective stages demonstrate the dynamics of the spider-dragline but the draglines are not overlapped for visualization purpose.

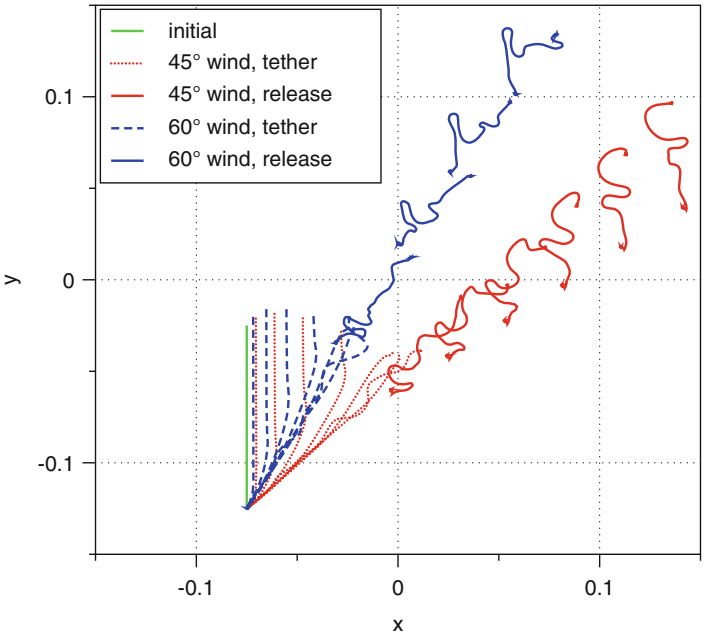


Fig. 20 Snapshots of the dragline at different instances in time for a 45° and a 60° background wind. The spider is released at $t = 1$ s. For both cases, the mass per unit length of the spider is set to $M = 2 \times 10^{-5}$ kg/m. The *green line* is the initial position. During the tether ($t \leq 1$), the time increment between draglines is $dt = 0.15$ s for both background winds. After the release ($t > 1$ s), the draglines are plotted at $t = 2.25$, 2.8, 2.9, 3, 3.1, 3.2, and 3.3 s in the 45° wind, and $t = 2.15$, 2.85, 3.05, 3.2, and 3.35 s in the 60° wind

3.3.2 Takeoff in a Cavity Flow

To further study the dynamics of takeoff, we also simulated the spider-dragline in nonuniform wind, i.e., cavity flow. The setup for cavity flow was the same as performed in Sect. 3.2.2: the flow starts at rest; the horizontal velocity on the top of the domain is set to $U = 3\tanh(100t)$ m/s and the velocities on the other three boundaries are all set to zero; and eddies are formed within the computational domain. For the spider-dragline system, the dragline was initially positioned vertically above the spider. At the beginning of the simulation, the spider was tethered at the center of the domain and was then released in the flow at t = 2 s. The spider mass per unit length in all cases was set to $M = 2 \times 10^{-5}$ kg/m.

Figures 21 and 22 show eight representative snapshots of the vorticity field with the dragline colored in pink. Figure 21 shows four snapshots of the vorticity field during the tether ($t \leq 2$ s). Figure 22 shows four snapshots after release ($t > 2$ s). While the spider-dragline system is tethered, the dragline waves around and interacts with the flow. The dragline sometime breaks up the vortices developed due to the background cavity flow, and the dragline itself sheds vortices in an alternate pattern,

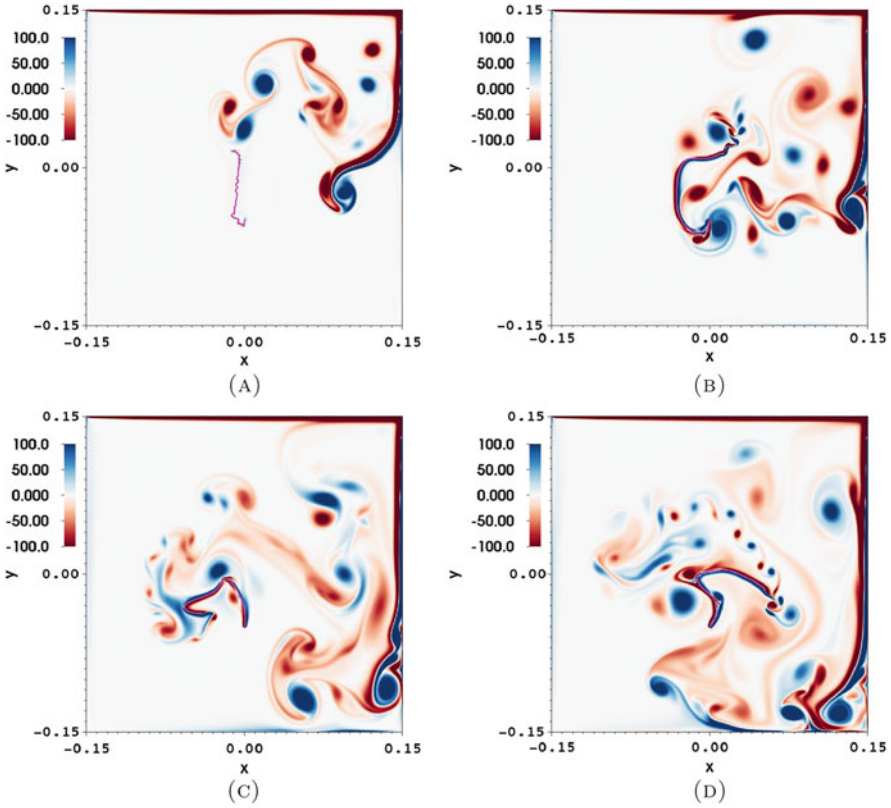


Fig. 21 Vorticity (s^{-1}) snapshots showing the spider-dragline system in the cavity flow. The spider-dragline is tethered for $t \leq 2$ s and released at $t = 2$ s. The dragline is shown in pink. (**a**) t = 0.8 s. (**b**) t = 1.2 s. (**c**) t = 1.6 s. (**d**) t = 2 s

as seen in Fig. 21d. Once the spider-dragline system is released, it free-falls and is advected in the cavity flow, which is dominated by the large eddies moving around the domain.

Figures 23 and 24 show the profiles of dragline during tether and release, respectively. Initially, the dragline is positioned vertically as shown by the green straight line near the center of the domain. From the profiles of the dragline at different instances in time during the tether (Fig. 23), we see that the dragline is swirled by the cavity flow due to the nonzero velocity imposed on the top boundary of the domain. Once the spider-dragline is released from the tether as shown in Fig. 24, it mostly moves with the cavity flow though there are some effects due to gravity. For a longer time simulation, we have observed that the spider continues to be advected round and round the large eddy produced by the cavity flow.

Figure 25 shows the force per unit length acting on tether during $t \leq 2$ s. When $t < 1$ s, the force is negligible. The massive spider is fixed and the massless dragline

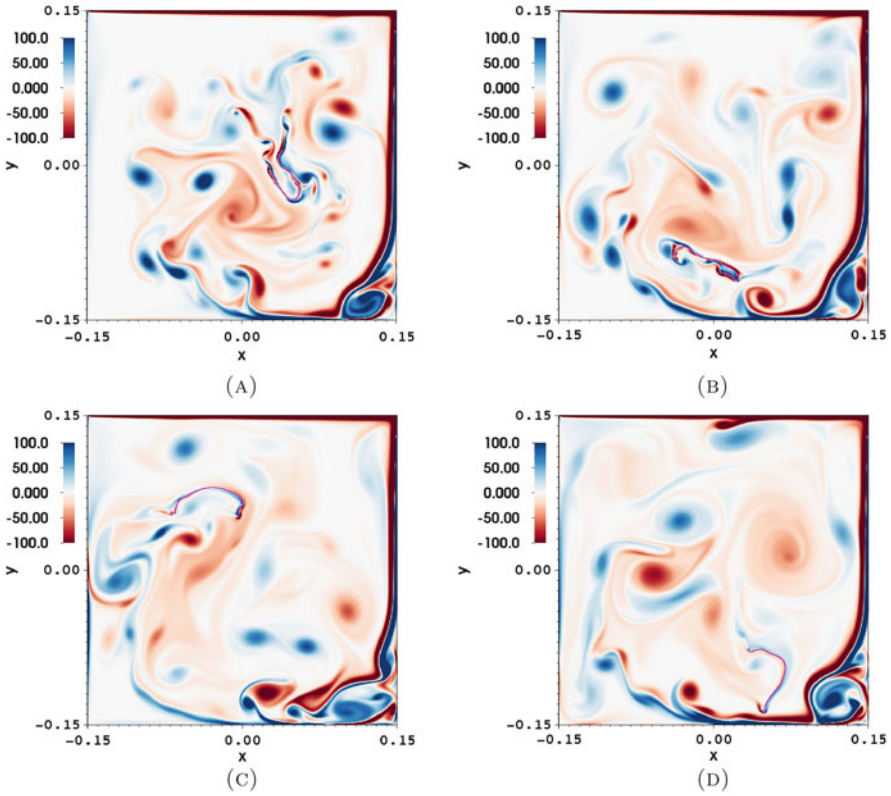


Fig. 22 Vorticity (s^{-1}) at various times in the cavity flow (continued). The spider-dragline is tethered for $t \leq 2\,s$ and released at $t = 2\,s$. (**a**) t = 2.4 s. (**b**) t = 2.8 s. (**c**) t = 3.2 s. (**d**) t = 4 s

does not move in a nearly "quiescent" fluid. The fluid motion produced by the nonzero velocity imposed at the top of the domain generates vortices as shown in Fig. 21a. After $t > 1\,s$, the interaction between the dragline and the flow is intense. Near $t = 1.2\,s$, a vortex with positive vorticity directly reaches the spider. The force is dramatically increased at that time. The instability of the flow field causes large variations in the tether force. When spiders in the tiptoe position can sense this flow force acting on them, they might utilize the force as a signal for further unsteady fluid motion and eventually takeoff given some threshold.

4 Conclusions

By numerically solving the fully coupled fluid–structure interaction problem of a flexible dragline in a viscous fluid, we have revealed new phenomena that cannot be captured by simpler models that neglect how the presence of the dragline affects

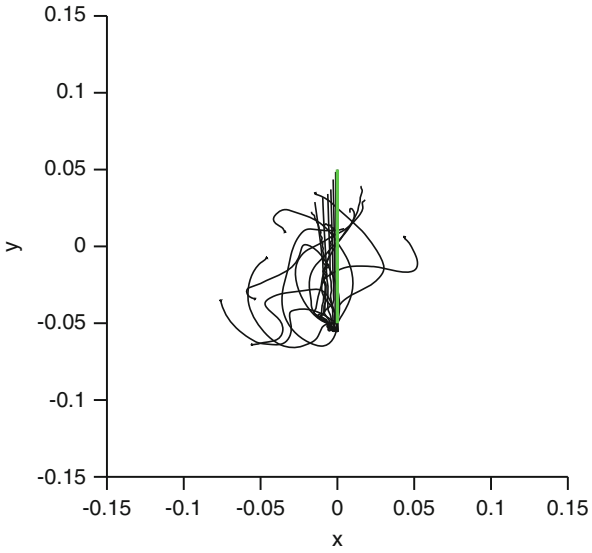


Fig. 23 Temporal snapshots showing the position of the spider-dragline while it is tethered. The spider is tethered for $t \leq 2$ s. The *green line* displays the initial position of the dragline, and the dragline deforms as the cavity flow develops. The time increment is $dt = 0.1$ s between two successive draglines

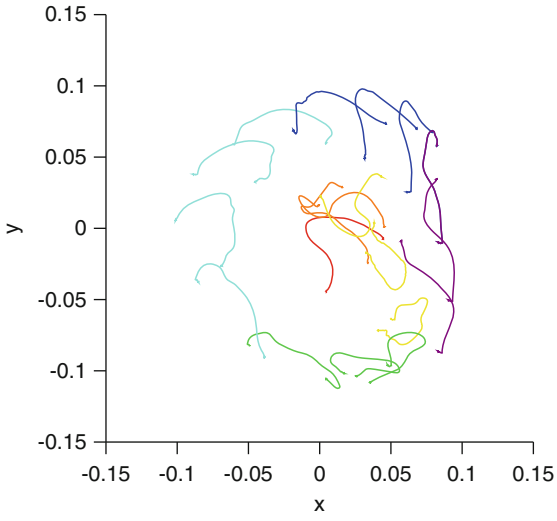


Fig. 24 Temporal snapshots showing the position of the dragline after release ($t > 2$ s). The rainbow colors correspond to advancement in time, with the *black line* showing the position at the end of the simulation. The time increment is $dt = 0.1$ s as Fig. 23

the motion of the air. Our results show that for $Re > 10^3$, strong vortex shedding occurs at the end of the dragline, resulting in oscillations of the dragline itself and some horizontal movement of the spider as it falls through a quiescent fluid. Strong vortex shedding is also present before takeoff, which may generate higher transient

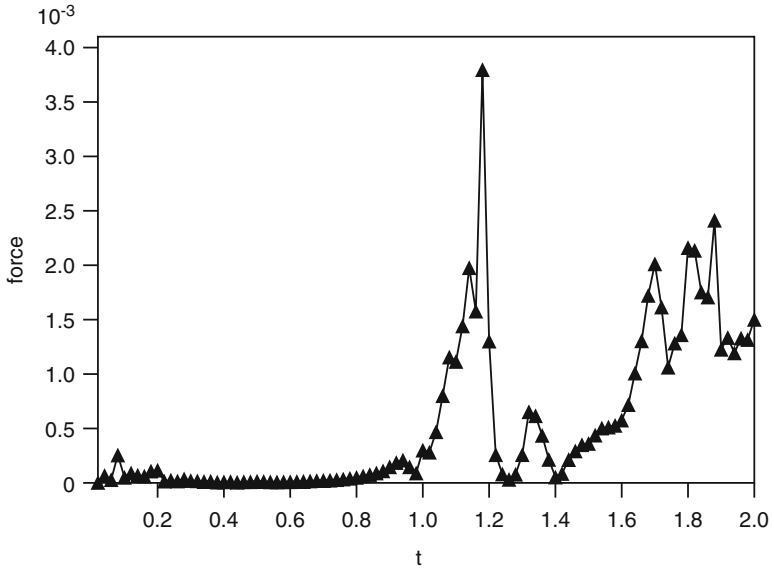


Fig. 25 Force per unit length (N/m) acting on the point where the spider is tethered vs. time (s). The spider is released at $t = 2$ s. The large forces beginning after $t = 1$ s are due to vortices produced by the cavity flow interacting with the dragline

forces to lift the spider into the air. It is also possible that the spiders can sense the vortex shedding frequency and use it to inform whether or not to take off since shedding frequency will vary directly with wind speed. This information could also be used to determine how much longer to make the dragline since the dynamics of the oscillations will also depend upon dragline length.

Our results show that for the parameters considered, the settling velocity varies linearly with the length of the dragline and nonlinearly with the mass of the spider. At the settling velocity, the gravitational force, Mg, balances the drag. Drag varies linearly with length and linearly with velocity for $Re << 1$ and quadratically for $Re >> 1$. For the set of parameters considered, the observed linear relationship suggests a lower Re scaling between force and velocity. When varying the mass of the spider and keeping the length of the dragline constant, a nonlinear relationship is observed between force and velocity since the gravitational force and drag are balanced. This suggests that the larger masses and resulting higher settling velocities push the system to a higher Re scaling.

Direct comparison of settling velocities resulting from the two-dimensional simulations and those of actual three-dimensional spiders is not straightforward. In a two-dimensional simulation, we are essentially modeling an infinitely long sheet which will have higher drag than a one-dimensional line. The mass of the spider must be scaled accordingly, but the relationship between the drag produced by a one-dimensional string and a two-dimensional sheet across intermediate Reynolds

numbers in unsteady flow is not obvious. As a crude estimate, we divide the mass of an actual spider by the diameter of the spider to obtain a mass per unit length. The resulting settling velocities are within the range of those observed for actual spiders [30, 31].

Interestingly, in many simulations, the dragline bends at the tip where the spider is attached. This is perhaps not surprising given the strong vorticity that forms at this leading tip and the low resistance to bending of the dragline. It is not clear if such strong bending would occur in three dimensions or in the presence of electrostatic forces. Similarly, the entire dragline becomes "entangled" in cases where the flow is unsteady. This is particularly true for the movement of the spider within a cavity and also within a crosswind. The tangling of the dragline was also predicted by Reynolds et al. [23]. It is not clear if this phenomenon occurs during actual spider ballooning.

Complex transport dynamics are observed in updrafts and eddies. When eddies are present in the background flow, the dragline may quickly become entangled. It is also possible in these cases for the spider to swirl through the air and remain suspended in the air column as in Fig. 24. In uniform background flows, strong vortex shedding from the tip of the dragline can result in tangling of the dragline after takeoff. Depending upon the entanglement pattern, the dragline may effectively act as a bluff body with finite width, potentially increasing the drag coefficient and lowering the settling velocity.

A natural next step for this work is to move into three dimensions. As mentioned above, the two-dimensional simulations essentially represent a sheet that is infinitely long in the direction moving into and out of the two-dimensional plane. It is likely that the interactions of a sheet with a fluid and its dynamics would be rather different from a one-dimensional dragline. This extension would also allow us to consider multiple draglines that are used by some species of spider for ballooning.

Acknowledgements We are grateful to the National Institute for Mathematical and Biological Synthesis (NIMBioS), which is sponsored by the National Science Foundation (NSF: award DBI-1300426) and The University of Tennessee, Knoxville, for hosting our working group as part of the Research Collaboration Workshop for Women in Mathematical Biology. We especially thank Dr. Anita Layton for organizing the NIMBioS workshop. Additional funding was provided by NSF to KSS (Postdoctoral Research Fellowship 1306883), LAM (CBET 1511427), AC (Graduate Research Fellowship 201315897), and LB (Reeds and Edgerton Funds).

References

1. Batchelor, G.K.: Introduction to Fluid Mechanics. Cambridge University Press, Cambridge (1999)
2. Bell, J.R., Bohan, D.A., Fevre, R.L., Weyman, G.S.: Can simple experimental electronics simulate the dispersal phase of spider ballooners? J. Arachnol. **33**(2), 523–532 (2005)
3. Berger, M.J., Colella, P.: Local adaptive mesh refinement for shock hydrodynamics. J. Comput. Phys. **82**(1), 64–84 (1989)

4. Berger, M.J., Oliger, J.: Adaptive mesh refinement for hyperbolic partial-differential equations. J. Comput. Phys. **53**(3), 484–512 (1984)
5. Bonte, D., Vandenbroecke, N., Lens, L., Maelfait, J.: Low propensity for aerial dispersal in specialist spiders from fragmented landscapes. Proc. R. Soc. B **270**(1524), 1601–7 (2003)
6. Bonte, D., Van Dyck, H., Bullock, J.M., Coulon, A., Delgado, M., Gibbs, M., Lehouck, V., Matthysen, E., Mustin, K., Saastamoinen, M., Schtickzelle, N., Stevens, V.M., Vandewoestijne, S., Baguette, M., Barton, K., Benton, T.G., Chaput-Bardy, A., Clobert, J., Dytham, C., Hovestadt, T., Meier, C.M., Palmer, S.C.F., Turlure, C., Travis, J.M.J.: Costs of dispersal. Biol. Rev. **87**(2), 290–312 (2012)
7. Childs, H., Brugger, E., Whitlock, B., Meredith, J., Ahern, S., Pugmire, D., Biagas, K., et al.: VisIt: an end-user tool for visualizing and analyzing very large data (2012). http://www.osti.gov/scitech/servlets/purl/1170761
8. Coyle, F., Greenstone, M.H., Hultsch, A.-L., Morgan, C.E.: Ballooning mygalomorphs: estimates of the masses of Sphodros and ummidia ballooners. J. Arachnol. **13**, 291–296 (1985)
9. De Meester, N., Bonte, D.: Information use and density-dependent emigration in an agrobiont spider. Behav. Ecol. **21**(5), 992–998 (2010)
10. Glick, P.A.: The distribution of insects, spiders, and mites in the air. Tech. Bull. U.S. Dept. Agric. **673**, 1–150 (1939)
11. Gorham, P.W.: Ballooning spiders: the case for electrostatic flight. ArXiv:1309.4731v1 (2013)
12. Gressitt, J.L.: Biogeography and ecology of land arthropods of Antarctica. In: Mieghem, J., Oye, P. (eds.) Biogeography and Ecology in Antarctica. Monographiae Biologicae, vol. 15, pp. 431–490. Springer, Netherlands/Dordrecht (1965)
13. Griffith, B.: An adaptive and distributed-memory parallel implementation of the immersed boundary (IB) method (IBAMR). https://github.com/IBAMR/IBAMR (2014)
14. Herschlag, G., Miller, L.A.: Reynolds number limits for jet propulsion: a numerical study of simplified jellyfish. J. Theor. Biol. **285**(1), 2369–2381 (2011)
15. Humphrey, J.A.C.: Fluid mechanic constraints on spider ballooning. Oecologia **73**, 469–477 (1987)
16. Kim, Y., Peskin, C.S.: Penalty immersed boundary method for an elastic boundary with mass. Phys. Fluids **19**, 053103 (18 pages) (2007)
17. Miller, L.A., Peskin, C.S.: Flexible clap and fling in tiny insect flight. J. Exp. Biol. **212**, 3076–3090 (2009)
18. Mittal, R., Iaccarino, G.: Immersed boundary methods. Annu. Rev. Fluid Mech. **37**, 239–61 (2005)
19. Nathan, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A., Tsoar, A.: Mechanisms of long-distance seed dispersal. Trends Ecol. Evol. **23**, 638–647 (2008)
20. Peskin, C.S.: Flow patterns around heart valves: a numerical method. J. Comput. Phys. **10**, 252–271 (1972)
21. Peskin, C.S.: The immersed boundary method. Acta Numer. **11**, 479–517 (2002)
22. Reynolds, A.M., Bohan, D.A., Bell, J.R.: Ballooning dispersal in arthropod taxa with convergent behaviours: dynamic properties of ballooning silk in turbulent flows. Biol. Lett. **2**(3), 371–3 (2006)
23. Reynolds, A.M., Bohan, D.A., Bell, J.R.: Ballooning dispersal in arthropod taxa: conditions at take-off. Biol. Lett. **3**(3), 237–40 (2007)
24. Richter, C.J.: Aerial dispersal in relation to habitat in eight wolf spider species. Oecologia **214**, 200–214 (1970)
25. Ronce, O.: How does it feel to be like a rolling stone? ten questions about dispersal evolution. Annu. Rev. Ecol. Evol. Syst. **38**(1), 231–253 (2007)
26. Shao, Z., Hu, X.W., Frische, S., Vollrath, F.: Heterogeneous morphology of Nephila edulis spider silk and its significance for mechanical properties. Polymer **40**(16), 4709–4711 (1999)
27. Sheldon, K.S., Zhao, L., Chuang, A., Panayotova, I.N., Miller, L.A., Bourouiba, L.: Revisiting the physics of spider ballooning. In: Layton, A.T., Miller, L.A. (eds.) Women in Mathematical Biology. Association for Women in Mathematics Series, vol. 8 (2017). doi:10.1007/978-3-319-60304-9_9, 125–139

28. Sobey, I.J.: Oscillatory flows at intermediate Strouhal number in asymmetric channels. J. Fluid Mech. **125**, 359–373 (1982)
29. Stauffer, S.L., Coguill, S.L., Lewis, R.V.: Comparison of physical properties of three silks from Nephila clavipes and Araneus gemmoides. J. Arachnol. **22**(1), 5–11 (1994)
30. Suter, R.B.: Ballooning in spiders: results of wind tunnel experiments. Ethol. Ecol. Evol. **3**(1), 13–25 (1991)
31. Suter, R.B.: Ballooning: data from spiders in freefall indicate the importance of posture. J. Arachnol. **20**, 107–113 (1992)
32. Thomas, C.F.G., Hol, E.H.A., Everts, J.W.: Modelling the diffusion component of dispersal during recovery of a population of linyphiid spiders from exposure to an insecticide. Funct. Ecol. **4**(3), 357–368 (1990)
33. Thomas, C.F.G., Brain, P., Jepson, P.C.: Aerial activity of linyphiid spiders: modelling dispersal distances from meteorology and behaviour. J. Appl. Ecol. **40**(5), 912–927 (2003)
34. Tytell, E.D., Hsu, C.-Y., Fauci, L.J.: The role of mechanical resonance in the neural control of swimming in fishes. Zoology **117**(1), 48–56 (2014)
35. Weyman, G.S.: Laboratory studies of the factors stimulating ballooning behavior by linyphiid spiders (Araneae, Linyphiidae). J. Arachnol. **23**(25), 75–84 (1995)

On the Dynamic Suction Pumping of Blood Cells in Tubular Hearts

Nicholas A. Battista, Andrea N. Lane, and Laura A. Miller

Abstract Around the third week after gestation in embryonic development, the human heart consists only of a valveless tube, unlike a fully developed adult heart, which is multi-chambered. At this stage in development, the heart valves have not formed and so net flow of blood through the heart must be driven by a different mechanism. It is hypothesized that there are two possible mechanisms that drive blood flow at this stage—Liebau pumping (dynamic suction pumping (DSP) or valveless pumping) and peristaltic pumping. We implement the immersed boundary method (IBM) with adaptive mesh refinement (IBAMR) to numerically study the effect of hematocrit on the circulation around a valveless tube. Both peristalsis and DSP are considered. In the case of DSP, the heart and circulatory system is simplified as a flexible tube attached to a relatively rigid racetrack. For some Womersley number (*Wo*) regimes, there is significant net flow around the racetrack. We find that the addition of flexible blood cells does not significantly affect flow rates within the tube for $Wo \leq 10$, except in the case for $Wo \approx 1.5$ where we see a decrease in average flow with increasing volume fraction. On the other hand, peristalsis consistently drives blood around the racetrack for all *Wo* and for all hematocrit considered.

1 Introduction

The Liebau pump (dynamic suction pump), first described in 1954 [37], was studied as a novel way to pump water. It has not been until the past 20 years that scientists started looking at the pump as a valveless pumping mechanism in many biological

N.A. Battista (✉) • L.A. Miller
Departments of Mathematics and Biology, University of North Carolina, Chapel Hill, NC 27599, USA
e-mail: nick.battista@unc.edu; lam9@unc.edu

A.N. Lane
Departments of Biostatistics and Mathematics, UNC Gillings School of Global Public Health, Chapel Hill, NC 27599, USA
e-mail: anlane@live.unc.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_11

systems and biomedical applications, including microelectromechanical (MEM) systems and micro-fluidic devices. Direct applications of such pumps include tissue engineering, implantable microelectrodes, and drug delivery [10, 29, 30, 37].

With extensive industrial applications, dynamic suction pumping (DSP) has proven to be a suitable means of transport for fluids and other materials in a valveless system, for scales of $Wo > 1$ [6]. DSP can be most simply described by an isolated region of actuation, located asymmetrically along a flexible tube with stiffer ends. Flexibility of the tube is required to allow passive elastic traveling waves, which augment bulk transport throughout the system. The rigid ends of the tube cause reflections of the elastic waves, which when coupled with an asymmetric actuation point, can promote unidirectional flow. DSP is illustrated in Fig. 1.

Due to a coupling between the system's geometry, material properties of the tube wall, and pumping mechanics, there is a complex, nonlinear relationship between volumetric flow rate and pumping frequency [6, 9, 20]. Analytic models of DSP have been developed to address this relationship [2, 4, 9, 35, 43, 49]. Most models use simplifications such as the inviscid assumption, long wave approximation, small contraction amplitude, and one-dimensional flow. Furthermore, no analytical model has described flow reversals, which can occur with changes in the pumping frequency. Relaxing many of these assumptions, physical experiments have been performed to better understand DSP [9, 19, 20, 37], as well as in silico investigations [3, 5, 6, 24, 25]. Most of these experimental and computational studies focus on the "high" *Wo* regime ($Wo \gg 1$).

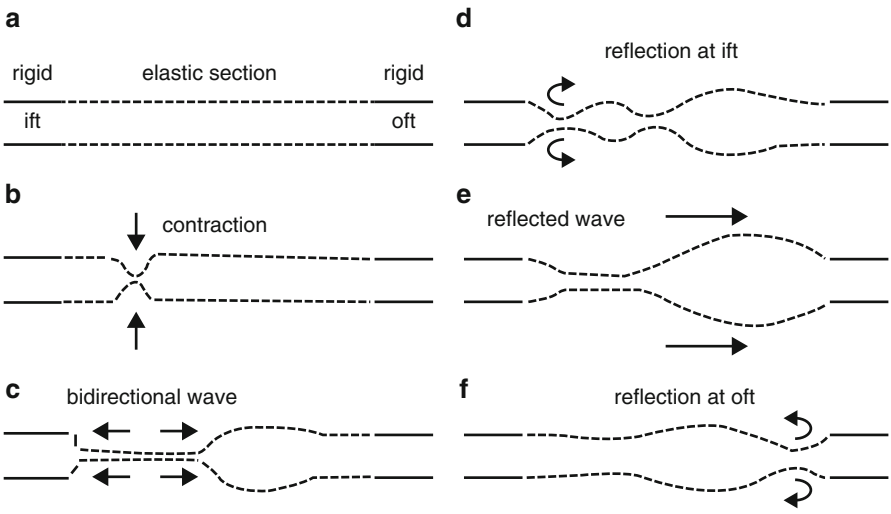


Fig. 1 Schematic diagram illustrating dynamic suction pumping (DSP) [50]. (**a**) The flexible tube is at rest. (**b**) Active contraction of the tube in a noncentral location along the tube. (**c**) Contraction induces an elastic passive bidirectional wave to propagate along the tube. (**d**) Wave reflects off rigid portion of the tube on side nearest to contraction point. (**e**) The reflected wave travels down the tube. (**f**) The waves reflect off the rigid section at the far side of the tube. Notice that the reflected wave amplitude is smaller than the reflected wave off the other end

The vertebrate embryonic heart is a valveless tube, similar to those in various invertebrates, such as urochordates and cephalochordates [27, 47]. Historically, the pumping mechanism in these hearts has been described as peristalsis [27, 50]. More recently, DSP has been proposed as a novel cardiac pumping mechanism for the vertebrate embryonic heart by Kenner et al. in [26] and was later declared the main pumping mechanism in vertebrate embryonic hearts by Fourhar et al. in [14]. Debate over which is the actual pumping mechanism of the embryonic heart continues today, with the possibility that the mechanism may vary between species or may be some hybrid of both mechanisms [34, 53].

Although the size of the blood cells during the tubular heart stage is on the same order of magnitude as the tube itself, previous work with numerical, analytical, and physical have not considered their presence. Given their size ($d \approx 4\ \mu m$) and volume fraction (hematocrit) that ranges from 0% to 40%, it is likely that the blood cells are having some effect on the flow. When the first coordinated myocardial contractions begin to drive blood flow, the embryonic blood lacks blood cells. However, as the heart tube stage progresses, the hematocrit (the volume fraction of blood cells) becomes present, as seen in Fig. 2, and increases linearly during development [1]. Hematocrit may play a role in the distribution of forces along the endothelial lining that contribute to the shaping and growth of the heart.

The purpose of this paper is to explore the performance of DSP and peristalsis when blood cells are added to the flow. In particular, a central goal is to quantify the relationship between the magnitude of flow and the hematocrit in tubular hearts over a range of Womersley numbers, *Wo*. While the vertebrate tubular heart is on the order of tens of microns ($Wo < 1$) [6], the tubular hearts of many invertebrates span from tens to hundreds of microns ($Wo < 1$), e.g., sea squirts, to salps hearts on the order of millimeters ($Wo > 1$) [27]. These ranges of *Wo* naturally lend themselves to numerical study via the immersed boundary method (IBM). The particular geometry for the computational models will be based upon experimental data from zebrafish, *Danio rerio*, embryonic tubular hearts.

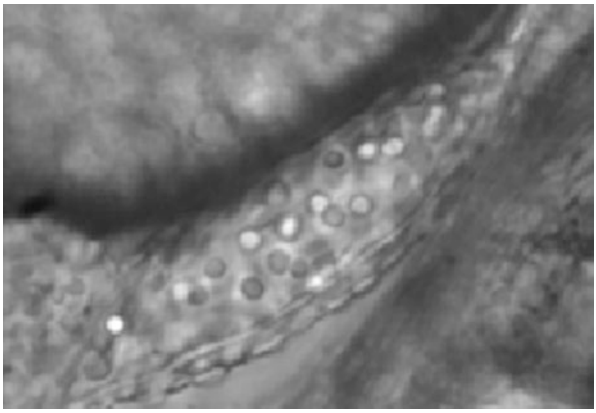


Fig. 2 The embryonic heart tube of a zebrafish 30 hpf courtesy of [32]. Spherical blood cells are seen within the tubular heart. The heart tube is roughly 5 blood cells thick in diameter

2 The Immersed Boundary Method

The IBM is a numerical method developed to solve viscous incompressible fluid dynamic problems with an immersed elastic structure [41, 45]. Since its development in the 1970s by Charles Peskin [44], it has been applied to a wide spectrum of biomathematical models, ranging from blood flow through the heart [44, 45], aquatic locomotion [21], and insect flight [23, 38, 39] to plant biomechanics [40, 55].

The power of this method is that it can be used to describe flow around complicated time-dependent geometries using a regular Cartesian discretization of the fluid domain. The elastic fibers describing the structure are discretized on a moving curvilinear mesh defined in the Lagrangian frame. The fluid and elastic fibers constitute a coupled system, in which the structure moves at the local fluid velocity and the structure applies a singular force of delta-layered thickness to the fluid.

We used an adaptive and parallelized version of the immersed boundary method, immersed boundary method with adaptive mesh refinement (IBAMR) [16, 18]. IBAMR is a C++ framework that provides discretization and solver infrastructure for partial differential equations on block-structured locally refined Eulerian grids [7, 8] and on Lagrangian (structural) meshes. Adaptive mesh refinement (AMR) allows for better resolved dynamics between the fibers and the fluid by increasing grid resolution in areas of the domain that contain an immersed structure or where the vorticity exceeds some threshold. AMR also improves computational efficiency by decreasing grid resolution in areas of the domain that do not require a high level of resolution. The AMR does not occur in every time-step, but rather every n time-steps for efficiency, where n is given as input for the code. IBAMR also includes infrastructure for coupling Eulerian and Lagrangian representations of the fluid and structure, respectively.

The Eulerian grid on which the Navier–Stokes equations were solved was locally refined near the immersed boundaries and regions of vorticity with a threshold of $|\omega| > 0.20$. This Cartesian grid was organized as a hierarchy of four nested grid levels, and the finest grid was assigned a spatial step size of $dx = D/1024$, where D is the length of the domain. The ratio of the spatial step size on each grid relative to the next coarsest grid was 1:4. The temporal resolution was varied to ensure stability. Each Lagrangian point of the immersed structure was chosen to be $\frac{D}{2048}$ apart, that is, twice the resolution of the finest fluid grid.

2.1 Equations of the Immersed Boundary Method

Assume that the immersed boundary is described on a curvilinear, Lagrangian mesh, S, that is free to move. The fluid is described on a fixed Cartesian, Eulerian grid, Ω, that has periodic boundary conditions. Given the size of the domain and the localization of the flow to the tube, the boundary conditions do not significantly

affect the fluid motion. The governing equations for the fluid, the Navier–Stokes equations, are given by:

$$\rho\left[\frac{\partial \mathbf{u}}{\partial t}(\mathbf{x},t)+\mathbf{u}(\mathbf{x},t)\cdot\nabla\mathbf{u}(\mathbf{x},t)\right]=-\nabla p(\mathbf{x},t)+\mu\Delta\mathbf{u}(\mathbf{x},t)+\mathbf{f}(\mathbf{x},t) \tag{1}$$

$$\nabla\cdot\mathbf{u}(\mathbf{x},t)=0. \tag{2}$$

Equations (1) and (2) are the Navier–Stokes equations written in Eulerian form, where Eq. (1) is the conservation of momentum for a fluid and Eq. (2) is the conservation of mass, i.e., incompressibility condition. The two constant parameters in these equations are the fluid density, ρ, and the dynamic viscosity of the fluid, μ. The fluid velocity, $\mathbf{u}(\mathbf{x},t)$, pressure, $p(\mathbf{x},t)$, and body force, $\mathbf{f}(\mathbf{x},t)$, are unknown spatial time-dependent functions of the Eulerian coordinate, $\mathbf{x}$, and time, t. The body force describes the transfer of momentum onto the fluid due to the restoring forces arising from deformations of the elastic structure. It is this term, $\mathbf{f}(\mathbf{x},t)$, that is unique to the particular model being studied.

The material properties of the structure may be modeled to resist to bending, stretching, and displacement from a tethered position. Other forces that have been modeled include the action of virtual muscles, electrostatic (contact) forces, molecular bonds, and other external forces [13, 36, 45, 52]. The immersed structure may deform due to bending forces and/or stretching and compression forces. In this paper, elastic forces are calculated as beams that may undergo large deformations and Hookean springs, i.e.,

$$\mathbf{F}_{beam}=-k_{beam}\frac{\partial^4}{\partial s^4}\Big(\mathbf{X}(s,t)-\mathbf{X}_B(s)\Big) \tag{3}$$

$$\mathbf{F}_{spring}=-k_{spring}\left(1-\frac{R_L}{||\mathbf{X}_S-\mathbf{X}_M||}\right)\cdot(\mathbf{X}_M-\mathbf{X}_S)\,, \tag{4}$$

Equation (3) is the beam equation, which describes forces arising from bending of the elastic structure. Equation (4) describes the force generated from stretching and compression of the structure. The parameters, k_{beam} and k_{spring}, are the stiffness coefficients of the beam and spring, respectively, and R_L is the resting length of the Hookean spring. The variables $\mathbf{X}_M$ and $\mathbf{X}_S$ give the positions in Cartesian coordinates of the master and slave nodes in the spring formulations, respectively, $\mathbf{X}_B(s)$ describes the deviation from the preferred curvature of the structure. In all simulations, $\mathbf{X}_B(s)=0$ along the straight portion of the tube.

A target point formulation can be used to tether the structure or subset thereof in place, holding the Lagrangian mesh in a preferred position that may be time dependent. An immersed boundary point with position $\mathbf{X}(s,t)$ that is tethered to a target point with position $\mathbf{Y}(s,t)$ undergoes a penalty force that is proportional to the displacement between them. The force that results is given by the equation for a linear spring with zero resting length,

$$\mathbf{F}_{target}=-\kappa_{target}\left(\mathbf{X}(s,t)-\mathbf{Y}(s,t)\right), \tag{5}$$

where k_{target} is the stiffness coefficient of the target point springs. k_{target} can be varied to control the deviation allowed between the actual location of the boundary and its preferred position. The total deformation force that will be applied to the fluid is a sum of the above forces,

$$\mathbf{F}(s,t) = \mathbf{F}_{spring} + \mathbf{F}_{beam} + \mathbf{F}_{target} \tag{6}$$

Once the total force from Eq. (6) has been calculated, it needs to be spread from the Lagrangian frame to the Eulerian grid. This is achieved through an integral transform with a delta function kernel,

$$\mathbf{f}(\mathbf{x},t) = \int \mathbf{F}(s,t)\delta(\mathbf{x} - \mathbf{X}(s,t))ds. \tag{7}$$

Similarly, to interpolate the local fluid velocity onto the Lagrangian mesh, the same delta function transform is used,

$$\mathbf{U}(s,t) = \frac{\partial \mathbf{X}}{\partial t}(s,t) = \int \mathbf{u}(\mathbf{x},t)\delta(\mathbf{x} - \mathbf{X}(s,t))d\mathbf{x}. \tag{8}$$

Equations (7) and (8) describe the coupling between the immersed boundary and the fluid, e.g., the communication between the Lagrangian framework and Eulerian framework. The delta functions in these equations make up the heart of the IBM, as they are used to spread and interpolate dynamic quantities between the fluid grid and elastic structure, e.g., forces and velocity. The quantity $\mathbf{X}(s,t)$ gives the position in Cartesian coordinates of the elastic structure at local material point, s, and time t. In approximating these integral transforms, a discretized and regularized delta function, $\delta_h(\mathbf{x})$ [45], is used,

$$\delta_h(\mathbf{x}) = \frac{1}{h^2}\phi\left(\frac{x}{h}\right)\phi\left(\frac{y}{h}\right), \tag{9}$$

where $\phi(r)$ is defined as

$$\phi(r) = \begin{cases} \frac{1}{4}\left[1 + \cos\left(\frac{\pi r}{2}\right)\right] & |r| \leq 2 \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

2.2 Numerical Algorithm

As stated above, we impose periodic boundary conditions on the rectangular domain. To solve Eqs. (1), (2), (7), and (8), we need to update the velocity, pressure, position of the boundary, and force acting on the boundary at time $n+1$ using data from time n. IBM does this in the following steps [45], with an additional step (4*b*) for IBAMR [15, 17]:

Step 1:
Find the force density, $\mathbf{F}^n$, on the immersed boundary, from the current boundary configuration, $\mathbf{X}^n$.

Step 2:
Use Eq. (7) to spread this boundary force from the curvilinear mesh to nearby fluid lattice points.

Step 3:
Solve the Navier–Stokes equations, Eqs. (1) and (2), on the Eulerian domain. In doing so, we are updating $\mathbf{u}^{n+1}$ and p^{n+1} from $\mathbf{u}^n$ and $\mathbf{f}^n$. Note: because of the periodic boundary conditions on our computational domain, we can easily use the Fast Fourier Transform (FFT) [11, 46], to solve for these updates at an accelerated rate.

Step 4:

4a. Update the material positions, $\mathbf{X}^{n+1}$, using the local fluid velocities, $\mathbf{U}^{n+1}$, using $\mathbf{u}^{n+1}$ and Eq. (8).

4b. Refine Eulerian grid in areas of the domain that contain an immersed structure or where the vorticity exceeds a predetermined threshold, if on a selected time-step for AMR.

The above steps outline the process used by the IBM to update the positions and velocities of both the fluid and elastic structure. We note that since we are using IBAMR additional steps are used for AMR. A more detailed discussion of IBM and IBAMR is found in [45] and [16], respectively.

2.3 Model Geometry

We numerically model a $2D$ closed racetrack where the walls of the tube are modeled as $1D$ fibers. The closed tube is composed of two straight portions, of equal length, connected by two half circles, of equal inner and equal outer radii. The tube, or racetrack, has uniform diameter throughout. The geometry of the racetrack is given in Fig. 3.

This study goes beyond previous work [5, 6, 25] through the addition of deformable blood cells, composed of springs connecting adjacent and opposite side Lagrangian nodes. The blood cells are modeled circular, in agreement with in vivo imaging illustrating their spherical geometry in embryonic blood [32], rather than biconcave [12].

All of the mock blood cells in our simulations have the same radii. The diameter of the blood cells was set to $d/5$ [31]. The flexible cells were modeled via attaching springs between adjacent Lagrangian points for each cell, i.e., beams and target points are not used. The geometry of the heart tube with mock blood cells is illustrated in Fig. 3 with all parameter values listed in Table 1. It is important to note that everywhere within our rectangular domain, the fluid has *constant* density ρ and viscosity μ, even within our elastic structures.

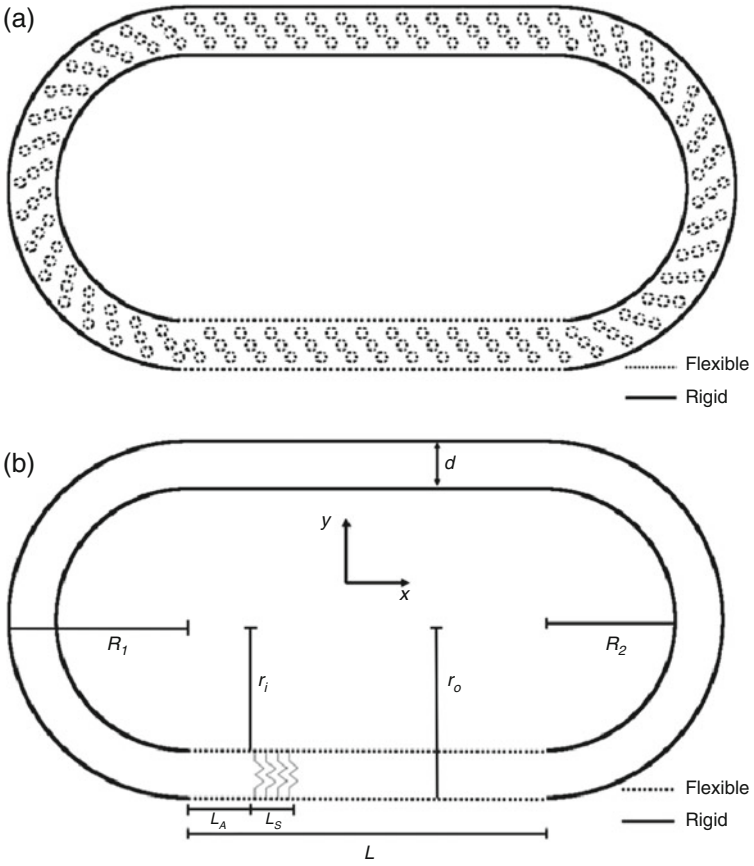


Fig. 3 (**a**) The racetrack geometry, which is held rigid except for the bottom of the tube which is flexible. It also includes flexible blood cells, here illustrating the initial position for a volume fraction of 15%. (**b**) The geometrical features of the racetrack

2.3.1 Dynamic Suction Pumping Model

In the DSP model, the straight portion on the bottom of the racetrack geometry is flexible, e.g., it is composed of beams and springs and is not tethered to target points. All other sides of the tube are held nearly rigid in a fixed position using target points, as well as springs and beams. There are also springs attached over a finite actuation region from the inner to outer boundary in the bottom elastic section of the tube. These springs are used to actuate the tube, modeling DSP. We model the action of "muscles" with linear springs, whose resting lengths change in time. These springs are attached between the inner and outer Lagrangian boundaries of the heart tube.

Table 1 Geometric parameters used in the numerical experiments

Parameter	Value
d	1
$R_1 = r_o$	3.75
$R_2 = r_i$	2.75
L	7.5
L_A	0.9375
L_S	0.75
r_C	0.1

d is the diameter of the tube, R_1 and r_o give the outer radius (or distance from the centerline) of the tube, R_2 and r_i give the inner radius (or distance from the centerline) of the tube, L is the length of the flexible section for DSP and contractile wave section for peristalsis, L_A is the length of straight tube before the actuation section for DSP, L_S is the size of the actuation section for DSP, and r_C is the radii of a blood cell

Resistance to stretching is included in the tethered portion of the tube and in the sections with preferred motion to reduce high frequency oscillations in the boundary. Small bending resistance is added to the simulations to: (1) better approximate heart tubes that have some resistance to bending, (2) reduce high frequency oscillations in the tethered portion of the tube, (3) smooth the transition from the flexible portion of the tube to the tethered portion, e.g., smooth the connection points, and (4) eliminate any kinks in the elastic section of the tube.

Rather than attaching these muscles between all points within this region, we choose a region that is 10% of the length of the flat portion, $L_S = L/10$, which is also translated at a distance of $L_A = L/8$ along the tube from the beginning of the flat portion from the left. This model was selected since traditional DSP only assumes an off-center region of active contraction. The resting lengths of these springs were changed according to:

$$R_L(s,t) = d\Big(1 - \frac{8.5}{10}\Big|\sin(2.3\pi t)\Big|\Big) \tag{11}$$

2.3.2 Peristalsis Model

A prescribed motion of the actuation region along the bottom straight portion of the tube is used to drive peristalsis. To permit volume conservation in the closed racetrack, the top straight section of the racetrack is modeled using springs and beams and is allowed to expand. The reasons for using both springs and beams, which allow for stretching and bending, respectively, here are the same as in the DSP model.

The rest of the racetrack geometry composed of target points is held nearly rigid, similarly to Sect. 2.3.1. There are also springs connecting the outer and inner layer of the top of the tube for additional support. The peristaltic wave of contraction is prescribed by interpolating between multiple positions as described below.

Phase 1 is defined by the position of the relaxed, straight tube. Phase 2 is defined as a fully pinched tube at the initial position of contraction. Phase 3 is defined as a full pinched tube at the end of the peristaltic wave. The initial contraction (pinching) of the tube was prescribed by interpolating between Phase 1 and Phase 2. Similarly, the contractile release was performed by interpolating back between Phase 3 and Phase 1. This is illustrated in Fig. 4. The traveling wave of contraction was performed by translating the pinch along the length of the contractile section of the tube.

The motion of the actual immersed boundary is driven by changing the position of the target points, which are tethered to each immersed boundary point along the racetrack. The times of each phase (contraction, translation, and relaxation) are seen in Table 2, where T is the period of one complete peristaltic wave. The following function was used to induce the traveling peristaltic wave between Phases 2 and 3,

$$X_{target} = \begin{cases} \pm\tilde{A}(x - x_L(t))^2 (x_R(t) - x)^2 \exp -\frac{(x - x_C(t))^2}{(0.5w)^2} - R_{o/i}, & x \in [x_L(t), x_R(t)] \\ 0 & \text{elsewhere} \end{cases}, \tag{12}$$

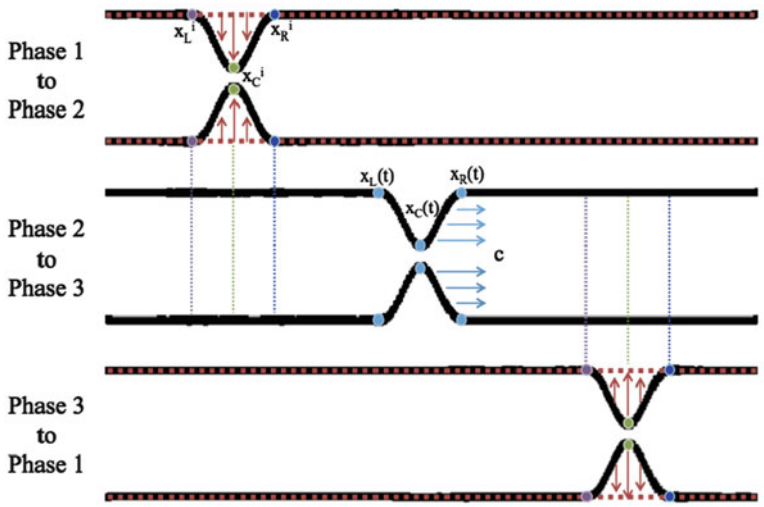


Fig. 4 Interpolation phases for the traveling contraction wave along the bottom portion of the racetrack geometry. From Phase 1 (*straight red tube*) to Phase 2, the tube gets pinched on the left side. From Phase 2 to Phase 3, the occlusive pinch travels down the tube at speed c. From Phase 3, the pinch is released and goes back to the straight tube (*in red*)

Table 2 Table of temporal parameters for the prescribed peristaltic wave

Parameter	Time
T	0.435
T_1	$0.025 \times T$
T_2	$0.95 \times T$
T_3	$0.025 \times T$

T is the nondimensional period. T_1 and T_3 gives the nondimensional period for the initial pinching and release of the tube. T_3 gives the nondimensional translation time of the peristaltic wave

and

$$x_L(t) = x_L^i + c(t - T_1),$$
$$x_C(t) = x_C^i + c(t - T_1),$$
$$x_R(t) = x_R^i + c(t - T_1),$$
$$c = -\frac{2x_C^i}{T_2},$$
$$\tilde{A} = 850.0,$$

where x_L^i, x_C^i, and x_R^i are the left-most, center, and right-most points associated with the first pinch. These points are illustrated in Fig. 4. The parameters c and $\tilde{A}$ are the wave speed and amplitude, respectively. $\pm\tilde{A}$ and $R_{o/i}$ correspond to the bottom and the top wall of the tube, respectively.

2.3.3 Determining Biologically Relevant Parameter Values

To determine the lower range of *Wo* within the heart tube, we take characteristic values for zebrafish embryonic hearts between 26 and 30 hpf and match our nondimensionless model parameters accordingly. The characteristic frequency, f_{zf}, was measured in vivo, and the characteristic length, L_{zf}, was taken as the diameter of the heart tube. The *Wo* was then calculated as:

$$Wo = L_{zf}\sqrt{\frac{2\pi \cdot f_{zf} \cdot \rho_{zf}}{\mu_{zf}}} = 0.15, \tag{13}$$

where $f_{zf} = 2.2\,\mathrm{s}^{-1}$ [33], $\rho_{zf} = 1025\,\mathrm{kg/m}^3$ [50], $\mu_{zf} = 0.0015\,\mathrm{kg/(m\,s)}$ [42], and $L_{zf} = 0.05\,\mathrm{mm}$ [6]. The occlusion ratio is assumed to be $occ = 0.85$ [31]. We take the characteristic velocity to be $V_{pump} = f_{zf} \cdot occ \cdot \frac{L_{zf}}{2} = 0.047\,\mathrm{mm/s}$. The dimensionless frequency may then be calculated as:

$$\tilde{f} = \frac{L_{zf}}{V_{pump}} \cdot f_{zf} = 2.3. \tag{14}$$

For the mathematical model, the parameter values were chosen to keep the dimensionless frequency fixed at $f_{sim} = 2.3$, and hence we get the dimensionless pumping velocity of $V_{sim} = f_{sim} \cdot occ \cdot \frac{d}{2} = 0.978$.

The *Wo* was varied by changing the dynamic viscosity, μ. For the simulations, the Wo_{sim} is calculated using a characteristic length of d, the width of the tube, and dimensionless density of $\rho = 1000$. The simulations were performed for $Wo_{sim} = \{0.2, 0.4, 0.6, 0.8, 0.9, 1.0, 1.5, 2, \ldots, 9, 10, 15, 20, 30\}$. Note that the higher end of these values describes a fully inertial regime which may be outside of what is found in nature. The stiffness of the target tethering points was chosen to minimize the deformations of the boundary, i.e., to keep it rigid, and was directly correlated to *Wo*. The motivation for the wide range of *Wo* considered is that we want to compare parameter values relevant to other types of tubular hearts, such as salps, tunicates, and insects. We also want to compare our results to the *Wo* range considered in most previous DSP studies (typically, $Wo > 1$). The other mechanical parameters were chosen to allow deformation and reexpansion of the heart tube on relevant timescales. Our parameter choices are given in Table 3, where they have been nondimensionalized, using the following relations for springs (and target points) and beams, respectively,

$$\tilde{k}_{spring/target} = \frac{k_{spring/target}}{\rho * d * V_{sim}^2} \tag{15}$$

$$\tilde{k}_{beam} = \frac{k_{beam}}{\rho * V_{sim}^2 * d^3}. \tag{16}$$

Note that:

- The top portion of the tube in the peristaltic cases is allowed to be flexible to conserve volume (e.g., this section expands as the tube compresses). This is not necessary in the case of DSP because the lower portion of the tube is flexible and can expand to conserve volume.
- In the case of peristalsis, the entire motion of the boundary is moved with a preferred motion (with the exception of the top which simply expands to conserve volume). In this setup, the elastic forces only serve to minimize deviations from that preferred motion and to reduce high frequency oscillations. As long as the actual motion is sufficiently close to the preferred motion, the parameters chosen do not alter the results.
- In the case of DSP, the stretching stiffness of the bottom, untethered portion of the tube, is a primary factor that influences the average velocity. We selected this value to give reasonable bulk flow at higher *Wo*. In future studies, we will characterize how this stiffness affects average flow.

All of the reported values have been nondimensionalized according to Eq. (15), (16). The viscosity may be used to change the *Wo* without affecting the dimensionless stiffnesses. In other words, viscosity can be used to vary the *Wo* without significantly effecting the dynamics of the structure. Note that the *Wo* considered fall within the range of many biological pumps, see [50].

Table 3 Table of mechanical parameters used in the computational model

Mechanical parameters	Symbol	DSP value	Peristalsis value
Stretching stiffness of the tube	$\tilde{k}_{s_{tube}}$	$1.3e7$	$3.3e6$
Stretching stiffness of springs across tube	$\tilde{k}_{s_{btwn}}$	$2.3e3$	$6.5e2$
Stretching stiffness of target points	$\tilde{k}_{target}$	$1.3e5$	$1.3e5$
Bending coefficient of the tube	$\tilde{k}_{beam}$	$3.3e2$	$3.3e10$
Stretching stiffness of blood cells	$\tilde{k}_{s_{cell}}$	$2.8e5$	$2.8e5$

Note that $\tilde{k}_{s_{btwn}}$ gives the stiffness coefficient of the actuating springs in the DSP model, while it describes the stiffness coefficients of springs connecting the outer and inner layer of the top of the tube in the peristalsis model. $\tilde{k}_{s_{tube}}$ is the stretching stiffness between adjacent points along the tube while $\tilde{k}_{s_{btwn}}$ gives the stretching stiffness between opposite points along a cross-section of the tube. $\tilde{k}_{target}$ is the tethering stiffness of target points, $\tilde{k}_{beam}$ gives the bending stiffness between adjacent points along the tube, and $\tilde{k}_{s_{cell}}$ gives the stretching stiffness between points making up the blood cells

3 Results

In this paper, we present simulations of DSP and peristalsis within a closed racetrack containing flexible blood cells of varying volume fractions but uniform geometry. The simulations were run for a range of Womersley number, $Wo \in [0.1, 30]$, and hematocrit, $VF = \{0\%, 5\%, 10\%, 15\%, 20\%, 25\%\}$. Examples of the locations of the blood cells and boundaries at different points in time are seen in Fig. 5.

3.1 Dynamic Suction Pumping Results

Figure 5 shows snapshots from simulations of DSP at five different Womersley numbers, $Wo = \{0.2, 2.0, 6.0, 10.0, 20.0\}$, where volume fraction is held constant at $VF = 15\%$. The images were taken after at $11.5, 22.5, 33.5$, and 44.5 heartbeats. In the cases of $Wo = 0.2$ thru $Wo = 2.0$, there is no significant net transport for the mock blood cells as evidenced by the negligible movement of the blood cells (note the color coding of blood cells in each quadrant). There is, however, clear transport when $Wo \geq 6.0$. Moreover, in the cases when $Wo \geq 6.0$, the blood cells begin to clump together, rather than move uniformly throughout the tube.

Keeping the volume fraction constant, we compared the spatially averaged velocity across a cross-section in the center of the top of the tube, for three different Womersley numbers, $Wo \in \{0.2, 2.0, 6.0, 10.0, 20.0\}$. Note that deformations of this section of the tube are negligible such that the average velocity is directly proportional to the volumetric flow rate. Figure 6 illustrates this for the case of $VF = 15\%$. From the figure, it is evident that the lower Wo case induces less net flow than the other two higher Wo cases. However, we can also deduce that the direction of flow is a nonlinear function of Wo. Note that for $Wo = 10.0$, flow is moving in the opposite direction to that of the $Wo = 20.0$ case.

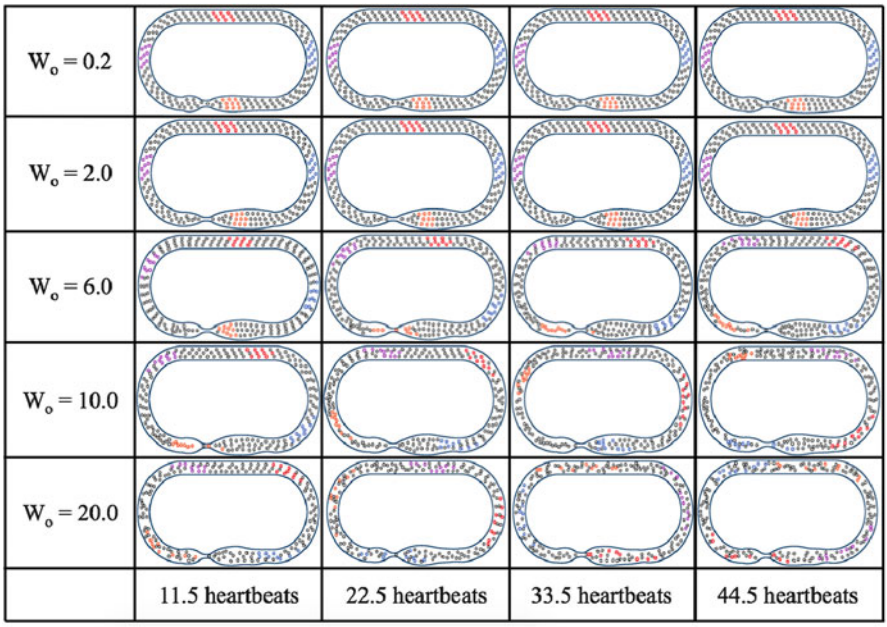


Fig. 5 A comparison of simulations with different Womersley number, $Wo = \{0.2, 2.0, 6.0, 10.0, 20.0\}$, but same amount of blood cells, $VF = 15\%$. The images were taken after at 11.5, 22.5, 33.5, and 44.5 heartbeats during the simulations. In the case of $Wo = 0.2$ and $Wo = 2.0$, there is no visual transport for the mock blood cells; however, there is clear transport when $Wo \geq 6.0$. Moreover, in the cases when $Wo \geq 6.0$, the hematocrit begins to clump together, rather than move uniformly throughout the tube

This study has considered two orders of magnitude in *Wo*. Given that *Wo* scales like the square root of the *Re*, note that our study spans four orders of magnitude in *Re*. For the range of *Wo* considered, higher volume fractions result only in flow in the negative direction. It would be interesting to see if this is the case for all *Wo*. Future studies will consider a larger and more detailed range of *Wo* to better resolve the changes in flow direction.

Moreover, an example comparison of the spatially averaged velocity vs. time for three different volume fractions, $VF = \{5\%, 15\%, 25\%\}$, for three specific Womersley numbers, $Wo = \{0.2, 2.0, 6.0\}$, is shown in Fig. 7. Figure 7a,b shows the similarity of the waveforms illustrating little effect of blood cells on bulk flow patterns, for all three selected *Wo*. Figure 7c–e gives the spatially averaged velocities (in diameters/heartbeat) vs. time over the course of the simulation. Time is given in number of heartbeats. In the $Wo = 0.2$ case, the average velocities asymptotically increase until they reach a periodic cycle. It is clear that the maximal flow rates in both the $Wo = 0.2$ and $Wo = 2.0$ cases are multiple orders of magnitude below one diameter/heartbeat. These are well below the experimentally observed velocity of $\sim$0.9 diameters/heartbeat recorded in zebrafish [14].

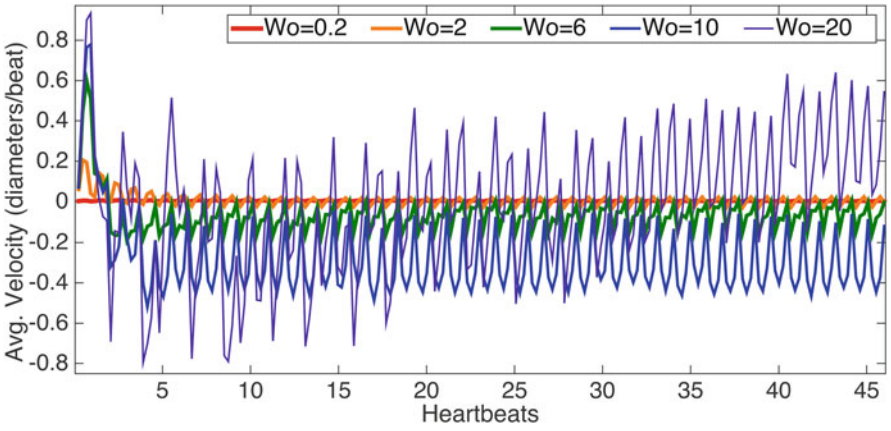


Fig. 6 A comparison of the spatially averaged velocity vs. time over the course of the simulation, for five cases with uniform hematocrit ($VF = 15\%$), but varying Womersley number, $Wo = \{0.2, 2, 6, 10, 20\}$. The average velocity was spatially computed across a cross-section in the center of the top of the tube. As Wo increases, the amplitude of oscillations in average velocity also increases. In the biologically relevant case, $Wo = 0.2$, there are slight oscillations; however, bulk net flow is insignificant

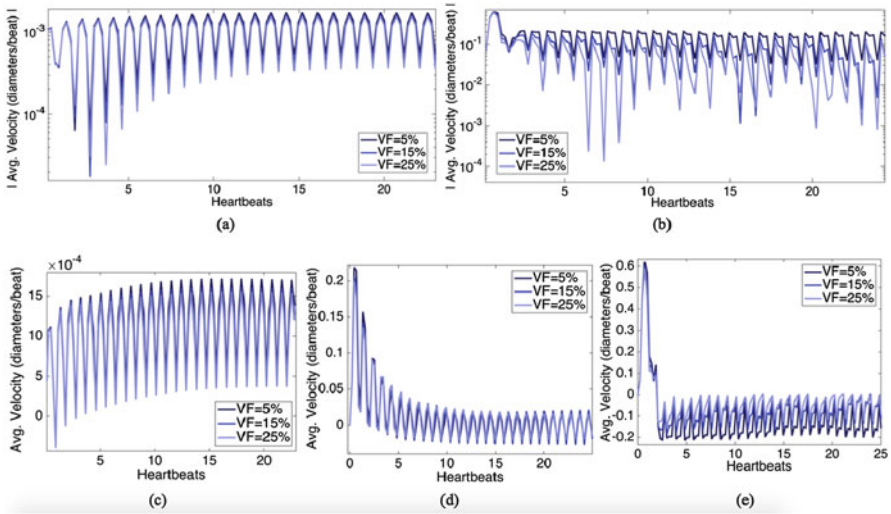


Fig. 7 A comparison of the spatially averaged velocity vs. time over the course of the simulation, for varying hematocrit, $VF = \{5\%, 15\%, 25\%\}$, for three different Womersley numbers, $Wo = 0.2$ (**a**,**c**), $Wo = 2.0$ (**d**), and $Wo = 6.0$ (**b**,**e**). The average velocity was spatially computed across a cross-section in the center of the top of the tube. (**a**,**b**) illustrate how similar the waveforms are for varying volume fractions of Wo, 0.2 and 6.0, respectively. (**c**,**d**,**e**) give the average velocities, in diameters/heartbeat, over the course of the simulation, in heartbeats

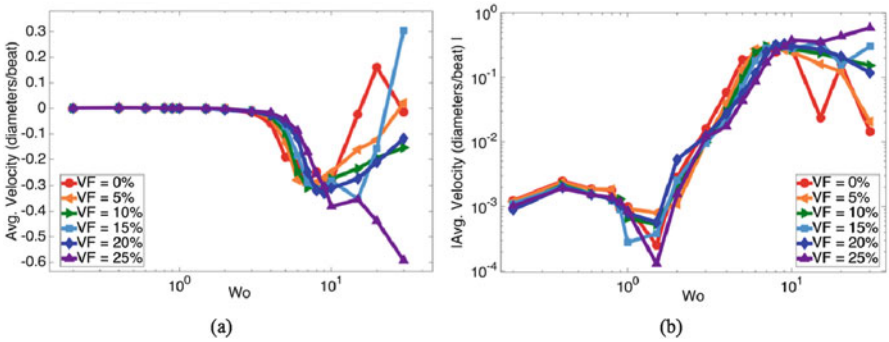


Fig. 8 (**a**) shows the spatially and temporally averaged velocity for each simulation vs. Womersley number for a hematocrit range of $[0\%, 25\%]$. (**b**) shows the spatially and temporally averaged magnitude of velocity vs. *Wo* for a hematocrit range of $[0\%, 25\%]$

To quantify the effect of blood cells further, spatially and temporally averaged velocities for various *Wo* and hematocrits were compared. This is illustrated in Fig. 8. From Fig. 10a, it is clear that flow rates are a nonlinear function of *Wo*. The case with zero hematocrit is in agreement with previous results reported in [6]. Moreover, the addition of hematocrit does not significantly perturb flow rates for $Wo \lesssim 10$, as seen in Fig. 10b, except for the case where $Wo = 1.5$. In the case of $Wo = 1.5$, the addition of hematocrit affects flow rates; however, absolute bulk flow rates are minimal over the range of *Wo* considered. Furthermore, for $Wo \gtrsim 10$, the addition of hematocrit affects flow rates in a nonlinear fashion.

3.2 Peristalsis Results

Figure 9 shows snapshots from simulations for two different Womersley numbers, $Wo = \{0.2, 2.0\}$, where hematocrit is held constant at $VF = 15\%$. The images were taken after 1.5, 2.5, 3.5, and 4.5 heart beats during the simulations. It is clear from both simulations that there is significant bulk flow throughout the tube. Moreover, significant mixing is observed in both cases. Note that although the top portion of the racetrack is elastic, it has a relatively high bending rigidity. The top of the tube expands just enough to compensate for the volume of the bottom of the tube that is lost due to the presence of the peristaltic wave.

The volume fraction was kept constant, at $VF = 15\%$, in Fig. 10a to explore the effect of scaling on bulk flow for $Wo = 0.2, 2.0, 20.0$. The spatially averaged velocity across the top of the tube, given in diameters/heartbeat, is similar between all three cases of *Wo*. Furthermore the direction of flow is consistent in all cases, with bulk flow moving counterclockwise around the tube, with a sharp decrease in velocity, showing flow going in the opposite direction, at the end of each heartbeat.

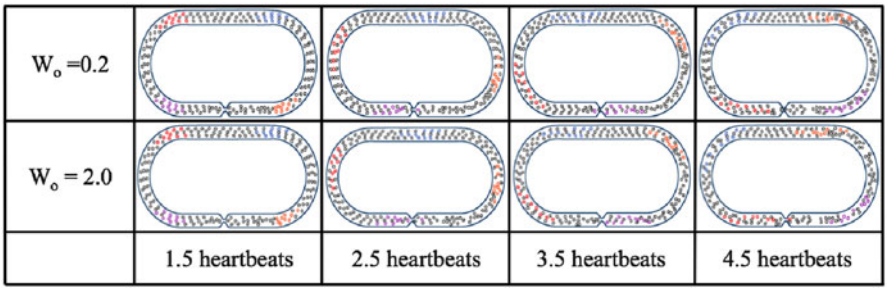


Fig. 9 A comparison of simulations for two different Womersley numbers, $Wo = \{0.2, 2.0\}$, but same hematocrit, $VF = 15\%$. The images were taken after at $1.5, 2.5, 3.5$, and 4.5 heartbeats during the simulations. It is clear that there is significant mixing of the blood cells with peristalsis, as the colored sections begin to mix

Figure 10b and c illustrates the effect of varying hematocrit for simulations with $Wo = 0.2$ and $Wo = 20$, respectively. In both cases, the waveforms look similar, suggesting that the addition of blood cells does not significantly affect bulk flow rates. However, we note that the sharp decrease in velocity at the end of the heartbeat is more pronounced in the $Wo = 0.2$ case, than in the $Wo = 20$ case.

4 Conclusions

In this paper, two-dimensional immersed boundary simulations were used to model DSP and peristalsis for a single actuation frequency over a range of Womersley numbers and hematocrits relevant to valveless, tubular hearts. When strong net flow was generated in the tube at higher Wo, blood cells clumped together and did not flow uniformly throughout the tube. The spatially and temporally averaged velocities across a cross-section along the top of the tube showed a nonlinear relationship between net flow rates and Wo for DSP. The effect of hematocrit on the net flow rate was significant for $Wo \gtrsim 10$ and was nonlinear. In particular, the varying levels of hematocrit changed the direction of flow for DSP for Wo on the order of 10. The addition of blood cells did not enhance the weak net flows produced for $Wo < 1$. These results highlight the complex dynamics governing DSP.

For DSP at low Wo and for the range of tube material properties considered here, the fluid is nearly reversible. This reversibility may explain in part why there is little net flow in the tube for the case of DSP (a reversible motion) at $VF = 0$. This result is in agreement with [5, 6]. Previous studies have shown enhanced fluid transport and animal locomotion in non-Newtonian fluids at low Re and Wo [28]. Since the addition of blood cells in a Newtonian fluid makes the bulk fluid effectively non-Newtonian, it is possible that the addition of blood cells could make the flow in tubular hearts irreversible. For the parameters considered here, any such effect was negligible.

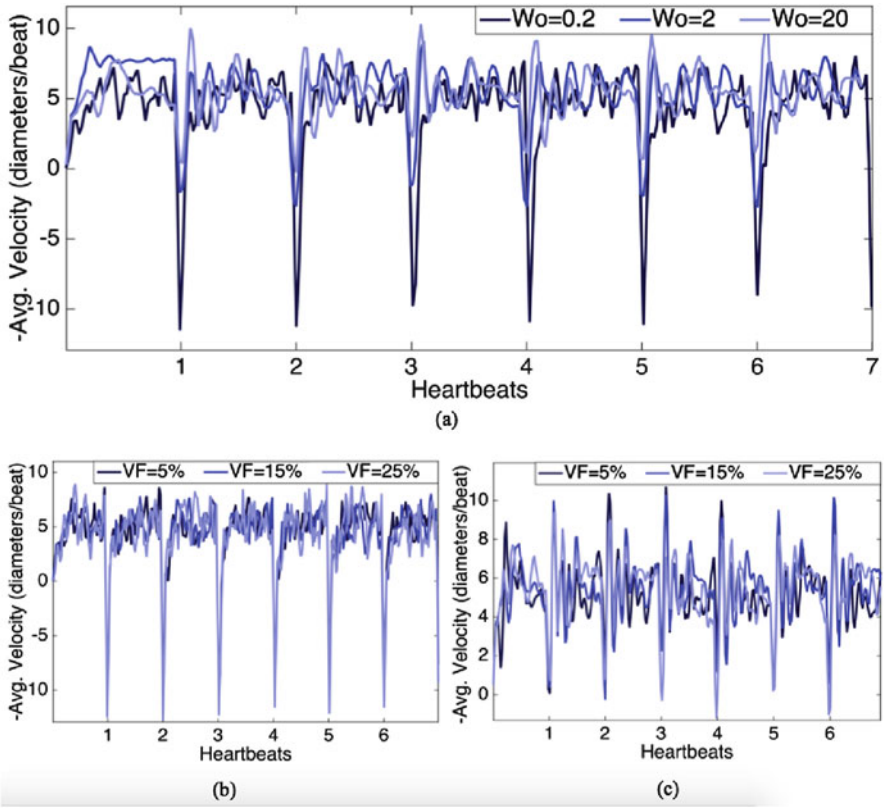


Fig. 10 A comparison of the spatially averaged velocity vs. time over the course of the simulation, for hematocrit, $VF = 15\%$, for three different Womersley numbers, $Wo = 0.2, 2, 20$ is shown in (**a**). The spatially averaged velocity was computed across a cross-section in the center of the top of the tube, given in diameters/heartbeat. (**b**) and (**c**) give the spatially averaged velocities for $Wo = 0.2$ and $Wo = 20$, respectively, for three volume fractions, $VF = 5\%, 15\%, 25\%$

For the case of peristalsis, flow was consistently driven around the racetrack for all *Wo* and for all hematocrits. Similar to DSP, the addition of hematocrit did not significantly change net flow rates at low *Wo*. Unlike the case of DSP, the addition of hematocrit also did not significantly alter the velocity waveform or the net flow at higher *Wo*.

Although the bulk transport of fluid was not significantly changed, the addition of blood cells may affect the shear stresses experienced by the cardiac cells and the amount of mixing within the heart tube. The peristalsis simulations show enhanced mixing as compared to that of DSP at the same *Wo* and *VF*. Furthermore, for $Wo = 0.2, 2$, peristalsis was able achieve similar levels of blood cell mixing at an order of magnitude faster than the DSP simulation at $Wo = 20$. These results are important when considering the role that fluid mixing and shear stress may play in cardiogenesis.

Experimental evidence has shown that blood flow, and more specifically hemodynamic forces, is essential for proper heart morphogenesis [22]. Furthermore, it is evident that there is a strongly coupled relationship between the underlying hemodynamics, cardiac electrophysiology, and activation of some genetic regulatory networks. For example, hemodynamics is thought to regulate the development of the pacemakers and the conduction of action potentials in the heart [48, 51]. Since there is direct feedback between the underlying electrophysiology and the flow induced by muscle contraction, changes in traveling action potentials will affect the hemodynamic forces felt at the endothelial layer, e.g., shear stress and pressure. These changes may then result in changes in gene expression via epigenetic signaling mechanisms, e.g., mechanotransduction. However, the exact pipelines that contribute to mechanotransduction are not completely understood [54].

Acknowledgements The authors would like to thank Steven Vogel for conversations on scaling in various hearts. We would also like to thank Lindsay Waldrop, Austin Baird, Jiandong Liu, Leigh Ann Samsa, and William Kier for discussions on embryonic hearts. This project was funded by NSF DMS CAREER #1151478 awarded to L.A.M. Funding for N.A.B. was provided from an National Institutes of Health T32 grant [HL069768-14; PI, Christopher Mack].

References

1. Al-Roubaie, S., Jahnsen, E.D., Mohammed, M., Henderson-Toth, C., Jones, E.A.: Rheology of embryonic avian blood. Am. J. Physiol. Heart Circ. Physiol. **301**(6919), 2473–2481 (2011)
2. Auerbach, D., Moehring, W., Moser, M.: An analytic approach to the Liebau problem of valveless pumping. Cardiovasc. Eng. Int. J. **4**, 201–207 (2004)
3. Avrahami, I., Gharib., M.: Computational studies of resonance wave pumping in compliant tubes. J. Fluid Mech. **608**, 139–160 (2008)
4. Babbs, C.: Behavior of a viscoelastic valveless pump: a simple theory with experimental validation. BioMed. Eng. Online **9**(42), 19832–19837 (2010)
5. Baird, A.J.: Modeling valveless pumping mechanisms (Ph.D. thesis). Univ. N. C. Chapel Hill **628**, 129–148 (2014)
6. Baird, A.J., King, T., Miller, L.A.: Numerical study of scaling effects in peristalsis and dynamic suction pumping. Biol. Fluid Dyn. Model. Comput. Appl. **628**, 129–148 (2014)
7. Berger, M.J., Oliger, J.: Adaptive mesh refinement for hyperbolic partial-differential equations. J. Comput. Phys. **53**(3), 484–512 (1984)
8. Berger, M.J., Colella, P.: Local adaptive mesh refinement for shock hydrodynamics. J. Comput. Phys. **82**(1), 64–84 (1989)
9. Bringley, T., Childress, S., Vandenberghe, N., Zhang, J.: An experimental investigation and a simple model of a valveless pump. Phys. Fluids **20**, 033,602 (2008)
10. Chang, H.T., Lee, C.Y., Wen, C.Y.: Design and modeling of electromagnetic actuator in MEMS-based valveless impedance pump. Microsyst. Technol. Micro Nanosyst. Inf. Storage Process. Syst. **13**, 1615–1622 (2007)
11. Cooley, J., Tukey, J.W.: An algorithm for the machine calculation of complex Fourier series. Math. Comput. **19**, 297–301 (1965)
12. Crowl, L.M., Fogelson, A.L.: Computational model of whole blood exhibiting lateral platelet motion induced by red blood cells. Int. J. Numer. Methods Biomed. Eng. **26**, 471–487 (2009)
13. Fogelson, A.L., Guy, R.D.: Immersed-boundary-type models of intravascular platelet aggregation. Comput. Methods Appl. Mech. Eng. **197**, 2087–2104 (2008)

14. Forouhar, A.S., Liebling, M., Hickerson, A., Nasiraei-Moghaddam, A., Tsai, H.J., Hove, J.R., Fraser, S.E., Dickinson, M.E., Gharib, M.: The embryonic vertebrate heart tube is a dynamic suction pump. Science **312**(5774), 751–753 (2006)
15. Griffith, B.E.: Simulating the blood-muscle-vale mechanics of the heart by an adaptive and parallel version of the immersed boundary method (Ph.D. thesis). Courant Institute of Mathematics, New York University (2005)
16. Griffith, B.E.: An adaptive and distributed-memory parallel implementation of the immersed boundary (ib) method (2014). URL https://github.com/IBAMR/IBAMR
17. Griffith, B.E., Peskin, C.S.: On the order of accuracy of the immersed boundary method: higher order convergence rates for sufficiently smooth problems. J. Comput. Phys. **208**, 75–105 (2005)
18. Griffith, B.E., Hornung, R., McQueen, D., Peskin, C.S.: An adaptive, formally second order accurate version of the immersed boundary method. J. Comput. Phys. **223**, 10–49 (2007)
19. Hickerson, A.I.: An experimental analysis of the characteristic behaviors of an impedance pump (Ph.D. thesis). Calif. Inst. Technol. **608**, 139–160 (2005)
20. Hickerson, A., Rinderknecht, D., Gharib, M.: Experimental study of the behavior of a valveless impedance pump. Exp. Fluids **38**, 534–540 (2005)
21. Hieber, S., Koumoutsakos, P.: An immersed boundary method for smoothed particle hydrodynamics of self-propelled swimmers. J. Comput. Phys. **227**, 8636–8654 (2008)
22. Hove, J.R., Koster, R.W., Forouhar, A.S., Acevedo-Bolton, G., Fraser, S.E., Gharib, M.: Intracardiac fluid forces are an essential epigenetic factor for embryonic cardiogenesis. Nature **421**(6919), 172–177 (2003)
23. Jones, S.K., Laurenza, R., Hedrick, T.L., Griffith, B.E., Miller, L.A.: Lift- vs. drag-based for vertical force production in the smallest flying insects. J. Theor. Biol. **384**, 105–120 (2015)
24. Jung, E.: Two-dimensional simulations of valveless pumping using the immersed boundary method (Ph.D. thesis). Courant Inst. Math. N. Y. Univ. **608**, 139–160 (1999)
25. Jung, E., Peskin, C.: 2-d simulations of valveless pumping using immersed boundary methods. SIAM J. Sci. Comput. **23**, 19–45 (2001)
26. Kenner, T., Moser, M., Tanev, I., Ono, K.: The Liebau-effect or on the optimal use of energy for the circulation of blood. Scr. Med. **73**, 9–14 (2000)
27. Kriebel, M.E.: Conduction velocity and intracellular action potentials of the tunicate heart. J. Gen. Physiol. **50**, 2097–2107 (1967)
28. Lauga, E.: Propulsion in a viscoelastic fluid. Phys. Fluids **19**, 083,104 (2007)
29. Lee, D.S., Yoon, H.C., Ko, J.S.: Fabrication and characterization of a bidirectional valveless peristaltic micropump and its application to a flow-type immunoanalysis. Sensors Actuators **103**, 409–415 (2004)
30. Lee, C.Y., Chang, H.T., Wen, C.Y.: A MEMS-based valveless impedance pump utilizing electromagnetic actuation. J. Micromech. Microeng. **18**, 225–228 (2008)
31. Maes, F., Chaudhry, B., Ransbeeck, P.V., Verdonck, P.: Visualization and modeling of flow in the embryonic heart. IFMBE Proc. **22**(6919), 1875–1878 (2008)
32. Maes, F., Chaudhry, B., Ransbeeck, P.V., Verdonck, P.: The pumping mechanism of embryonic hearts. IFMBE Proc. **37**, 470–473 (2011)
33. Malone, M., Sciaky, N., Stalheim, L., Klaus, H., Linney, E., Johnson, G.: Laser-scanning velocimetry: A confocal microscopy method for quantitative measurement of cardiovascular performance in zebrafish embryos and larvae. BMC Biotechnol. **7**, 40 (2007)
34. Manner, J., Wessel, A., Yelbuz, T.M.: How does the tubular embryonic heart work? looking for the physical mechanism generating unidirectional blood flow in the valveless embryonic heart tube. Dev. Dyn. **239**, 1035–1046 (2010)
35. Manopoulos, C.G., Mathioulakis, D.S., Tsangaris, S.G.: One-dimensional model of valveless pumping in a closed loop and a numerical solution. Phys. Fluids **18**, 201–207 (2006)
36. Mathur, S.R., Sun, L., Das, S., Murthy, J.Y.: Application of the immersed boundary method to fluid, structure, and electrostatics interaction in MEMS. Numer. Heat Transfer Part B Fundam. Int. J. Comput. Methodol. **62**, 399–418 (2012)
37. Meier, J.: A novel experimental study of a valveless impedance pump for applications at lab-on-chip, microfluidic, and biomedical device size scales (Ph.D. thesis). Calif. Inst. Technol. 8636–8654 (2011)

38. Miller, L.A., Peskin, C.S.: When vortices stick: an aerodynamic transition in tiny insect flight. J. Exp. Biol. **207**, 3073–3088 (2004)
39. Miller, L.A., Peskin, C.S.: A computational fluid dynamics of clap and fling in the smallest insects. J. Exp. Biol. **208**, 3076–3090 (2009)
40. Miller, L.A., Santhanakrishnan, A., Jones, S.K., Hamlet, C., Mertens, K., Zhu, L.: Reconfiguration and the reduction of vortex-induced vibrations in broad leaves. J. Exp. Biol. **215**, 2716–2727 (2012)
41. Mittal, R., Iaccarino, C.: Immersed boundary methods. Annu. Rev. Fluid Mech. **37**, 239–261 (2005)
42. Mohammed, M., Roubaie, S., Jahnsen, E., Jones, E.: Drawing first blood: Measuring avian embryonic blood viscosity. SURE Poster Presentation **61**, 33–45 (2011)
43. Ottesen, J.: Valveless pumping in a fluid-filled closed elastic tube-system: one-dimensional theory with experimental validation. J. Math. Biol. **46**, 309–332 (2003)
44. Peskin, C.: Numerical analysis of blood flow in the heart. J. Comput. Phys. **25**, 220–252 (1977)
45. Peskin, C.S.: The immersed boundary method. Acta Numer. **11**, 479–517 (2002)
46. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Fast Fourier transform. Ch. 12 in Numerical Recipes in FORTRAN: The Art of Scientific Computing **2**, 490–529 (1992)
47. Randall, D.J., Davie, P.S.: The hearts of urochordates and cephalochordates. Comp. Anat. Dev. **1**, 41–59 (1980)
48. Reckova, M., Rosengarten, C., deAlmeida, A., Stanley, C.P., Wessels, A., Gourdie, R.G., Thompson, R.P., Sedmera, D.: Hemodynamics is a key epigenetic factor in development of the cardiac conduction system. Circ. Res. **93**, 77 (2003)
49. Samson, O.: A review of valveless pumping: History, applications, and recent developments (2007). URL http://www.researchgate.net/publication/267300626_A_Review_of_Valveless_Pumping_History_Applications_and_Recent_Developments
50. Santhanakrishnan, A., Miller, L.A.: Fluid dynamics of heart development. Cell Biochem. Biophys. **61**, 1–22 (2011)
51. Tucker, D.C., Snider, C., Woods Jr., W.T.: Pacemaker development in embryonic rat heart cultured *in oculo*. Pediatr. Res. **23**, 637–642 (1988)
52. Tytell, E., Hsu, C., Williams, T., Cohen, A., Fauci, L.: Interactions between internal forces, body stiffness, and fluid environment in a neuromechanical model of lamprey swimming. Proc. Natl. Acad. Sci. U. S. A. **107**, 19832–19837 (2010)
53. Waldrop, L.D., Miller, L.A.: Large-amplitude, short-wave peristalsis and its implications for transport. Biomech. Model. Mechanobiol. **15**, 629–642 (2016)
54. Weinbaum, S., Zhang, X., Han, Y., Vink, H., Cowin, S.: Mechanotransduction and flow across the endothelial glycoalyx. Proc. Natl. Acad. Sci. U. S. A. **100**, 7988–7995 (2003)
55. Zhu, L., He, G., Wang, S., Miller, L.A., Zhang, X., You, Q., Fang, S.: An immersed boundary method by the lattice Boltzmann approach in three dimensions. Comput. Math. Appl. **61**, 3506–3518 (2011)

Undergraduate Research Highlight: Modeling Movement Behavior Among Interacting Species

Anne Talkington

Abstract Migrating species are affected by their interactions with one another, the biotic environment, and the abiotic environment. The ability to understand and predict their patterns of movement is of particular interest for purposes of conservation, development planning, and resource management. Researchers have made attempts to model the direction in which they move, and the extent of their movement. However, current models focus on a specific factor that influences the movement within ecosystems, such as climate or land use. The complexity of the organism's interactions and instinctual drive is often simplified. I present a framework for a matrix-based model that expands on previous models and allows the researcher to describe how an organism interacts with its biotic and abiotic environment, in as much detail as the research demands. This model can describe the strength of an organism's attraction to a particular place, or the relative speed at which it will migrate there. The matrix modeling framework is generalized to be applicable to any species; yet, it can be tailored to the biology and ecology of specific organisms. It predicts the movement of organisms along a gradient on the physical landscape, based on their needs, within a given time. Understanding changing behavioral influences on individuals is a significant step in making an educated decision regarding human intervention in a natural migration pattern.

1 Introduction

For the purposes of modeling discussed in this paper, we consider migration as an organism's movement from one place to another. Migration can be a singular movement event, such as a range shift in response to changing environmental conditions, or a repeated movement event, such as the seasonal migration of a species as part of its circannual rhythm. An organism's biotic environment

A. Talkington (✉)
Department of Mathematics, Duke University, Durham, NC, USA

Department of Mathematics, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
e-mail: anne.talkington@unc.edu

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9_12

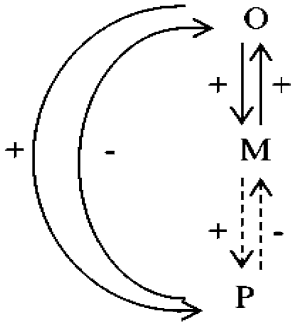


Fig. 1 The variety of species that an organism may interact with. This example simplifies an ecosystem to the network of relationships between an organism (O), a mutualist (M), and a predator (P), three possible types of interactions. The attractive or repulsive nature of interaction between the species is indicated by the + and − signs. This graphic focuses on the relationship between the predator and the organism, and the relationship between the predator and mutualist (*dotted lines*) is weaker and less direct

encompasses factors, including symbiotic species, that could affect an organism's movement patterns [3, 5, 9]. These species may be mutualistic, predatory, or parasitic and may include members of the selected organism's own species (Fig. 1). The organism's abiotic environment encompasses factors such as food, climate, soil quality, terrain, proximity to a preferred or unsuitable ecosystem, or habitable space [1, 12]. Species may migrate according to seasonal or annual changes in these ecosystem factors, in order to live in their most suitable niche at any given time. More permanent changes, such as habitat encroachment or global warming, may influence the viability of the ecosystem and eventually necessitate species range shift, another context of migration [5].

Migration is a primary area of ecological interest because of its implications for potential species interactions, conservation, and resource management. Modeling or projecting migration patterns offers the potential to predict the migrations, and intervene if appropriate [5]. Historically, research in migration modeling has focused on particular environmental factors that affect the strength of an organism's attraction to a place, or its motivation to move. Several studies have encompassed either abiotic or biotic factors, but not both [1, 2, 12]. Barbet-Massin and Virkkala discuss the importance of simple climatic and topological models, building up to consideration of land use and vegetation [1, 12]. Bestley conducted research on the behavioral motivations behind migration, suggesting a common goal or instinct to find a specific environmental condition, as in the motivation behind herd-like migration patterns and the apparent cohesiveness that results from such migrations [2]. More comprehensive studies of species distribution focus on species interactions and interdependence. For example, Jaeschke suggests a layering, feedback method that determines the range of one species as dependent on the range of another [6]. Kissling presents a matrix method for species interaction distribution modeling (SIDM) [9]. These interactions can affect and are affected by migration.

However, questions about examining the interactions dynamically, and the relative complexity or flexibility of the models, are largely unaddressed, and the authors acknowledge that more investigation is needed [3, 6, 9].

Linear models address an array of abiotic contributors to migration at different states in a Markov chain, which can then be evaluated to produce the probabilistic evolution of movement patterns. Jonsen considers both a hidden Markov and Bayesian framework [7]. However, the states in models such as Jonsen's do not discuss interactions between species as a contributing factor to an organism's decision to move. Rather, behavioral inference is treated as an output. The model I am proposing accounts for an organism's interactions with other organisms occupying the environment, as well as aspects of the physical landscape.

Migration models that focus on a general approach cannot be specifically applied to one organism in particular because of the individual variables that affect the patterns of each. Models that analyze one particular species cannot be generalized to other species because of their dependence and construction on a selected number of individualized variables, often gathered empirically and deemed significant. I am proposing a matrix-based migration model, based on movements due to species interactions, that can be tailored to the level of detail required for ecological studies. This model is flexible, so it can be made specific to a focal species but also generalizable across species and timescales. The model can be evaluated as time progresses, by tracking organisms' movement or motion in a broad sense across the landscape of environmental characteristic and physical space. It also introduces a behavioral aspect which has not been addressed in previous studies. The variety of applications of the model are subsequently illustrated in three case studies, including the response of salamanders to deforestation, the seasonal migration of white sharks, and the response of perch to an overflowing river as they make their way to breeding grounds. This model has the potential for broad use in ecological research.

2 Methods

A system of matrices was developed to describe the abundance and relative significance of the biotic and abiotic components affecting an organism's movement. An organism would be more strongly affected (attracted or repelled) by factors occurring in greater amounts or carrying a higher weighting factor of relative importance (see Fig. 2). Refer to Table 1 for a guide to abbreviations used in the matrices as I describe the process of setting up a sample of the model.

As illustrated in Fig. 2, this organism's instinctual attention was shown to be divided between the components of a particular ecosystem that rendered it a suitable (positive) or unsuitable (negative) migration destination. Weightings were assigned to each factor (biotic and abiotic, named in the row headings) with a positive or negative value, where the absolute value represented strength or proportion of interaction, and the sign served as an indicator. The weightings were normalized so that the sum of the absolute values of the entries down each of the columns

OMP (Organism-Mutualist-Predator) Sample Matrices

Biotic: Weightings **W**

	O	M	P
O	0	.5	.6
M	.3	0	.1
P	-.3	-.1	0

Biotic: Amounts **a**

	O	M	P
Amount	1	2	2

Abiotic (FS, or Food-Space): Weightings **X**

	O	M	P
F	.2	.3	.1
S	.2	.1	.2

Abiotic: Amounts **b**

	F	S
Amount	3	5

Net Biotic Motion **N=aW**

O	M	P
0	.3	.8

Net Abiotic Motion **M=bX**

O	M	P
1.6	**1.4**	**1.3**

Fig. 2 A general example of the matrix organization for an ecosystem, analyzed for a general system of focal organisms, a mutualistic species, and a predatory species

Table 1 Key abbreviations in the matrix model

Abbreviation	Interpretation
O	Organism
M	Mutualist
P	Predator
F	Food
S	Space
W	Biotic weightings
a	Biotic amounts
X	Abiotic weightings
b	Abiotic amounts
N	Net biotic motion
M	Net abiotic motion
S	Scaled biotic motion
T	Total motion and migration indices

(here Organism, Mutualist, or Predator) totaled 1. In biotic weighting matrices, a lone organism could not affect itself. A 0 in the matrix indicated complete lack of interaction. However, it was possible for a group of organisms to affect one another. Thus, when more than one organism of a selected group or species was analyzed (and the system was being analyzed as a whole rather than a single organism and its mutualistic cohorts), a nonzero number was entered in the respective element of the

biotic weighting matrix. This value for intraspecific interactions could be positive, to represent grouping or cohesion, or negative, to represent territoriality. A 1 in one entry of the matrix would indicate that no other factor affects the organism. Relevance in the relative weighting system was determined by proximity to a resource or a desperate drive to obtain the resource. For example, the 0.3 at the intersection of row M and column O in the biotic weighting matrix of Fig. 2 indicated that 30% of this particular organism's instinctual attention was given to the mutualist. They interacted positively and attracted one another. By contrast, the −0.3 at the intersection of row P and column O indicated that 30% of the mutualist's attention was given to its predator, which repelled the organism and drove it in the opposite direction. The relationship between the organism and the mutualist or predator may have been more meaningful to one group than another, if the degree of interaction or dependence was greater. Thus, the intersection element of row M and column O may not necessarily be the same as the intersection of row O and column M.

The biotic "amounts" vectors were absolute and represent the number of organisms of a particular group or species. The biotic groups could be simplified as "lone organism," "other organisms of the same species," "all organisms of one species," "mutualistic species," or "predatory species." In Fig. 2, the O, M, and P columns each represent a group by this definition, though the scope of this model is not limited to OMP systems. Abiotic amounts were quantified relatively because the abiotic factors could have ranged from food supply, to water temperature, to amount of precipitation. They were therefore marked on a scale of prevalence from 1 (barely present) to 100 (very prevalent) in the environment, for consistency of units. In the sample system illustrated in Fig. 2, I have analyzed 1 organism, and populations of 2 mutualists (of one species) and 2 predators (of one species). I considered the food and available space in the ecosystem, here denoted as F and S, as my abiotic components of choice. This system was small and did not contain a large food supply (values of 5 and 3, respectively). However, positive weighting values in the abiotic table indicated that the quality of the food and the space available were still attractive to its members. Factors that limited or depended on one another were grouped into one column, determined by the limiting factor. (For example, the organisms may have needed shelter as well as space, but their instinct to migrate to a particular location was exclusively dependent on available space, regardless of the amount of shelter in the habitat. Thus, "space" was deemed a category in this example.) The designation of categories is left to the discretion of the researcher in each application of the model.

The effect of each biotic or abiotic factor on an organism or group's overall movement was determined through matrix multiplication, which multiplied the amount or prevalence of a variable by its relative importance. Each "amounts" row vector (**a** and **b**) multiplied its respective "weightings" matrix (**W** and **X**). The sum of the weighted values developed an index of biotic movement and an index of abiotic movement, represented as net movement values. Each column in the movement vectors (**N** and **M**) coincided with an organism, group, or species (the same biotic groups, or columns, initially defined in the amounts matrix) and was interpreted as the amount that each group would be driven to move based

OMP System

Net Biotic Motion **N=aW**

O	M	P
0	.3	.8

Net Abiotic Motion **M=bX**

O	M	P
1.6	**1.4**	**1.3**

Abiotic Scaling

O	M	P
3.2	5.6	6.5

Scaled Biotic Motion **S**

O	M	P
0	**1.68**	**5.2**

Total Motion and Migration Indices **T=S+M**

O	M	P
1.6	**3.08**	**6.5**

Dynamic of the System at this Instant in Time (sum of all matrix entries)	Group Average Movement in (Arbitrary) Positive Direction	Individual Average Movement	Weighted Average Movement
11.18	**3.72**	**2.24**	**4.15**

Fig. 3 Example of biotic and abiotic scaling. The scaling factors greater than 1 for each column reveal that species interactions are enhanced by a favorable external environment

on biotic and abiotic factors, respectively. Under the normalized scaling system, these "affinity units" represent drive as a property of the organism's behavioral preferences, or the strength of an organism's attraction to a new point. Continuing the OMP example, the organism in Fig. 3 would be driven 1.6 affinity units due to abiotic environmental factors, whereas the mutualist would be assigned 1.4 units, and the predator would be driven 1.3 units.

The biotic movement values were each multiplied by an abiotic scaling factor to account for enhanced or diminished interaction behavior, which is dependent on the particular set of abiotic factors in the system, as illustrated in Fig. 3. Physical characteristics of a particular environment could influence an organism's migrational behavior both directly, through habitat attractors and deterrents (measured by net abiotic motion), and indirectly, through its impact on the social and behavioral components of biotic motion (accounted for by abiotic scaling). For example, a temperate climate might encourage activity for some organisms while discouraging those who prefer it drier/wetter, or hotter/colder. Referring to the sample system provided serves to clarify the need for and implementation of these scaling factors. In this OMP system, the mutualist's scaling was a factor of 4. Each mutualist's motion was multiplied by a factor of 4 for each ± 1 it responded to the environment (net abiotic motion), indicating greater awareness of, response to, and interaction with the other species in the system given its environmental conditions. This means that in this particular environment, its drive to move based on biotic interactions was amplified by a factor of 4 times its drive to move based on the abiotic environment.

Its scaled biotic motion was thus $(1.4 \times 4) \times 0.3 = 1.68$. The same math applies in the columns representing the organism and the predator, accounting for biotic interactions enhanced by environmental sensitivity. Here, the predator's scaling was a factor of 5, so its scaled biotic motion was $(1.3 \times 5) \times 0.8 = 5.2$. The O column was less interesting than the M and P columns mathematically because of its factor of 0 in the net biotic motion entry. Had this been a nonzero value, it would have been multiplied by 3.2, or 1.6×2. However, the calculation was informative nonetheless because it illustrated that regardless of its environmental response, no degree of environmental acuteness can create a biotic interaction that did not initially exist.

The scaled biotic movement vector **S** was comparable to the net abiotic movement vector **M**, and the sum of the two vectors represented the total overall movement for each organism in the system ($\mathbf{T} = \mathbf{S} + \mathbf{M}$). In Fig. 3, the mutualist experienced a total of $1.4 + 1.68 = 3.08$ units of drive affinity. Again, each of the columns coincided with the associated organism, species, or group in question. The elements could be interpreted as the strength of attraction to a region, or as the speed of migration in a specified direction (Figs. 4, 5). Directed movement described the "push" away from the less-than-ideal conditions in the current environment, and destination attraction reflected a "pull" toward the ideal niche. Numbers with a larger magnitude relative to the other entries in **T** indicated a greater attraction.

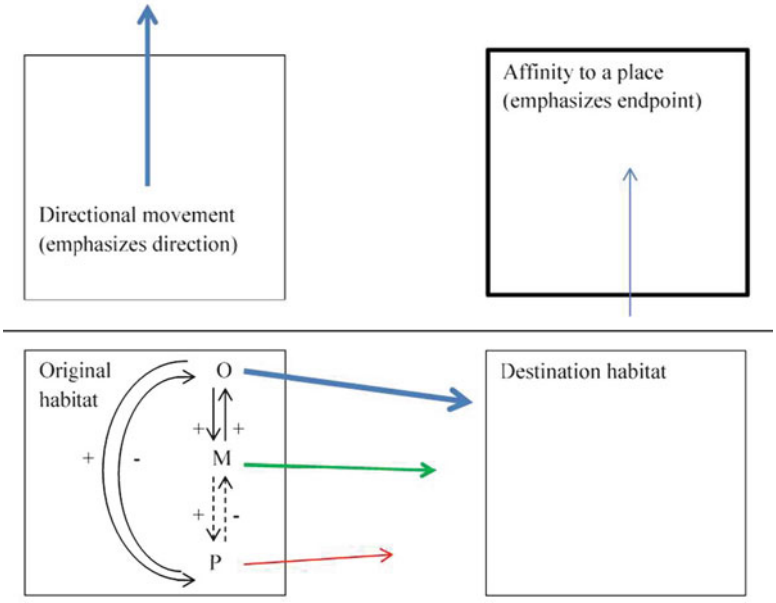


Fig. 4 Conceptual interpretation of migration analysis. The matrix method accounting for species interactions can be used to determine the extent of a population's movement in a specified direction (*top left*), or to determine the population's attraction to one particular location (*top right*). Together, the two measurements illustrate the overall migration in the ecosystem, as shown for the OMP system (*bottom*). The *thickness and length of the arrows* represent the strength of attraction to a place, or the drive to migrate there

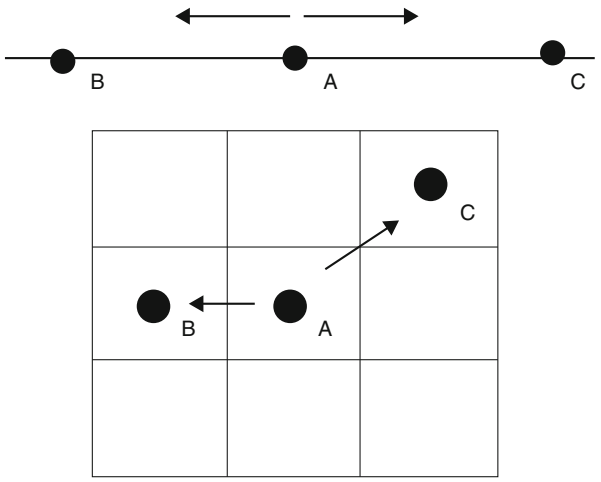


Fig. 5 Intuitive spatial illustration of the model. An organism at point A will evaluate the biotic and abiotic conditions at its current state as well as points B and C. The organism will then decide where to move based on its affinity to the different points. This can be represented linearly, or on a grid over a two-dimensional surface

3 Metrics of System Motion

Refer to Table 2 for a guide to abbreviations used in the matrices as I describe the process of analyzing the model.

The dynamic of the system was the sum of all movement elements, for all organisms (d = summed elements of **T**; in the OMP example: $1.6 + 3.08 + 6.5 = 11.18$). The average movement could be interpreted in terms of three indices: group average movement, individual average movement, and weighted average movement. Group average movement represented migration simply as a system of groups or species, each of which carried equal weight in the environment. Thus, it was defined as the dynamic of the system divided by the number of biotic groups ($g = d/$length of **a**; for OMP: a dynamic of $11.18/3$ groups). Individual average movement focused on the contribution of each individual to the dynamic. The number of organisms in each group was more important than the group to which each belonged. The organisms interacted and moved together on the individual level, and each one contributed both to the total overall dynamic and the element of the dynamic attributed to its group or species. Individual average movement was thus defined as the dynamic of the system, divided by the total number of organisms in the system (sum of the entries in the biotic "amounts" matrix, or $i = d/$summed elements of **a**). For OMP, this value was obtained from $11.18/(1 + 2 + 2)$. Weighted average movement analyzed the system from a center of mass perspective. This index accounted for both the number of individual organisms and the relative impact of the group or species to which each belonged. Thus, the weighted average index was

Table 2 Abbreviations describing metrics of system motion

Abbreviation	Interpretation
d	Dynamic of the system
g	Group average movement
i	Individual average movement
w	Weighted average movement

determined by multiplying each entry in the total overall movement matrix by the number of organisms in the associated column, taking the sum across all columns, and dividing by the total number of organisms ($w = (\mathbf{a}_1\mathbf{T}_1 + \mathbf{a}_2\mathbf{T}_2 + \cdots)/$summed elements of $\mathbf{a}$, where $\mathbf{a}_k$ is the kth element of the vector $\mathbf{a}$, or more succinctly: $\mathbf{a} \cdot \mathbf{T}/$summed elements of $\mathbf{a}$). The value for the OMP case is determined by the following calculation: $(1 \times 1.6 + 2 \times 3.08 + 2 \times 6.5)/(1 + 2 + 2)$. The indices were systematically calculated through a program in Mathematica software.

The indices can be summarized as follows:

- *Dynamic of the System:* sum of all movement elements
- *Group Average Movement:* dynamic, divided by the number of groups (columns)
- *Individual Average Movement:* dynamic, divided by the number of individuals (organisms of all species)
- *Weighted Average Movement:* sum of all total movement vectors multiplied by the number of organisms in a column, divided by the total number of organisms

4 Case Studies

The following three case studies, based on past ecological research, represent applications that highlight specific points of interest in the development and interpretation of the new migration model, and interesting results we can derive by constructing and implementing our theoretical framework around the real system. Each case represents a different element in which the matrix method is useful for analyzing the ecosystem. Case 1, salamanders, explicates an individual–group relationship. Case 2, Pacific white sharks, utilizes the method with respect to two groups, or a "split population." Finally, Case 3 uses golden perch to illustrate the difference between the three migration indices and the usefulness of each to answer particular research questions. The analyses are scenarios derived from the published literature, rather than explicitly taken from previously obtained data, but illustrate the potential of the model if more data were known. The initial input values come from applied knowledge of the dynamics in species interactions. Running the model confirms the expected behavior in each situation; indeed, the results map qualitatively onto conclusions drawn about the species in question.

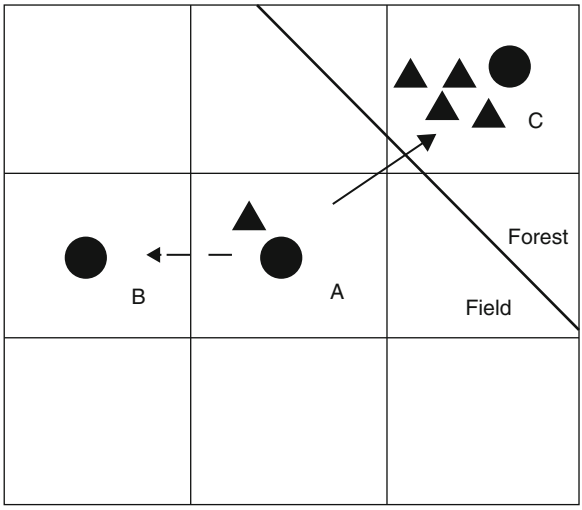


Fig. 6 Illustrated interpretation of the salamander case study. The lone salamander (represented by the *triangle*) at point A is attracted to the safety of the forest vs. exposure in the field, and the larger group of salamanders at point C. Upon evaluating its options, such as the habitat at point B or elsewhere in the surrounding grid, it is not attracted to the open space or greater distance from the group and is driven to move toward (has the greatest affinity for) point C in this instant

4.1 Case 1: Salamanders

Rittenhouse and Semlitsch have conducted research pertaining to salamander populations (*Ambystoma maculatum*) in the context of deforestation [11]. Here, we are considering seasonal migration for the purpose of breeding. This particular example, taken from a situation in their work, illustrates an outside, lone salamander who is attracted to the group and a part of the physical environment (Fig. 6). According to the findings of Zamudio and Wieczorek, this grouping or clustering is a realistic scenario [13]. Specifically, the salamander prefers the safety away from the boundary of its ideal habitat. This is represented by biotic weightings of 0.5 toward other salamanders, and abiotic weightings of 0.25 to the forest with a complementary weighting of −0.25 away from the edge. The abiotic amounts in this matrix reflect the size of the system, where there is more "safe" area (given as 100 units of space or squares on a grid) than boundary "unsafe" zone (given as 50 units of space). The focal salamander will exhibit strong tendencies to migrate toward the larger central group, quickly (total motion index of 81.25). The salamanders in a larger cluster, however, will stay generally where they are (total motion index of 19.38). They will accept the salamander into their breeding group, but the lone will migrate to the group before the group moves to the loner. Their directed movement analysis (Fig. 7) reveals the group's minimal preference to any particular location. Rather, they are strongly attracted to one another and their environment, and pursuing their instinct to come together as one group. Their low overall movement value indicates stirring around the group's center, represented as the mean.

When the matrices are analyzed and indices are reinterpreted to represent attraction to the current location of the main group, all of the salamanders exhibit high overall totals (Fig. 8). The difference in overall movement values suggests that

Salamander Migration Matrices, Directed Movement

Biotic: Weightings

	Lone Salamander	Group of Salamanders
Lone Salamander	0	.5
Group of Salamanders	.5	0

Biotic: Amounts

	Lone Salamander	Group of Salamanders
Amount	1	10

Abiotic: Weightings

	Lone Salamander	Group of Salamanders
Edge of habitat	-.25	-.25
Forest "safety zone"	.25	.25

Abiotic: Amounts

	Edge of Habitat	Forest "Safety Zone"
Amount	50	100

Total Motion and Migration Indices

Lone Salamander	Group of Salamanders
81.25	**19.38**

Dynamic of the System	Group Average Movement	Individual Average Movement	Weighted Average Movement
100.63	**50.31**	**9.15**	**25.00**

Fig. 7 Matrix analysis of salamander population. The input values for the matrix elements reflect relative speed, strength, and amount of movement toward a particular location

the salamanders will prefer migrating in one direction, but only to an extent. As an organism or population comes closer to its goal, it will have less need to go as far or as fast to reach the preferred condition. At this stage in their migration, the organisms tend to have more affinity for staying in a particular habitat with the desired relevant resources, and coexisting with the other organisms nearby. They have less affinity for the biotic and abiotic factors beyond this environment, which become less relevant or less favorable as they wander past their ideal. Thus, the organisms express less desire to migrate further. This phenomenon of picking a habitat destination is explored as a series of time steps with changing destinations. When migrating organisms overshoot their ideal, the matrix "amount" and "weighting" values in the next time step evaluation reflect deterrents along their migration path. These values are considered relative to one another, and normalized at each step so that the column sum remains 1. As a result, the model can illustrate that wandering salamanders will be drawn back to the balance of their ideal location. The two analyses complement one another: if organisms are not attracted elsewhere, then they will tend to stay in their current location. Once the organisms are in their ideal location, they will not be attracted elsewhere.

Salamander Migration Matrices, Destination Attraction

Biotic: Weightings

	Lone Salamander	Group of Salamanders
Lone Salamander	0	.1
Group of Salamanders	.1	.2

Biotic: Amounts

	Lone Salamander	Group of Salamanders
Amount	1	10

Abiotic: Weightings

	Lone Salamander	Group of Salamanders
Edge of habitat	-.45	-.35
Forest "safety zone"	.45	.35

Abiotic: Amounts

	Edge of Habitat	Forest "Safety Zone"
Amount	50	100

Total Attraction and Migration Indices

Lone Salamander	Group of Salamanders
47.25	**57.93**

Dynamic of the System	Group Average Migratory Attraction	Individual Average Migratory Attraction	Weighted Average Migratory Attraction
105.18	**52.59**	**9.56**	**56.95**

Fig. 8 Matrix analysis of salamander population. The input values for the matrix elements reflect attraction to the immediate surroundings of the primary group

4.2 Case 2: White Sharks

The matrix method is also applicable in studies of migratory populations, to describe movement that recurs temporally (Fig. 9). The principal motivation behind the movement remains the same. The migration patterns of white sharks in the Pacific Ocean (*Carcharodon carcharias*) illustrate an example of a "split group" [4, 8]. The model simplifies the scenario, with proximity to the coastline, coastal traffic, and water temperature as primary variables (Fig. 10). The biotic set was divided into two specific groups: sharks near the coast of the western USA, and sharks near the center of the ocean. The sharks' preferred location depends on season. In general, the model shows that the white sharks will distinctly prefer one extreme over another. They will convene and remain in this location for the duration of several months. The groups, though they are represented as balanced in the model, will not "meet in the middle" (represented by a mean movement of 0, preserving the existing population center) because the environment is not as favorable as one extreme or the other. The biotic factors determine that the sharks will gather and potentially compete for prime territory as more sharks arrive. The abiotic factors determine where they will gather.

The destination attraction analysis yielded results that were similar for both groups because the white sharks share common instincts; the only differences were environmental. The ocean sharks are seasonally attracted toward the inland sharks,

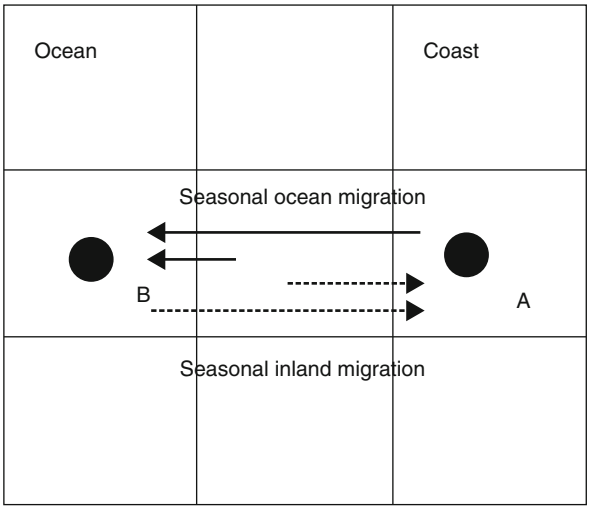


Fig. 9 Illustrated interpretation of the white shark case study. The sharks on the grid will have greater affinity for a coastal habitat (with A as the ultimate destination, an affinity maximum) or the open ocean (with B as the ultimate destination, an affinity maximum), according to the season

Pacific White Sharks Migration Matrices, Directed Movement

Biotic: Weightings

	Ocean Sharks	Coastal Sharks
Ocean Sharks	0.3	-.2
Coastal Sharks	-.2	.7

Biotic: Amounts

	Ocean Sharks	Coastal Sharks
Amount	100	100

Abiotic: Weightings

	Ocean Sharks	Coastal Sharks
Coastal Traffic	-.15	-.075
Water Temperature	.35	.025

Abiotic: Amounts

	Coastal Traffic	Water Temperature
Amount	25	100

Total Attraction and Migration Indices

Ocean Sharks	Coastal Sharks
500.00	**47.50**

Dynamic of the System	Group Average Movement	Individual Average Movement	Weighted Average Movement
547.50	**273.75**	**2.74**	**273.75**

Fig. 10 Matrix analysis of Pacific white shark population. The groups of organisms represent the division between two extreme preferred locations, each more likely to be habited seasonally. This analysis describes two evenly split groups that intuitively come together near the shore or in the middle of the ocean, depending on the preference for the current half the year

but further migration would be detrimental for the inland sharks. Their instinct to remain in their current habitat is represented by the low and cancelling relevance values of the abiotic factors as they travel outside this location. If the sharks travel too far, the amounts and relevance of the abiotic factors will force them to turn around and thus "meet" the immigrating ocean sharks. The amount of correction in the amount and relevance values corresponds to the amount the white sharks stray from their ideal course.

4.3 Case 3: Golden Perch

A migratory analysis of the fish populations of the Murray–Darling River System in Australia signifies the importance of distinguishing between the average movement indices of the migration model [10]. The model yields a projection of migration in response to flood conditions. It generally tells the story of one golden perch (*Macquaria ambigua*) in a crowd where the trend is upstream migration for the physically capable fish (Fig. 11).

Analyzing the averages gives a distinct picture of the migration interactions occurring on a smaller scale within the system. The individual average assumes that the movement in the system is an aggregate measurement, and that each perch is only contributing a small part to the dynamic. By this analysis, the fish exhibit relatively weak attraction along the stream and are not particularly directed, individually or as a group. By contrast, the group and weighted averages yield similar, considerably higher values (Fig. 12). They represent each group member as moving or being attracted by the full value calculated in the appropriate column of the overall total matrix. The disparity between the two index results is caused by the factor of size difference in biotic groups, as accounted for in the weighted average.

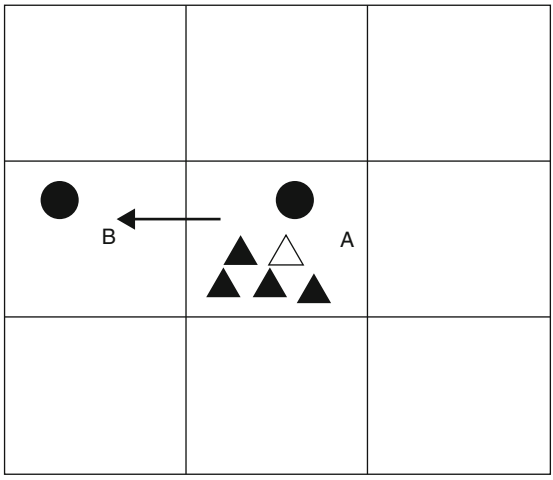


Fig. 11 Illustrated interpretation of the golden perch case study. The lone perch (*open triangle*) exhibits grouping behavior with the larger school of fish (*filled triangles*), while all fish instinctually swim upstream, shown here from point A to point B

Golden Perch Migration Matrices, Destination Attraction

Biotic: Weightings

	Lone Perch	Group of Perch
Lone Perch	0	.1
Group of Perch	.1	.1

Biotic: Amounts

	Lone Perch	Group of Perch
Amount	1	6

Abiotic: Weightings

	Lone Perch	Group of Perch
Water Flow and Temperature	.45	.4
Lack of Predators	.45	.4

Abiotic Amounts

	Water Flow and Temperature	Lack of Predators
Amount	50	100

Total Attraction and Migration Indices

Lone Perch	Group of Perch
112.05	**106.20**

Dynamic of the System	Group Average Migratory Attraction	Individual Average Migratory Attraction	Weighted Average Migratory Attraction
218.25	**109.13**	**31.18**	**107.04**

Fig. 12 Matrix analysis of population of perch in the Murray–Darling River. The summary indices each measure a different element of the population dynamics, on an individual or a group level. The researcher can determine which index is most reasonable for each particular study

Each index addresses a different question. The details of the input elements reveal that the perch have some instinct to stay together, but physical capability and desire to obtain shared resources results in the common migration.

5 Discussion

The matrices in the migration model allow for a system of organization with as much detail as the researcher finds necessary or knows. The basic Organism–Mutualist–Predator and Food–Space structure can be condensed, or expanded with more categories and subcategories to thoroughly describe the ecosystem and its connection to the human system. This flexibility allows the model to apply specifically to any species from its general form, at any level of biotic or abiotic complexity, for any duration of time. The decision between destination attraction versus directed movement analysis, as well as the index of choice, has been developed as a metric to be used as relevant to the specific ecological research in question. The matrix entries allow various environmental factors to be compared with one another and accounted for through a general units system. In applying the model to real-world scenarios, I have demonstrated its ability to contribute to the knowledge we can obtain about a system through previous models, such as interaction matrices and layered feedback loops [6, 9].

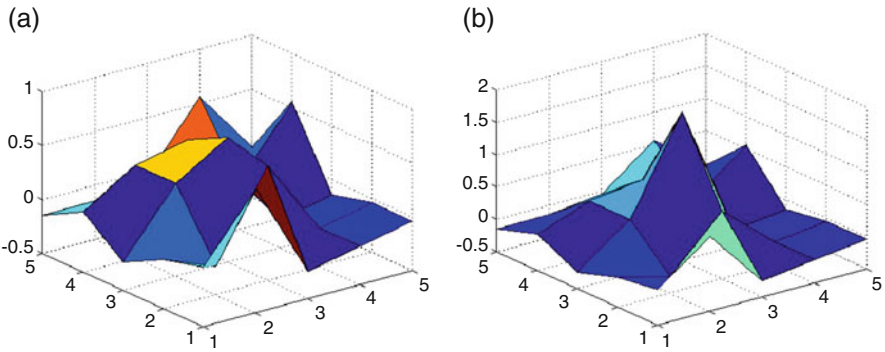


Fig. 13 Preliminary work on an individual-based model (IBM) visualization of the landscape presented by the matrix model. Over time, the organisms concentrate at the peak according to the affinity gradient. (**a**) Initial distribution of organisms. The *x* and *y* axes represent physical coordinates. The *z* axis represents the normalized initial population. (**b**) Organisms find a "peak" in the fitness landscape, and are most likely to move to this point in the physical landscape, thus creating a "peak" in population. After several iterations, the concentration of organisms at the peak continues to increase and may stabilize at a dynamic equilibrium

Its applicability could be further explored through programming an individual-based model (IBM), to visualize the implications suggested by this method and facilitate calculations over several time iterations. Spatially, if I consider the result of directed movement and destination attraction as a trajectory through a vector field, the entries in the total attraction vector **T** would physically represent individual movement vectors toward the equilibrium niche, and dictate the path traced and updated in the IBM. IBM would overlay the physical space with a grid of affinities, and enable researchers to better understand more complex systems over a longer duration of time using this model. The visualization in Fig. 13 presents the potential for an IBM approach.

The elements of the movement vector **T** account for the amount of attraction in one time-step, or iteration. Running the model over multiple iterations involves advancing the state of the system by the given vectors and reevaluating the matrices. Each endpoint becomes a new beginning point with each step in the model. The weighting factors update based on the organism's biological needs (e.g., fish require a certain amount of water at a certain temperature), the difference between its current state and its ideal state, the environmental factors in close proximity to the organism (e.g., a nearby predator will occupy the majority of an organism's attention), and any external changes that influence the system (e.g., the introduction of a pollution agent or the barrier of a new development plan). The conditions for an "ideal niche" form a fitness gradient that overlays the physical space along which the species move. Over time, I can see a trajectory form from the movement vectors along this gradient, outlining the path of migration toward the niche. In this interpretation of the ecosystem space, the niche is considered the equilibrium, minimum, or steady-state point of this system. If an organism "overshoots" its ideal location in

one time step, the next time step would result in a vector pointing back to the niche. The gradients formed by directed movement (push away from current location) and destination attraction (pull toward another location) analysis should result in a consistent picture.

The primary limitation of the model is currently the availability of behavioral knowledge on a wide range of species. Even so, the model in its present state can be used to answer interesting, if more theoretical, questions. For example, it is possible to explore the relative importance of different factors in determining the trajectory of an organism's potential migratory path, or to examine the number and location of the niches that form as potential endpoints, given an initial distribution of resources across a landscape.

6 Conclusions

The development of this new model provides an opportunity to reanalyze ecological research and data. The matrices highlight the relative impact of each biotic or abiotic factor, and the effect of changing one detail of the ecosystem. This model demonstrates that interactions are critical in assessing migration, and different levels of interaction contribute to different migration tendencies. Depending on an organism's needs, such conditions and interactions can change over time.

This method can be applied to reevaluate environmental phenomena and the consequences of the human role in the ecosystem, in the context of behavioral tendencies [4]. We can trace human intervention through the network of species interactions and interdependence. With the ability to evaluate the outcome of a future course of action, we can thus become more aware of the successes and shortcomings of conservation efforts. The migration model facilitates efforts and progress in migration research.

Acknowledgements The author would like to thank the McShea Laboratory and the Biology Department at Duke University for support in this research project, and Julie Tuttle for guidance.

Competing Interests The author has no competing interests.

References

1. Barbet-Massin, M., Thuweller, W., Jiguet, F.: The fate of European breeding birds under climate, land-use and dispersal scenarios. Glob. Chang. Biol. **18**(3), 881–890 (2012). doi:10.1111/j.1365-2486.2011.02552.x
2. Bestley, S., Jonsen, I.D., Hindell, M.A., Guinet, C., Charrassin, J.-B.: Integrative modeling of animal movement: incorporating in situ habitat and behavioural information for a migratory marine predator. Proc. R. Soc. B Biol. Sci. **280**(1750) (2013). doi:10.1098/rspb.2012.2262
3. Davis, A.J., Jenkinson, L.S., Lawton, J.H., Shorrocks, B., Wood, S.: Making mistakes when predicting shifts in species range in response to global warming. Nature **391**(6669), 783–786 (1998). doi:10.1038/35842

4. Domeier M.L., Nasby-Lucas, N.: Migration patterns of white sharks Carcharodon carcharias tagged at Guadalupe Island, Mexico, and identification of an eastern Pacific shared offshore foraging area. Mar. Ecol. Prog. Ser. **370**, 221–237 (2008). doi:10.3354/meps07628
5. Hellmann, J.J., Prior, K.M., Pelini, S.L.: The influence of species interactions on geographic range change under climate change. In: Ostfeld, R.S., Schlesinger, W.H. (eds.) Year in Ecology and Conservation Biology, vol. 1249, pp. 18–28. Blackwell Science, Oxford (2012)
6. Jaeschke, A., Bittner, T., Jentsch, A., Reineking, B., Schlumprecht, H., Beierkuhnlein, C.: Biotic interactions in the face of climate change: a comparison of three modeling approaches. PLoS One **7**(12), e51472-Article No.: e51472 (2012). doi:10.1371/journal.pone.0051472
7. Jonsen, I. (2016). Joint estimation over multiple individuals improves behavioural state inference from animal movement data. Sci. Rep. **6**, 20625. http://doi.org/10.1038/srep20625
8. Jorgensen, S.J., Reeb, C.A., Chapple, T.K., Anderson, S., Perle, C., Van Sommeran, S.R., Fritz-Cope, C., Brown, A.C., Klimley, A.P., Block, B.A.: Philopatry and migration of Pacific white sharks. Proc. R. Soc. B **277**, 679–688 (2010). doi:10.1098/rspb.2009.1155
9. Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kuehn, I., McInerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.-C., Zimmermann, N.E., O'Hara, R.B.: Towards novel approaches to modeling biotic interactions in multispecies assemblages at large spatial extents. J. Biogeogr. **39**(12), 2163–2178 (2012). doi:10.1111/j.1365-2699.2011.02663
10. Reynolds, L.F.: Migration patterns of five fish species in the Murray-Darling River System. Aust. J. Mar. Freshwat. Res. **34**, 857–871 (1983)
11. Rittenhouse, T.A.G., Semlitsch, R.D.: Grasslands as movement barriers for a forest-associated salamander: migration behavior of adult and juvenile salamanders at a distinct habitat edge. Biol. Conser. **131**(1), 14–22 (2006). doi:10.1016/j.biocon.2006.01.024
12. Virkkala, R., Marmion, M., Heikkinen, R.K., Thuweller, W., Luoto, M.: Predicting range shifts of northern bird species: influence of modeling technique and topography. Acta Oecol. **36**(3), 269–281 (2010). doi:10.1016/j.actao.2010.01.006
13. Zamudio K.R., Wieczorek A.M.: Fine-scale spatial genetic structure and dispersal among spotted salamander (Ambystoma maculatum) breeding populations. Mol. Ecol. **16**, 257–274 (2007). doi:10.1111/j.1365-294X.2006.03139.x

Index

A
Abiotic interactions, 234, 235, 237
Acute experimental pain, 3
Araneae, 164, 180
Autoregulation, 64, 71, 75–99, 104–105, 108–109

B
Bile acid dynamics, 140, 144
Biotic interactions, 238, 239

C
Circadian rhythm, 2, 4, 8, 14–17, 23–46, 50–52, 54, 58, 59, 61
Clostridium difficile, 137–159

D
Diabetes, 63, 71, 72, 99, 101, 102, 108–111, 135
Differential equations model, 80
Dispersal, 164–166, 168, 170, 172, 173, 176, 180–182
Dorsal horn neural network, 26

F
Fluid mechanics, 170
Fluid-structure interaction, 175, 176, 181, 182, 185, 187, 205

G
Gap junctions, 15, 76–79, 86, 104, 105

H
Heart development, 214
Heart tube, 213, 217–219, 221, 222, 228
Hematocrit, 213, 223–228
Hemodynamics, 64, 70, 76, 87, 111, 229
Homeostatic sleep drive, 4, 8–17, 25, 40, 46, 50–53, 55, 56, 58–61

I
Immersed boundary method (IBM), 213, 214, 216, 217, 248

K
Kidney, 63–72, 75–99, 101–111, 115–135

L
Loops of Henle, 102, 116–118, 120, 125, 128, 134

M
Mathematical model, 2, 16, 17, 24, 25, 27–31, 40, 50, 59, 63–72, 76–81, 86, 102–108, 117–122, 132, 137–159, 175, 176, 214, 222
Matrix model, 236, 248
Metabolism, 107, 111, 137–159
Microbial interactions in the gut, 140
Migration, 233–236, 239, 241–249
Myogenic response, 64–70, 76, 80, 84–87, 102, 105

A.T. Layton, L.A. Miller (eds.), *Women in Mathematical Biology*, Association for Women in Mathematics Series 8, DOI 10.1007/978-3-319-60304-9

N
Neural firing rate model, 16
Neuropathic pain, 17, 24, 26, 36, 39, 40, 46
Non-linear model, 80

R
Renal hemodynamics, 64, 70, 87

S
Sleep deprivation, 2, 9–16, 40, 46, 50, 52–56, 60
Sleep restriction, 15, 17, 50, 55, 57–58, 60
Smooth muscle, 64, 65 67, 68, 76, 77, 79, 81, 86–88, 99, 104, 105, 109
Spider, 163–176, 179–208

T
Tubuloglomerular feedback, 64, 69–70, 76, 85, 87, 102, 104, 105

U
Urine concentrating mechanism, 102, 116, 117, 119, 134

V
Vasa recta 102, 110, 116–119, 121, 123, 127–132, 134