Dynamics of protein-mediated polymer coupling and their implications in antibody production and emergent patterning

Thesis by
Soichi Hirokawa

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended July 12, 2022

Soichi Hirokawa
ORCID: 0000-0001-5584-2676

ACKNOWLEDGEMENTS

They say it takes a village to raise a child, and having been given the nickname of "Growing Boy" for much of my 20s, one can imagine how large this village must have been. And yet I can truly say that any growth I have experienced, both scientifically and personally, I owe to the friends and family in this village.

I would like to start by first thanking Rob. I started at Caltech with a great deal of uncertainty and self-doubt about what I wanted to do in graduate school. It was while listening to Rob give a talk about how we should be guided by our curiosity that I decided to drop by his office to get to know him. Had he not been in his office later that afternoon, I may be living an entirely different life now, but I am grateful that as fate would have it, I would meet the person who would become not just my advisor but a close mentor in life as well. I am grateful for the opportunities Rob has given me, from traveling and teaching in New Zealand, South Korea, France, and the Marine Biological Laboratory in Woods Hole, MA, to spending time in other labs and talking to incredible scientists from around the world. I appreciate his sentiment that the research that we do today should become a lecture topic in a classroom tomorrow and it has been a guiding principle that I intend to carry with me. And while the independence that Rob gave me to approach my research in my own way was challenging early on, I can say that it has also led me to grow as a scientist and as an individual. It is with gratitude that I say that my recent and future research directions are being guided by a combination of curiosity and Rob's support.

I would also like to thank several professors both at Caltech and at other institutes. It has truly been a pleasure to learn from David Baltimore, who has always been direct but incredibly supportive. I appreciate all he has done over my early years of graduate school from helping me get up to speed on immunology to offering advice about the academic path. I would also like to thank David Schatz, who over my many years of working with him has been an amazing wealth of knowledge and dedication in the area of V(D)J recombination. His constant support and enthusiasm for the latest updates on my research had been a huge motivation in times when I was scratching my head over the work I was doing.

I would also like to thank members on my thesis committee. I have interacted with Keith Schwab over my many years as a graduate student in the Applied Physics department and while my conversations with him have been numerous, the one that sticks with me is on a drive to Utah on the way to a lab group retreat in my first year of school (sorry I navigated us in the opposite direction for half an hour). My conversation during that drive was a key factor in solidifying my intent on joining Rob's lab. I would also like to thank Matt Thomson, whose excitement whenever I had new data about the projects I was working on led to many meaningful conversations about how I could continue to advance to the next steps. I also thank Dave Hsieh for his deep desire to learn more about biophysics and for his constant questions to understand my work better.

I certainly could not have gotten as far as I have on any of my projects without the constant support and feedback of my coauthors. Griffin, with whom I had shared an office for many years, has always been a great friend. His patience in teaching me when I knew so little and constant willingness to provide feedback on the problems I had been facing led me to learn a great deal in molecular biology, data analysis, and statistics. Nathan was one of the first collaborators I brought on board for the V(D)J work and his carefulness in all of the work that he did was a major reason we were able to explore so many different mutations with such ease. Had he not ruminated on a problem I was facing and offered some thoughts the next day, I might not have learned that it is worth expanding outside of the field to find people with knowledge on a particular problem. Geoff Lovely was in many ways my first true mentor in graduate school when he came back from the east coast three times to help teach me everything I needed to know to hit the ground running for my V(D)J work (and we set off at a blistering pace!). His intense focus could not be understated, but his endless dedication and support, best exemplified when he called me two minutes after I sent him a text at 4 am in his time zone, gave me the drive to keep working hard and for that I am deeply grateful. Mike has been an incredible source of knowledge when it comes to protein expression and purification systems. From helping me purify HMGB1 late into the night all the way to pulling back the curtain on the mysteries of purification, Mike has always been a great person to learn from both on a scientific and personal level.

I am deeply appreciative of the time that Heun Jin and Rachel had taken to help me make progress on my active matter work. Learning about optics and about building microscopes from Heun Jin has been one of the most useful and fun skills I have picked up as a student, and to be able to put together a custom microscope for the photobleaching experiments was not only immensely satisfying but further cultivated my enthusiasm for building things with my own hands. In addition to teaching me about microscopy, his hard work behind the scenes to ensure that I could continue to run experiments and his frequent willingness to listen to and offer ideas for next steps have helped me develop a greater sense of confidence in my ability to think critically and conduct research effectively. I would also like to thank Rachel whose hard work, determination, and skill in soaking up new information have been a source of inspiration over the years. I thank her for constantly pushing me to be a better scientist in the lab and considering the struggles of life while biking up 12+% grades. To all of these collaborators I wish to extend my sincerest thanks for their time in making me and my work better than they would have been had I attempted to fly solo.

One of the great things about Rob is his ability to attract a welcoming and brilliant group of scientists. I have spent countless hours reading active matter papers with Vahe, which has led me to think carefully about how to word my own writing. Working with Ana in the lab and pestering Tom about math and data analysis problems has led to many fruitful discussions. Learning from other deep thinkers like Muir and Manuel have all provided greater clarity in my own work. Discussions with other members of the group including Gita, Suzy, Grace, Gabe, Sara, Avi, and Scott, even if it was away from research, have always been a pleasure. And though not formally a part of the group, I would like to thank Nigel, Rob's illustrator, for teaching me how to design better figures in my work. While I still aspire to achieve the artistic ability that Nigel demonstrates in his work, I have found his perspectives on how to view art and the world to be some of the most memorable insights.

I would also like to extend my gratitude to Justin Bois. Being a teaching assistant under him for multiple years of Bi1x and to get a chance to talk to him about data analysis whenever he has a free moment has helped me become a better communicator and scientist as a result.

I would like to thank Dan Needleman for graciously allowing me to work in his lab at Harvard for several months. Although the original project plan we had envisioned did not pan out for reasons beyond our control, my time in his lab was nothing short of spectacular and helped me realize that the mistakes and failures that I had experienced before my visit were helping me to grow as a scientist. I would also like to thank Zev Bryant, who also allowed me to work with several members of his group at Stanford, and who had been incredibly insightful in offering advice about the scientific pursuit. I found a wonderful community of enthusiastic scientists in both of their groups, and would like to give special mention to Charlie, Athena, Easun, Xingbo, Will, Helen, Gloria, Brian, Marta, Colm, Maya, and Peter for the many fun discussions we shared, the hard work we did together, and for constantly making their labs a welcome place to work.

The work that I have been able to accomplish has also been aided by the support of several administrators. I would like to thank Jonathan, Celene, Kimberly, Pam, and Rui for their support as lab managers who constantly helped make sure we could get the research equipment that we needed to continue to run our research. I would also like to thank Jennifer Blankenship and Christy Jenstad for helping me get settled in at Caltech from day one.

I would not have been able to go through graduate school without the support of many dear friends and while I fear leaving out anyone who falls into this category, several groups of friends immediately come to mind. First, to the many housemates I have lived with, particularly up on Calaveras, at Steuben, and at Francesca. I would like to give particular shoutouts to Nick, Dylan, Andrew, Ryan, Phil, Kathryn, Juni, Patrick, Brenden, MK, Emmet, Alexis, and Matt, all of whom made home an amazing environment to talk about truly anything on my mind. I would also like to give special shoutouts to several friends who continued to help me get through graduate school, including Alistair, Daryl, Elise, Dan, Sam, Renée, Lillian, Jayden, Greg, Matt, and Evan. I would like to give special thanks to Sarah and Nate, Andrew, and Sarah and Michael for being there for me over the past couple of years as we navigated the pandemic. For those whom I may have left off by accident, know that your support has been no less valuable. My endless conversations with all of these friends on all manner of topics both scientific and personal have always made me feel comfortable discussing the challenges we face and figuring out how we

can continue to be better. I would also like to thank John Grotzinger, with whom I have literally ridden hundreds of miles on a bike. John has offered a wealth of interesting insights into the history of science and scientists and is always enthusiastically tackling new challenges head on. I can only hope to emulate that energy for years to come.

It goes without saying that the eternal support that I have received from my family has helped me get this far in my life. To my mom and dad, I cannot thank them enough for their constant love and dedication and for allowing me to take the path of my choosing. All of my victories in life I share with them. From the game nights online to weekend visits up to Santa Barbara, I thank Takako for being a great sister to help take my mind off of school and find a comfortable reprieve out of town. And to her partner and my close friend Josh, thank you for all of your warm support, advice on cycling, and intense focus on optimization, both in games and in life.

I would like to close my acknowledgements by thanking Steve Bloom, who lived larger than life and always defied the odds. But most important to me was his belief that I could always aspire to more and his tireless support to not only help me prepare my written statements for graduate school applications, but also push me to reflect on what I wanted out of my life. Without his help, I may not have ended up at Caltech in the first place.

ABSTRACT

Proteins serve a wide range of functions in and out of the cell, from signaling and gene regulation to transport and structural reinforcement. These functions are usually carried out from interactions with other molecules in the surrounding medium such as other proteins, small molecules, or DNA. One such class of proteins are what I will call polymer-coupling proteins: these proteins intentionally link identical polymers or two regions of the same polymer together so that their coupled interactions critically affect the state of the biological system. A vast array of such proteins exist in nature with roles such as the looping of DNA to physically inhibit the expression of a gene or the formation of the cytoskeleton which provides a cell with its shape. In this thesis, I use *in vitro* experimental methods to explore two cases of coupling proteins and understand their roles not only in reorganizing their complementary polymers but influencing the final state of their respective systems.

In Chapter 2, I examine the starting process for the assembly of an antibody-encoding gene in developing immune cells. Motivated by data suggesting that some antibodies are less likely to be made than others, I explore how the early steps of constructing an antibody-encoding gene affect this uneven frequency of assembly. To initiate recombination, the recombination-activating gene (RAG) protein complex simultaneously binds and cuts two well-recognized sequences neighboring two antibody-encoding gene segments in order to allow other proteins to combine these exposed segments together. The sequences to which the RAG protein performs its binding and cutting functions have certain identifiable sequence patterns but can still vary. Through a single-molecule experimental method known as tethered particle motion (TPM) I show how changes to the binding site sequence can enhance or diminish the propensity of the RAG protein to bind and cut the DNA and thus explore the consequences of these altered interactions in the unequal selection for certain antibody gene segments over others.

In Chapter 3, I turn to questions of the emergence of order from self-organization in biological systems. From the molecular to the population scale, biology constantly demonstrates that with an injection of energy, systems can be driven out of equilibrium and allow for the organization of its constituents. A case of such organization in cells is the coupling of microtubules by motor proteins to

create and maintain the mitotic spindle, a critical biological architecture for ensuring that each cell obtains a copy of the genome during division. *In vitro* experiments that exploit similar motor-microtubule interactions have become a convenient way to identify the effects of perturbing a key player such as motor properties or boundary conditions of the system on the spatiotemporal extent of organization. However, in many instances, the dynamics under which such cytoskeletal systems reduce their entropy over the course of creating order have not been carefully examined in experimental systems. Here, I use engineered light-dimerizable motors that can give rise to the formation of a highly connected network that compacts to form a dense, organized structure, and through the use of a noninvasive imaging technique observe how the polymers that make up the network continually reorganize in the bulk during a global contraction of the network.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] S. Hirokawa, H. J. Lee, R. A. Banks, and R. Phillips, "Grid FRAP patterning reveals a dispersive effect in the bulk of a linearly contracting microtubule network", in preparation (2022).
S.H. set up the experiments, provided reagents, collected data, analyzed the data, developed the theory, and participated in the writing of the manuscript.

[2] R. A. Banks, V. Galstyan, H. J. Lee, S. Hirokawa, A. Ierokomos, T. Ross, Z. Bryant, M. Thomson, and R. Phillips, "Motor processivity and speed determine structure and dynamics of motor-microtubule assemblies", bioRxiv (under review at eLife) (2021), DOI: 10.1101/2021.10.22.465381.
S.H aided in the molecular biology, supplied reagents, analyzed data for a section of the supplemental information, and provided feedback during the writing of the manuscript.

[3] S. Hirokawa, G. Chure, N. M. Belliveau, G. A. Lovely, M. Anaya, D. G. Schatz, D. Baltimore, and R. Phillips, "Sequence-dependent dynamics of synthetic and endogenous RSSs in V(D)J recombination", Nucleic Acids Research 48 (2020), 6726, DOI: 10.1093/nar/gkaa418.
S.H participated in the conception of the project, prepared reagents, performed the molecular biology and single-molecule experiments, analyzed most of the data, and participated in the writing of the manuscript.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Overview

Proteins are critical components for the overall function of biological cells. *E. coli*, budding yeast, and mammalian cells such as HeLa cells have genomes with 4×10^3, 6×10^3, and 2×10^4 protein-coding genes, respectively [1]. Each of these proteins serves a key function in the cell, including but not limited to expression and regulation of genes [2, 3], transport of molecules within or across the cell [4], and signaling to the cell to adapt due to changes in the environment [5].

In many cases, proteins take on the role of linking two similar molecules or environments. Some examples of proteins with coupling properties are highlighted in Fig. 1.1. As shown in Fig. 1.1A, some proteins or protein complexes can bring distinct regions of DNA into close proximity. A classic example of this is the lac repressor found in bacteria, which can simultaneously bind to two sites of a bacterial genome located $\backsim$100 base pairs away to form a loop in the DNA and physically inhibit RNA polymerase from transcribing the downstream gene [6–8], but as will be discussed in this dissertation, other proteins couple more disparate regions of DNA to deliberately cut and paste the genome. Proteins such as mechanosensitive channels [9] or proteins in the electron transport chain [10] couple two different regions separated by the lipid bilayer by sitting on the membrane and directing the flow of molecules into or out of the cell, thereby taking advantage of or driving chemical gradients. And motor proteins can bind to and move along two filaments, allowing the filaments to slide relative to each other and ultimately influence cell shape and locomotion (Fig. 1.1C).

In this dissertation, I examine in greater detail two demonstrations of dynamic restructuring of polymer systems when protein complexes bind them. Chapter 2 examines how the DNA-binding protein that initiates the early stages of gene recombination influences the ability of the cell to construct particular antibodies. In this study, I use a single-molecule assay known as tethered particle motion (TPM) to examine how changes to a particular DNA

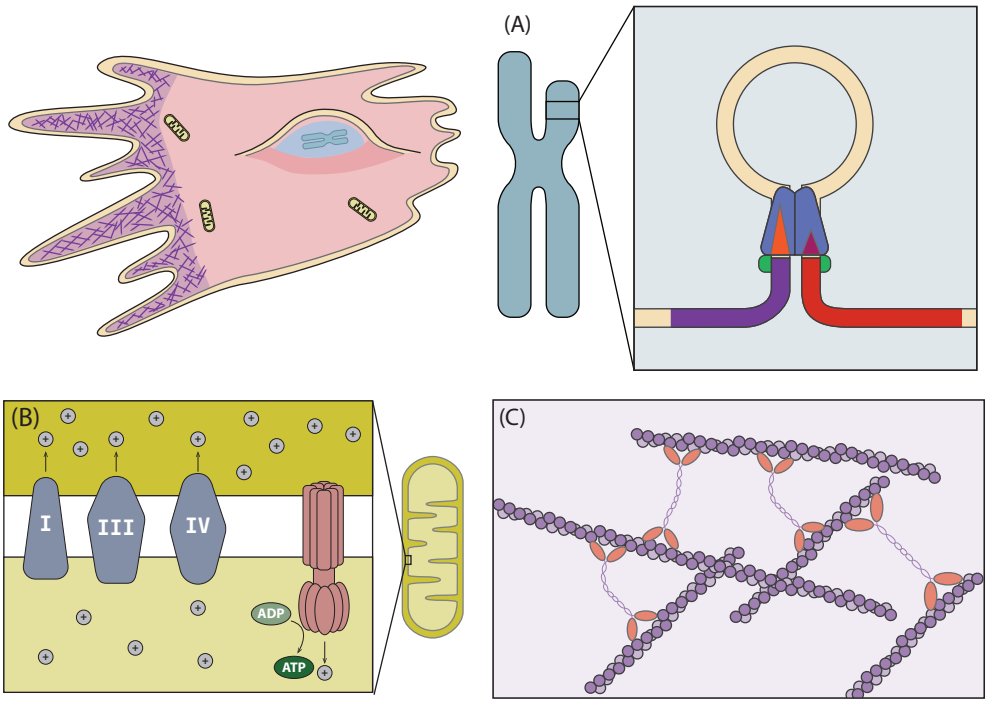


Figure 1.1: **Examples of proteins with coupling functions.** Cells utilize a range of proteins with coupling functions. (A) Some proteins bind to DNA at two sites in order to regulate or even aid in creating genes. (B) Proteins such as those found in the electron transport chain couple two environments to control the import and export of molecules. (C) Other proteins couple the filaments of the cytoskeleton to influence the shape or movement of the cell.

sequence affects the extent to which the recombination-activating gene (RAG) protein complex can bind and cut the DNA. After showing our findings on the sequence-dependent effects of this protein-DNA interaction, I will discuss our results in more detail by relating the data we have collected to known information about the physical mechanism of the binding and cutting effect as well as reflect on how our findings give us a better understanding of the unequal frequency of creating particular antibody-encoding gene combinations.

Chapter 3 looks at the bulk reorganization that occurs when many coupling proteins are interacting with a multitude of the polymers to which these proteins bind. In particular, I look at the case where a collection of motor proteins each containing a cross-linking domain couples a large field of microtubules and transforms the once disorganized microtubule array into a contracted, well-ordered system. Using a microscope constructed and programmed in-house, we examined how microtubules redistribute when the motors that couple and

move the filaments throughout the network macroscopically drive a global contraction of the network. When the filaments are propelled to self-organize, do they couple to new filaments along the way or maintain the same filament neighbors throughout the contraction process?

Before we discuss these two case studies, it is worth doing a deeper dive into the history of both of these respective fields. For the remainder of this chapter, I provide more context for these proteins through the cellular processes that require their functions and key early work done by experts that led to the scientific inquiries that I address in this thesis.

1.2 Antibody production starts with cutting and pasting within the genome.

One of the most fascinating aspects of jawed vertebrates lies in the ability of their immune systems to identify and discard a diversity of invasive bacteria or infected cells within their body. Such a system requires the flexibility to counter a vast array of infectious agents with the speed to quickly contain the threat. While the innate immune system provides the first line of defense and helps identify bacteria based on markers commonly found on the cell surface, other bacteria and viruses that have infected cells in the host organism may slip through the cracks and continue to propagate in the host organism. As a

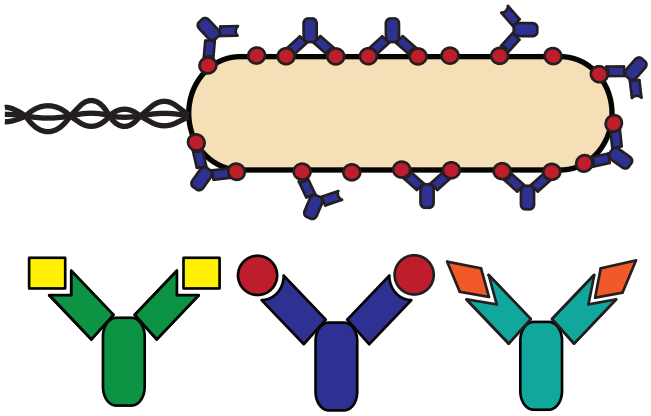


Figure 1.2: **Antibodies allow for the immune cells to identify invasive agents with greater specificity.** Antibodies differ from one another in their ability to bind different small protein markers called antigens on the surfaces of bacteria or infected cells. This identification allows other cells in the immune system to find infectious agents that have been tagged by the antibodies and dispose them appropriately.

result, a more methodical part of the immune system is called upon to produce an antibody that specifically identifies a small $\backsim$10 amino acid long portion of a protein on the surface of the infectious agent (Fig. 1.2). But with 20 different amino acids, there are $20^{10} \approx 10^{13}$ possible 10-amino acid combinations. As an antibody is a protein complex and the human genome only contains $\approx 2 \times 10^4$ protein-encoding genes, the ability to produce an antibody that identifies the particular polypeptide chain seems impossible.

While the idea of an antibody stems as far back as 1890 [11], and antibodies were known to be incredibly diverse in their structure during the coming years, their genetic origins would remain a mystery for over 80 years. The two hypotheses by the 1970s were (a) that the antibody genes are inherited or (b) that the genes are prepared in the immune cells possibly due to large-scale DNA mutation events [12, 13]. In what became an illuminating point for the field, Hozumi and Tonegawa showed through DNA hybridization experiments that whereas two genetic portions of an antibody appear in two spatially disparate places on the chromosome in mouse embryos, once the mouse can produce mature B cells these same genetic portions appear closer together, suggesting a rearrangement of the DNA to bring segments of an antibody-encoding gene into closer proximity [14]. This observation gave rise to a flurry of sequencing results of the genomic regions from which the antibody-encoding genes seemed to originate, revealing fragments of these genes clustered into subregions of the chromosome early on in the development of immune cells that then led to combinations of these fragments in developed B and T cells. Fig. 1.3 demonstrates this clustering of fragmented genetic information for the immunoglobulin κ (Igκ) locus on chromosome 2 in humans [15]. The growing evidence pointed to the idea that rather than having antibody genes readily transcribed, which would either lead to an unwieldy genome size or a set of antibodies with insufficient coverage, the genome contains gene segments that encode a portion of the antibody. By cutting and pasting gene segments together within the genome, a developing immune cell can in principle create an antibody combination that uniquely identifies the $\backsim$10 amino acid long marker. This process became known as V(D)J recombination.

As illustrated in Fig. 1.4, an antibody is composed of two heavy chains (the four-block pieces that make up the bottom stem and inside arms of the Y shape in the middle of the antibody) and two light chains (the two-block pieces on

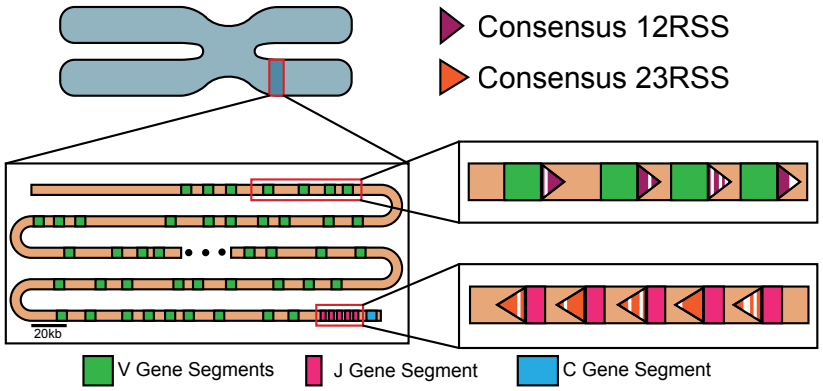


Figure 1.3: **Arrangement of antibody-encoding gene segments in the Igκ locus on chromosome 2 in humans.** The gene segments that make up an antibody were found to be fragmented in embryos and clustered according to the part of the antibody that they encode. Here, the variable (V) gene segments (green) are spread across roughly 1.7 million base pairs (Mbp) with their compatible joining (J) gene segments (pink) clustered in about a 100 thousand base pair (kbp) region. The constant (C) gene segment (blue) specifies the functional form of the antibody, such as for secretion or on the surface of a mature B cell. Each gene segment was found to have a particular sequence pattern adjacent to them. In the case of the Igκ locus in humans, the 12RSSs (purple) are adjacent to the V gene segments while 23RSSs (orange) are adjacent to the J gene segments. Dotted lines denote a large portion of the DNA region where no gene segments are found. Scale bar is meant to convey the general position of gene segments and not gene segment or RSS size. Schematic adapted from the ImMunoGeneTics® database [16, 17].

the outsides of the Y arms). These two components come from their respective heavy and light chain genes. The heavy chain gene requires the recombination of a variable (V) gene segment, a diversity (D) gene segment, and a joining (J) gene segment. The light chain gene requires the recombination of one V gene segment with a J gene segment found in a different part of the genome from the heavy chain. As there exist multiple V, D, and J gene segments in the chromosome for heavy and light chains as shown in Table 1.1, the number of combinations one can create rapidly expands. For example, there are nearly 6000 heavy chain gene combinations and nearly 300 light chain gene combinations (35 Vκ gene segments $\times$ 5 Jκ gene segments + 30 Vλ gene segments $\times$ 5 Jλ gene segments). By combining one heavy chain gene with one light chain gene, one can come up with roughly 1.8×10^6 antibodies by this set of combinations alone, which already exceeds the number of genes readily available in the genome. In addition, gene segments are imprecisely joined, with up to

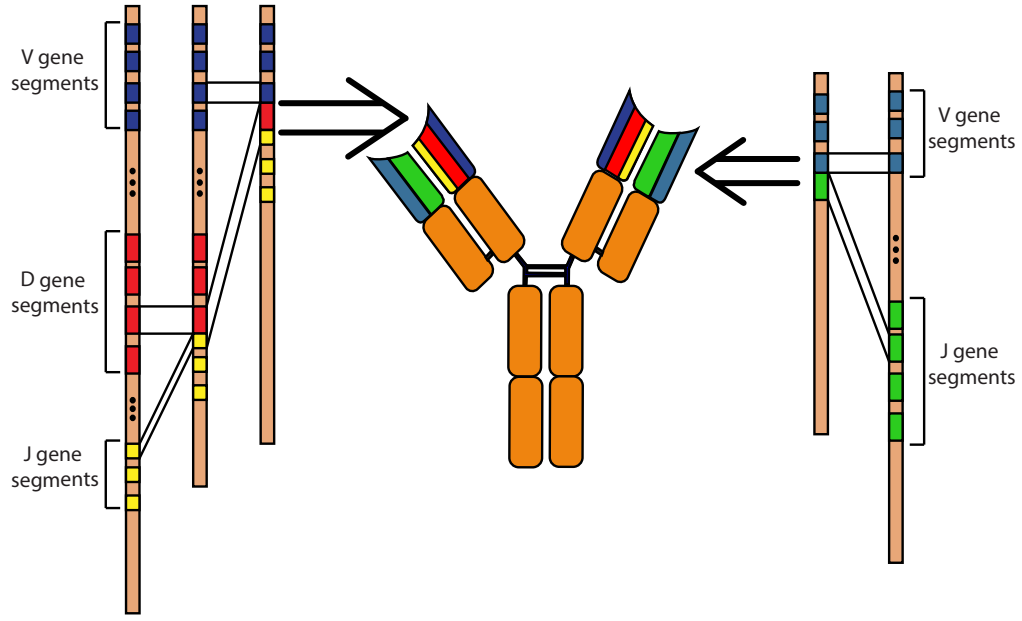


Figure 1.4: **Schematic of antibody gene encoding by V(D)J recombination.** Recombination occurs twice on one of the chromosomes to produce the heavy chain gene by combining a V (blue), D (red), and J (yellow) gene segment and once on another chromosome to produce the light chain gene by combining a V (aqua) and J (green) gene segment. Orange parts of the antibody come from constant gene segments and specify features of the antibody type such as whether it sits on the immune cell membrane or are secreted.

twenty nucleotides inserted at the junction, which can give rise to a few extra amino acids and thus a greater range of diversity in the gene segments. If on average about three amino acids worth of sequences are inserted between gene segments, this drives the number of possible antibodies to nearly 10^{10} possibilities! Throughout this recombination process, the cell passes through a series of intermediate checks that helps ensure that the recombined genes produce a functional protein and do not harm the host organism. Upon completion of all recombination events, the generated antibody is presented against a selected antigen containing the $\backsim$10-amino acid long marker. If the antibody binds sufficiently well to the antigen, the genes encoding the antibody regions are mutated at a higher rate to strengthen the antigen binding. A key piece that reinforced the idea of gene rearrangement was the identification of a pattern of sequences that consistently appeared adjacent to the gene segments of light chains in the embryonic mouse genome [19]. In particular, a well-conserved

seven base pair sequence (CACTGTG) was found immediately adjacent to the J gene segments and ten base pair sequence (GGTTTTTGTA) nearly 30 base pairs away, with greater variability in the "spacer" sequence in between. Similarly, the reverse complement of the conserved heptamer (CACAGTG) was found immediately adjacent to the V gene segment, suggesting a genomic recognition site for some enzyme to perform the recombination in a targeted fashion. These sequences, which would be later termed recombination signal sequences (RSSs), can be found in Fig. 1.3. These RSSs were found adjacent to heavy chain gene segments and upon closer inspection revealed that the sequence pattern appeared to involve the well conserved heptamer noted previously, a well conserved *nine* base pair sequence further away from the gene segment, and either a more variable 12- or 23-base pair long spacer sequence (± 1 base pair) in between depending on the gene segment it neighbors [20]. Of further note was the observation that in order to get the corrected VJ or VDJ combinations for the light and heavy chain genes, respectively, the recombination would typically involve combining a gene segment adjacent to an RSS containing a 12-bp spacer (12RSS) with one adjacent to an RSS containing a 23-bp spacer (23RSS), leading to what became known as the 12/23 rule.

Within a few years, the mysterious recombinase was revealed to involve two proteins, first the recombination-activating gene-1 (RAG-1) [21] followed shortly after by RAG-2, which enhanced recombination activity by at least three orders of magnitude when expressed in combination with RAG-1 [22]. In later works, the function of the RAG recombinases as coupling agents was revealed when RAG-1 was determined to be able to bind to the RSS sites, with stronger binding in the presence of RAG-2 [23], and eventually cleave the RSSs away

Table 1.1: **Number of functional gene segments in human immune system.** Gene segments are distinguished by whether they are part of the light or heavy chain gene and further classified by whether they belong in the variable (V), diversity (D), joining (J), constant (C) gene segment family. λ and κ denotes the two types of light chains, found on different loci and each containing their own set of V and J gene segments that combine exclusively within their designated loci. Table adapted from [18].

Segment	**Light chains**	**Heavy chain**
Variable (V)	35 (κ); 30 (λ)	40
Diversity (D)	0	23
Joining (J)	5 (κ); 5 (λ)	6

to expose the gene segments [24]. Other proteins were then found to be involved in holding and joining the gene segments together. Fig. 2.1 in Chapter 2 outlines the steps of the RAG-RSS interaction that initiates V(D)J recombination. Thus, the RAG1/2 complex was found to simultaneously bind and cleave a 12RSS and a 23RSS, revealing the role of the recombinase in removing large sections of genomic DNA within developing immune cells in pursuit of producing an antibody-encoding gene.

While the range of heavy and light chain gene segments have been well identified in the genomes of various standard jawed vertebrate organisms, including in mice, zebrafish, and humans, their frequency of usage as part of a complete antibody-encoding gene was found to differ widely. Large-scale sequencing studies performed on zebrafish and in mice revealed that some gene segments from the same locus were used in high frequency while others were almost never selected [25–27]. These findings suggested that not all antibodies are equally likely to be produced in the immune system, but the factors that determine the preference for some gene segments over others was not clear amid the myriad of players involved throughout V(D)J recombination. When considering the myriad of players that start V(D)J recombination and how they might affect antibody gene segment selection, a possible candidate is the RSSs adjacent to the gene segments. Indeed, early work suggested that even a single point mutation to an RSS might dramatically affect the production of the right antibody, a finding with unfortunate health implications for some people [28]. Fig. 1.5 outlines two hypotheses of how the RSSs may affect gene segment selection. For one, gene segments and by extension their partner RSSs can be spread across millions of base pairs suggesting a spatial dependence, leading to a diagram suggesting that the arrangement of gene segments in the locus influences their chances of being selected [26]. In addition, RSS sequence may affect the actual binding and cutting function of RAG and thus cause the first two steps of V(D)J recombination to be enhanced or hindered depending on the sequence.

RSSs were known to exhibit some sequence variance ever since their discovery [29]. Fig. 1.6 further highlights this point through an examination of the range of 12- and 23RSSs that are found in the mouse genome. Fig. 1.6A highlights the number of nucleotide mismatches that can be found between heptamers (left column) and nonamers (right column) of naturally-occurring

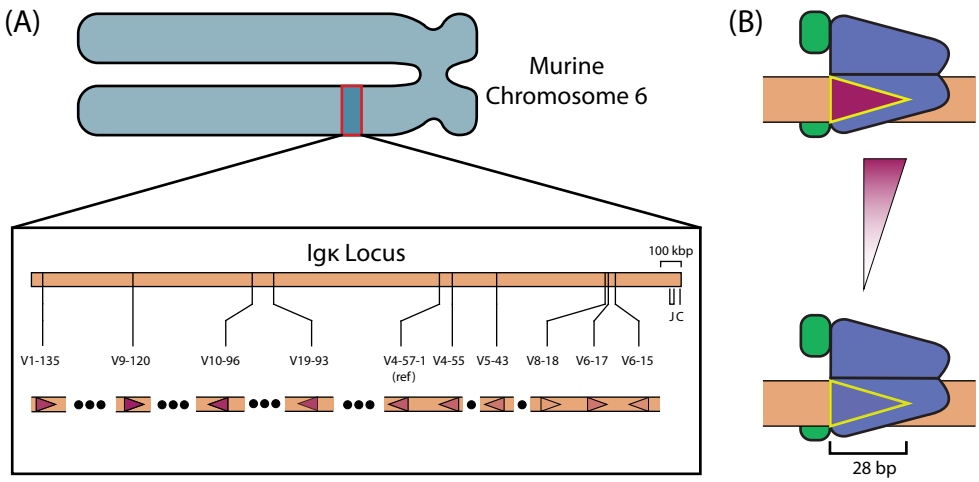


Figure 1.5: **Outline of different gene segments in the Igκ locus in chromosome 6 in mice.** (A) Gene segments can span about a few million base pairs, as shown for the Igκ locus in chromsome 6 in mice. (B) Each gene segment has a unique RSS sequence that can affect RAG binding and cutting. The triangles shown represent RSSs with the arrows pointing away from the gene segment that would sit immediately adjacent to them. Darkness of purple denotes the gene segment usage as found in [26] to suggest gene segment usage frequency may be tied to binding or cutting strength of RAG onto the RSS.

(endogenous) 12- (top row) and 23RSSs (bottom row) against their consensus sequence. Many nonamers can differ from their corresponding consensus sequences by one to three base pairs with some even differing by as much as six base pairs. Fig. 1.6B highlights the variation in 12RSS sequences at the single nucleotide level. With each position highlighted along the position axis, the larger the letter, the more frequently that nucleotide is found at that position among 12RSSs in the mouse genome. Here, one can more clearly identify the well-conserved heptamer sequence 'CACAGTG' at the start and similarly find the conserved nonamer 'ACAAAAACC' toward the end of the sequence, with a more variable spacer sequence. Bulk assays involving V(D)J recombination performed on plasmids rather than in the chromosome indicate that changes to either of the RSS sequences could affect the extent of recombination performed [29–31], but these assays made it difficult to determine the interaction between RAG and the RSSs. Does changing the RSS sequence affect RAG binding, cutting, or both? Do these changes in interaction depend on the position where a change in the RSS sequence is made? Furthermore, many of the bulk assays performed do not provide much insight into the dynamics of the RAG-RSS interaction, making it unclear how long RAG remains bound to

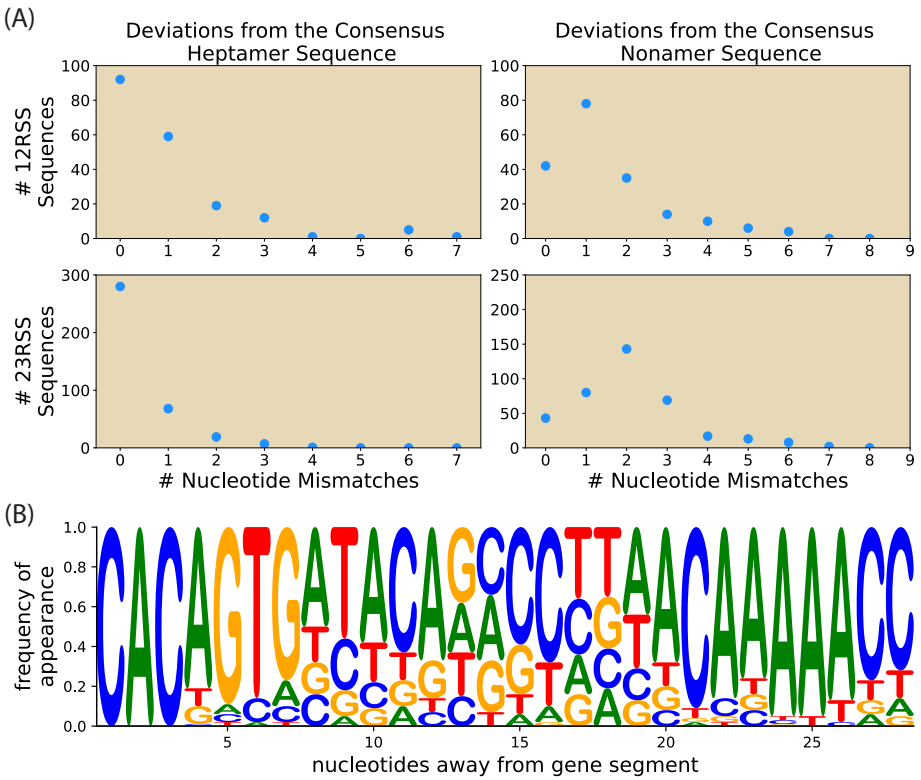


Figure 1.6: **Deviations in endogenous RSS heptamer and nonamer sequences from the consensus sequences in the mouse genome.** (A) Histograms of number of base pairs that each heptamer (left column) and nonamer (right column) deviates from their respective consensus sequences, with distinctions between 12RSSs (top row) and 23RSSs (bottom row). (B) Census of the 12RSS. Height of letters scales with the fraction of RSSs that have that nucleotide at the position. Sequences obtained using the ImMunoGeneTics® database [16, 17].

both 12- and 23RSS before cutting, or whether RAG commits to cutting the DNA once it is bound to a 12- and 23RSS. It was after the establishment of a single-molecule assay in which RAG-RSS interactions could be observed in real time [32] that the sequence-dependent effects of the RSS on the propensity of RAG to bind and cut the DNA could be scrutinized and by extension the subsequent impact on antibody gene segment selection could be better understood. This examination of the RSS sequence effect on RAG-RSS dynamics is elaborated in Chapter 2 with supplementary information in Appendix A.

1.3 Working as a collective to create organization

Coordinated self-organization occurs across vast scales in biology, from the molecular to the ecological. At the population level, there is the flying V formation of geese as they migrate for the cold season (Fig. 1.7A) or the line of ants directed toward a sink full of dirty dishes. At the level of individual organisms, the skin of tropical fish or the dazzling displays of butterfly wings suggest some careful coordination among constituent cells (Fig. 1.7B). And at the cellular and molecular levels, cells spatiotemporally compartmentalize molecules or functions (Fig. 1.7C); the genomic information in eukaryotes is neatly compacted down into multiple chromosomes; and the mitotic spindle emerges during cell division (Fig. 1.7D). The list goes on.

Biology with its abundant examples of order and organization offers some of the best opportunities to study the emergence of this phenomenon and has subsequently captured the curiosities of biologists, mathematicians, and physicists alike. Alan Turing, though more noted for his contributions in computer science and codebreaking, showed mathematically that a series of patterns can emerge when multiple chemical species responsible for this patterning diffuse and react with each other [33], which was confirmed through numerical simulations complementing experimental results for the stripes found on zebrafish [34, 35]. Tamás Vicsek showed that self-organization of initially disordered particles can be minimally simulated by adjusting the trajectory of each individual based on the orientations of its nearest neighbors [36]. And investigators like John Toner and Yuhai Tu have been carefully developing quantitative models that describe the collective organization of a population such as the flocking of birds [37, 38].

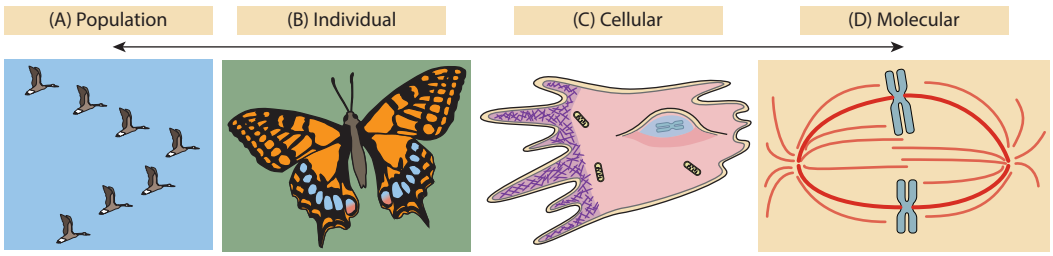


Figure 1.7: **Scales of patterning in biology.** The patterns that come from self-organization in biology can be found from the ecological level down to the molecular level. Examples range from (A) the migration of geese, (B) the patterning on butterfly wings, (C) the organization within a cell, and (D) the production and maintenance of the mitotic spindle within cells.

Over the past 20-30 years, experimentalists have been devising clever ways of probing and perturbing such active systems across these size scales. At the population level once again, work has been done to study the emergence of bird flocks in the sky by setting up cameras in a field [39]. Similarly, at the smaller scale, efforts are made to examine systems exhibiting self-organization in the lab. In some cases, these assays utilize granular rods or particles that are stimulated by mechanical or chemical energy [40, 41], while other works involve biological elements such as a collection of cells [42] or, as will be relevant for the remainder of this chapter, biomolecules produced in cells. By extracting cytoskeletal filaments, motor proteins that move along the filaments, and chemical energy such as ATP from the cell and mixing them in the right combinations, one can immediately begin to observe self-organization. To better understand the types of organized patterns found in these systems, it is worth touching on the relevant properties of the molecules involved.

Two well-studied cytoskeletal filaments are actin and microtubules, both of which serve equally vital but non-overlapping purposes in an organism. Actin is involved in the division, motility, or shape of a cell while microtubules act as causeways along which certain molecules can be actively transported or as part of the mitotic spindle that divides chromosomes once they have been duplicated prior to cell division. While actin and microtubules differ by their physical properties, such as their diameters (for actin it is a few$\backsim$10 nm in diameter while for microtubules it is $\approx$25 nm [43]) or their flexibility (actin has a persistence length of $\approx$10 µm [44] while a microtubule has a persistence length in the millimeter range [45]), two of their main similarities are (1) that their length changes based on the addition or removal of the individual monomers of which they are composed and (2) that they have a built-in asymmetry that allows for a specificity of orientation along their long axis. This asymmetry distinguishes the two ends of the filaments, with one end called the plus (+) end and the other the minus (-) end. This asymmetry is relevant not only for the preferential growth or shrinkage occurring at one end or another [46] but also with respect to the motors that walk along them.

A discussion about actin or microtubules inevitably involves mentioning motor proteins. Motor proteins bind and move unidirectionally along the aforementioned filaments through the consumption of chemical energy in the form of ATP. In doing so, they can transport cargo in a directed manner in the cell or

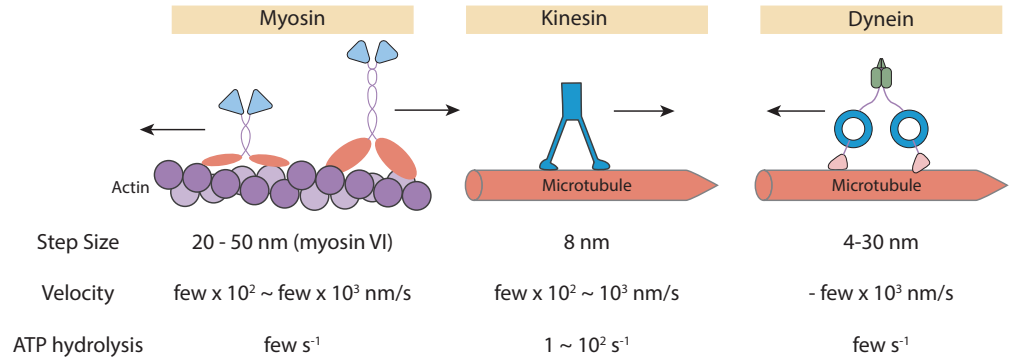


Figure 1.8: **Comparison of myosin, kinesin, and dynein motors.** While myosin motors walk along actin filaments, kinesin and dynein walk along microtubules. The direction that the myosin motor walks along the filament depends upon the specific motor while dynein motors process toward the minus end of microtubules and the majority of kinesin motors move toward the plus end of microtubules. Though variable depending on the specific motor, rough values for the step size, velocity, and ATP hydrolysis rate of each motor family are provided as obtained by [52].

help drive shape changes in the cell by generating forces on and subsequently sliding filaments relative to each other [47]. While there are a multitude of motors each with their own speed, direction that they move along their complementary filament, and rate of ATP consumption, they are typically broken down into three groups. Myosin motors move along actin filaments and depending on the type of myosin motor either move toward the plus end or the minus end. One of the more well known examples of myosin is myosin II which drives muscle contractions by pulling actin filaments toward each other [48, 49]. On the other hand, kinesin and dynein motors move along microtubules. While dynein motors move toward the minus end, most of the known kinesin motors move toward the plus end of microtubules, though some exceptions are known to exist such as kinesin-14, also commonly known as Ncd [50, 51]. These motor classes are summarized in Fig. 1.8.

The ability to take the filaments, corresponding motor proteins, and ATP out of the cell, mix them in a tube, and directly image these mixtures has opened up a new avenue for controllably studying the emergence of order in an energetically-driven population in the laboratory. In particular, early work showed that by designing kinesin so that they linked to one another and mixing these motors with microtubules and ATP, the microtubules could become coupled to each other through the multimerized motors and form a connected

network that could generate organized structures such as those found in Fig. 1.9 [53]. Depending on the concentration of motors, two patterns predominantly emerged: vortices where microtubules circulate about a common center (Fig. 1.9C) and asters where microtubules locally point toward a common center (Fig. 1.9D). By increasing or decreasing the concentration of motors by as little as 50%, one can generate only asters or only vortices, respectively. Follow up work demonstrated that by mixing motors that walk in one direction with those that move in the opposite direction and carefully adjusting the amount of both motor types and microtubules, a broader spectrum of patterns emerges [54]. While concentrations of motors and microtubules were tunable

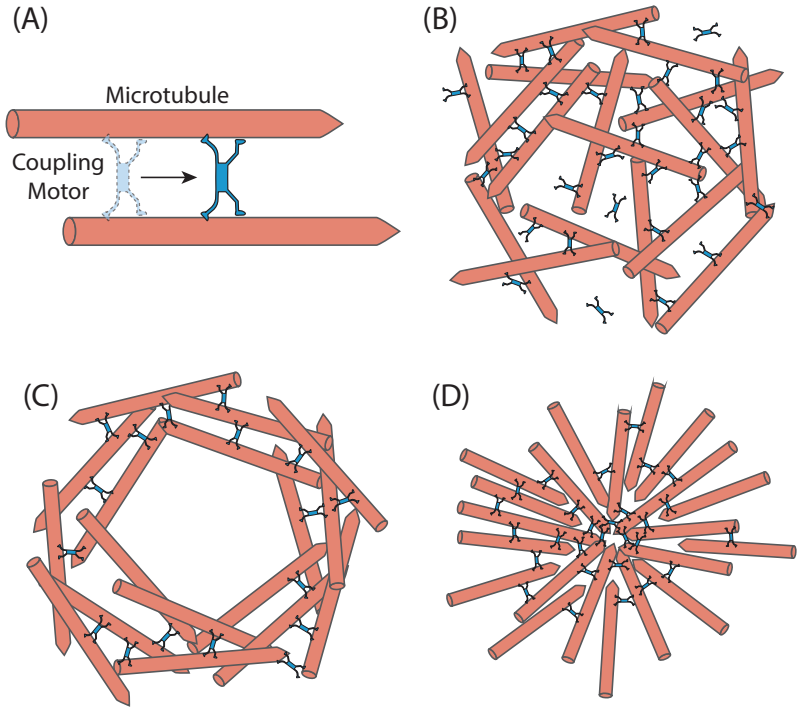


Figure 1.9: **Vortex and aster formed in *in vitro* motor-microtubule mixture.** (A) Motors were constructed such that they would multimerize, the result of which allowed them to simultaneously walk along two or more filaments. (B) Early experimental assays mixed these multimerized motors with microtubules and ATP (not shown), the chemical energy that allows the motors to traverse along the microtubules. (C) In some areas of the network, the motors induced vortex patterns where the microtubules circulated about a common center. (D) In other areas, motors generated asters where the microtubules pointed toward the common center. Lowering the concentration of motors caused vortices to predominantly emerge while raising the concentration made asters feature more prominently in the assay.

parameters, such assays could in principle produce different responses from modifying a range of other parameters, such as motor type (which could relate to processivity or speed) or the use of microtubules with dynamically changing length, thereby demonstrating the immense potential of *in vitro* active assays as testing grounds for better understanding the emergence of order.

Today, the list of ordered structures that have been observed and documented has expanded since these initial *in vitro* experiments through the development of other self-assembly assays [55–58] and in recent years have been found in *in vivo* systems ranging from morphogenesis to population-level organization [42, 59]. Furthermore, in much the same way that Turing thought carefully about how chemical diffusion and reactions can induce pattern formation or that Vicsek or Toner and Tu considered the theory of flocking, the different *in vitro* experimental efforts have motivated several pursuits through computer simulations and quantitative models to better understand the different ordered structures that emerge in these motor-filament systems. Agent-based simulations such as Cytosim and more recently aLENS show the time evolution of a cytoskeletal system based on user-defined details about the molecular players including the appropriate kinetic rates of filaments and crosslinkers, their initial spatial positions and orientations, and the conditions of the environment in which the particles reside [60–64]. In the case of developing quantitative models, while some theoretical work relies on identifying the microscopic rules of the molecular players in their simplest forms [62], many models take a more coarse-grained approach. Examples include spatiotemporal coupling of the motor density with the density or orientation field of the microtubule [65–68], spring-like models [69], hydrodynamic models [70, 71] or continuum mechanical models [72].

As will be more closely explored in this thesis, one behavior that occurs in many of these self-organization assays is a contraction of the network. In this case, filament arrays that are initially uniformly distributed and randomly oriented are combined with the motors that couple them, leading to large-scale reorganization of the connected array to create spatial inhomogeneities and locally concentrate. While work by Surrey and Nédélec suggest a local contraction of the network promoted the spatially nonuniform distributions of filaments and motors through the asters and vortices in their original work [53, 54], one of the biggest technical challenges of these self-organization assays

was the inability to observe the early stages of the self-organization process. Once the multimerized motors, microtubules, and energy source were mixed together, the reaction would begin before the sample could be mounted onto a microscope. In time, new clever assays would be designed that provided the user with better control over when and where self-organization could occur. One of the earliest implementations of control was in adding a component known as blebbistatin to inhibit movement of muscle myosin among a highly connected actin filament network [69]. Blebbistatin could be inactivated by shining blue light onto it, allowing the myosin motor to move along the filaments. By controlling the location and intensity of light on the sample, illuminated myosin motors move and deform the network while the unilluminated, blebbistatin-inhibited motors remain anchored in place along the actin filaments.

Controlling the start of the organization process can alternatively be achieved by controlling multimerization of the motors. In this case, motors can be designed to include one of two complementary pieces that link to each other when illuminated with blue light. One of these two additional components, called the improved light inducible dimer (iLid), undergoes a conformational change by light stimulation that allows for binding to its complement, known as micro [73]. In doing so, until an iLid-tagged motor and a micro-tagged motor dimerize through blue light illumination the two filaments remain uncoupled and organization does not occur. Fig. 1.10 illustrates this motor dimerization and filament coupling. With this assay, mixing of motors and microtubules in darkness allows for motors to remain separate and thus microtubules to remain uncoupled, preventing self-organization from starting. The assay can then be illuminated using a light projector displaying patterns of the users choosing to spatially limit where self-organization occurs. Simple geometries such as a circle reveal a contraction of the coupled filaments away from the unilluminated individual filaments to create aster-like structures with a dense core of filaments toward the center and radial splay of filaments at the edges [74]. Projecting different light patterns onto the network has also allowed for the ability to move these organized structures or to bring originally disparate asters together. With the ability to spatiotemporally control organization in the assay, one can observe the emergence and full trajectory of such self-assembly systems, thereby opening up a set of new questions about the initiation of and tendency toward order in inherently out-of-equilibrium systems. Fig. 1.10C-E

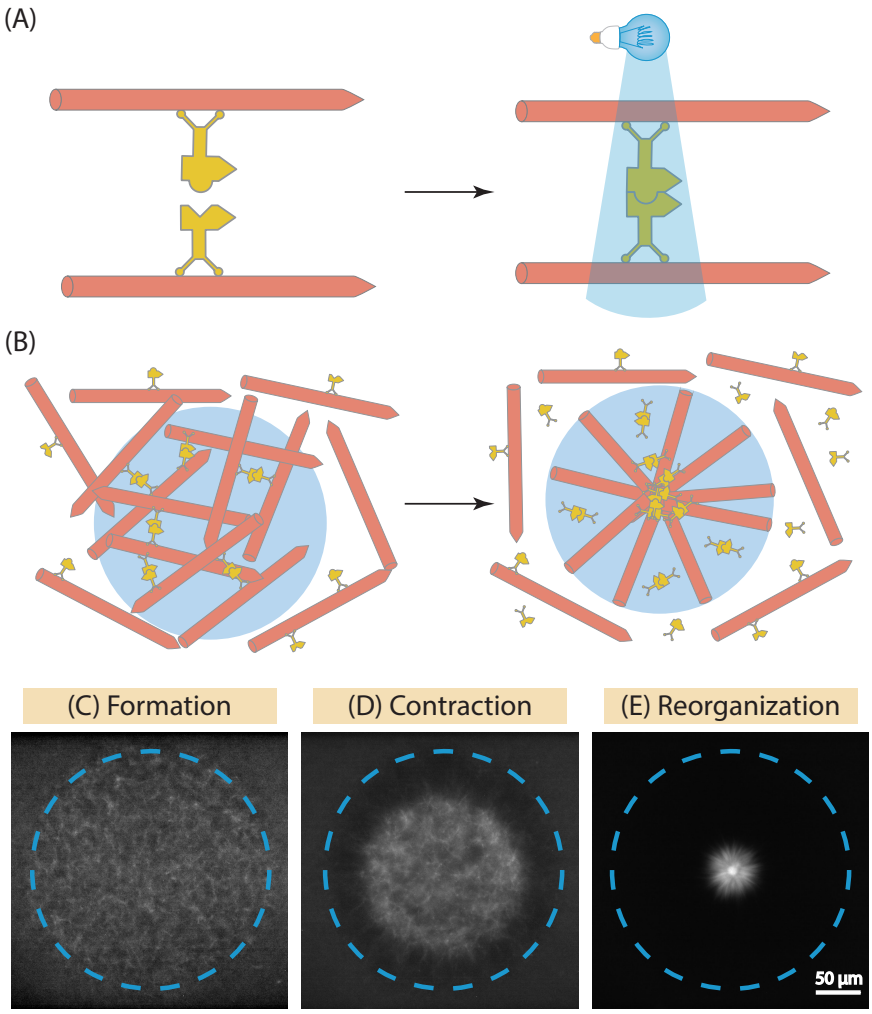


Figure 1.10: **Controlling filament coupling by light-induced motor dimerization.** (A) Motors are designed to include either an iLid or micro domain. When motors are exposed to blue light, the iLid domain undergoes a conformational change that allows for micro binding. The dimerized motors can then couple the movement of filaments. Note that motor binding onto microtubules is not a prerequisite for the motors to dimerize upon light-activation. (B) Projecting different light patterns on a collection of motors and microtubules spatially defines the regions where self-organization occurs. (C-E) Microscopy images of the microtubules under the light-activation scheme from projecting a circular pattern of light onto the network as outlined by the blue dashed circles. These images highlight the three qualitative phases that occur: (C) initial formation of the microtubule network by the dimerized motors, (D) contraction of the network away from the uncoupled reservoir of microtubules, and (E) final reorganization to create an aster-like structure. Image contrast differ for each image to more easily visualize features of the illuminated microtubule network. Ncd motors used in the assay not shown.

show sample microscopy images of the microtubule network during the initial network formation, network contraction, and final reorganization phases of the connected array.

There remain many fascinating questions in the field of such energetically-driven self-organizing systems (more commonly referred to as active matter systems) from the theoretical to the empirical that assays such as those mentioned so far have begun to address. How does the availability of energy influence the formation of structures, particularly if energy is a limited resource [75]? What are the forces that these dimerized motors exert on the network to locally reorient and redistribute filaments [76] and how can they be tied back to the energetics of the system? How does the speed and processivity of motors affect the size and distribution of the microtubule network [77]? What are the critical steps that help an initially disordered array transition to a contracting network? In Chapter 3 of this dissertation, we ask the question of how the coupled filaments as driven by motors under the iLid-micro dimerization scheme redistribute amongst themselves on their way to creating various structures. Put another way, as a network of microtubules begin to self-organize and contract toward a common center, for how long does each filament maintain the same nearest neighbors and to what extent is it encountering new individuals? To do so, we use a technique in fluorescence microscopy to visually (but not physically) remove some of the microtubules and see how the remaining visualizable microtubules move in the network. Based on these time lapse images, we can measure coarse-grained properties of the filament network and generate hypotheses for the behaviors that occur. To test these hypotheses, we then perturb the system using motors of different speeds to see if the coarse-grained properties change and compare the experimentally observed behaviors to a hypothesized theoretical model. These experimental perturbations and comparisons to theoretical predictions offer us a better understanding of the mechanism driving the observed microtubule distribution during the contraction phase, which we discuss at the end.

BIBLIOGRAPHY

[1] R. Milo and R. Phillips, *Cell Biology by the Numbers*, Garland Science, 2016.

[2] N. M. Belliveau et al., "Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria", Proc Natl Acad Sci 115 (2018), E4796.

[3] W. T. Ireland et al., "Deciphering the regulatory genome of *Escherichia coli*, one hundred promoters at a time", eLife 9 (2020), e55308.

[4] T. G. Fai et al., "Length regulation of multiple flagella that self-assemble from a shared pool of components", eLife 8 (2019), e42599.

[5] P. R. Taylor et al., "Macrophage receptors and immune recognition", Annu Rev Immunol 23 (2005), 901.

[6] S. Oehler et al., "The three operators of the *lac* operon cooperate in repression", EMBO J 9 (), 973.

[7] S. Johnson, M. Linden, and R. Phillips, "Sequence dependence of transcription factor-mediated DNA looping", Nucleic Acids Res 40 (2012), 7728.

[8] S. Johnson et al., "Multiple LacI-mediated loops revealed by Bayesian statistics and tethered particle motion", Nucleic Acids Res 42 (2014), 10265.

[9] G. Chure et al., "Connecting the dots between mechanosensitive channel abundance, osmotic shock, and survival at single-cell resolution", J Bacteriol 200 (2018), e00460.

[10] R.-Z. Zhao et al., "Mitochondrial electron transport chain, ROS generation and uncoupling", Int J Mol Med 44 (2019), 3.

[11] J. Lindenmann, "Origin of the terms 'antibody' and 'antigen'", Scand J Immunol 19 (1984), 281.

[12] H. Wigzell, "Antibody diversity: Is it all coded for by the germ line genes?", Scand J Immunol 2 (1973), 199.

[13] J. D. Capra and J. M. Kehoe, "Antibody diversity: Is it all coded for by the germ line genes?", Scand J Immunol 3 (1974), 1.

[14] N. Hozumi and S. Tonegawa, "Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions", Proc Natl Acad Sci 73 (1976), 3628.

[15] S. Tonegawa, "Somatic generation of antibody diversity", Nature 302 (1983), 575.

[16] M.-P. Lefranc, "Immunoglobulin (IG) and T cell receptor genes (TR): IMGT® and the birth and rise of immunoinformatics", Front Immunol 5 (2014).

[17] M.-P. Lefranc et al., "IMGT®, the international ImMunoGeneTics information system® 25 years on", Nucleic Acids Res 43 (2015), D413.

[18] K. Murphy, *Janeway's Immunobiology*, 8th ed., Garland Science, 2012.

[19] H. Sakano et al., "Sequences at the somatic recombination sites of immunoglobulin light-chain genes", Nature 280 (1979), 288.

[20] H. Sakano et al., "Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes", Nature 286 (1980), 676.

[21] D. G. Schatz, M. A. Oettinger, and D. Baltimore, "The V(D)J recombination activating gene, Rag-1", Cell 59 (1989), 1035.

[22] M. A. Oettinger et al., "RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination", Science 248 (1990), 1517.

[23] M. J. Difilippantonio et al., "RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination", Cell 87 (1996), 253.

[24] J. F. McBlane et al., "Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps", Cell 83 (1995), 387.

[25] J. A. Weinstein et al., "High-throughput sequencing of the zebrafish antibody repertoire", Science 324 (2009), 804.

[26] M. Aoki-Ota et al., "Skewed primary Igκ repertoire and V-J joining in C57BL/6 mice: Implications for recombination accessibility and receptor editing", J Immunol 188 (2012), 2305.

[27] N. M. Choi et al., "Deep sequencing of the murine Igh repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies", J Immunol 191 (2013), 2393.

[28] A. J. Feeney et al., "A defective Vkappa A2 allele in Navajos which may play a role in increased susceptibility to Haemophilus influenzae type b disease", J Clin Invest 97 (1996), 2277.

[29] D. A. Ramsden, K. Baetz, and G. E. Wu, "Conservation of sequence in recombination signal sequence spacers", Nucleic Acids Res 22 (1994), 1785.

[30] J. E. Hesse et al., "V(D)J recombination: A functional definition of the joining signals", Genes Dev 3 (1989), 1053.

[31] Y. Akamatsu et al., "Essential residues in V(D)J recombination signals", J Immunol 153 (1994), 4520.

[32] G. A. Lovely et al., "Single-molecule analysis of RAG-mediated V(D)J DNA cleavage", Proc Natl Acad Sci 112 (2015), E1715.

[33] A. M. Turing, "The chemical basis of morphogenesis", Phil Trans R Soc Lond B 237 (1952), 37.

[34] S. Kondo and T. Miura, "Reaction-diffusion model as a framework for understanding biological pattern formation", Science 329 (2010), 1616.

[35] S. Kondo, M. Watanabe, and S. Miyazawa, "Studies of Turing pattern formation in zebrafish skin", Phil Trans R Soc A 379 (2021), 20200274.

[36] T. Vicsek et al., "Novel type of phase transition in a system of self-driven particles", Phys Rev Lett 75 (1995), 1226.

[37] J. Toner and Y. Tu, "Long-range order in a two-dimensional dynamical XY model: How birds fly together", Phys Rev Lett 75 (1995), 4326.

[38] J. Toner and Y. Tu, "Flocks, herds, and schools: A quantitative theory of flocking", Phys Rev E 58 (1998), 4828.

[39] A. Cavagna and I. Giardina, "Bird flocks as condensed matter", Annu Rev Cond Matt Phys 5 (2014), 183.

[40] N. Kumar et al., "Flocking at a distance in active granular matter", Nat Commun 5 (2014), 4688.

[41] A. Kudrolli et al., "Swarming and swirling in self-propelled polar granular rods", Phys Rev Lett 100 (2008), 058001.

[42] K. Copenhagen et al., "Topological defects promote layer formation in *Myxococcus xanthus* colonies", Nat Phys 17 (2021), 211.

[43] B. Sodeik, "Mechanisms of viral transport in the cytoplasm", Trends Microbiol 8 (2000), 465.

[44] H. Isambert et al., "Flexibility of actin filaments derived from thermal fluctuations. Effect of bound nucleotide, phalloidin, and muscle regulatory proteins", J Biol Chem 270 (1995), 11437.

[45] J. van Mameren et al., "Leveraging single protein polymers to measure flexural rigidity", J Phys Chem B 113 (2009), 3837.

[46] L. Mohapatra et al., "Design principles of length control of cytoskeletal structures", Annu Rev Biophys 45 (2016), 85.

[47] B. Alberts et al., *Molecular Biology of the Cell, Sixth Edition*, Garland Science, 2015.

[48] A. F. Huxley and R. Niedergerke, "Structural changes in muscle during contraction: Inference microscopy of living muscle fibers", Nature 173 (1954), 971.

[49] H. Huxley and J. Hanson, "Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation", Nature 173 (1954), 973.

[50] H. B. McDonald, R. J. Stewart, and L. S. B. Goldstein, "The kinesin-like *ncd* protein of Drosophila is a minus end-directed microtubule motor", Cell 63 (1990), 1159.

[51] R. A. Walker, E. D. Salmon, and S. A. Endow, "The Drosophila claret segregation protein is a minus-end directed motor molecule", Nature 347 (1990), 780.

[52] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton*, Sinauer Associates, Inc., 2001.

[53] F. J. Nédélec et al., "Self-organization of microtubules and motors", Nature 389 (1997), 305.

[54] T. Surrey et al., "Physical properties determining self-organization of motors and microtubules", Science 292 (2001), 1167.

[55] T. Sanchez et al., "Spontaneous motion in hierarchically assembled active matter", Nature 491 (2012), 431.

[56] F. C. Keber et al., "Topology and dynamics of active nematic vesicles", Science 345 (2014), 1135.

[57] L. Huber et al., "Emergence of coexisting ordered states in active matter systems", Science 361 (2018), 255.

[58] A. Sciortino and A. R. Bausch, "Pattern formation and polarity sorting of driven actin filaments on lipid membranes", Proc Natl Acad Sci 118 (2021), e2017047118.

[59] Y. Maroudas-Sacks et al., "Topological defects in the nematic order of actin fibres as organization centres of *Hydra* morphogenesis", Nat Phys 17 (2021), 251.

[60] F. Nédélec, "Computer simulations reveal motor properties generating stable antiparallel microtubule interactions", J Cell Bio 158 (2002), 1005.

[61] F. Nédélec and D. Foethke, "Collective Langevin dynamics of flexible cytoskeletal fibers", New J Phys 9 (2007), 427.

[62] J. M. Belmonte, M. Leptin, and F. Nédélec, "A theory that predicts behaviors of disordered cytoskeletal networks", Mol Syst Biol 13 (2017), 1.

[63] J. Roostalu et al., "Determinants of polar versus nematic organization in networks of dynamic microtubules and mitotic motors", Cell 175 (2018), 796.

[64] W. Yan et al., "Toward the cellular-scale simulation of motor-driven cytoskeletal assemblies", eLife 11 (2022), e74160.

[65] F. Nédélec, T. Surrey, and A. Maggs, "Dynamic concentration of motors in microtubule arrays", Phys Rev Lett 86 (2001), 3192.

[66] H. Y. Lee and M. Kardar, "Macroscopic equations for pattern formation in mixtures of microtubules and molecular motors", Phys Rev E 64 (2001), 056113.

[67] S. Sankararaman and G. I. Menon, "Self-organized pattern formation in motor-microtubule mixtures", Phys Rev E 70 (2004), 031905.

[68] K. Husain and M. Rao, "Emergent structures in an active polar fluid: Dynamics of shape, scattering, and merger", Phys Rev Lett 118 (2017), 078104.

[69] M. Schuppler et al., "Boundaries steer the contraction of active gels", Nat Commun 7 (2016), 1.

[70] M. C. Marchetti et al., "Hydrodynamics of soft active matter", Rev Mod Phys 85 (2013), 1143.

[71] Z. Qu et al., "Persistent fluid flows defined by active matter boundaries", Comms Phys 4 (2021), 1.

[72] P. J. Foster et al., "Active contraction of microtubule networks", eLife 4 (2015), 1.

[73] G. Guntas et al., "Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins", Proc Natl Acad Sci 112 (2015), 112.

[74] T. D. Ross et al., "Controlling organization and forces in active matter through optically defined boundaries", Nature 572 (2019), 224.

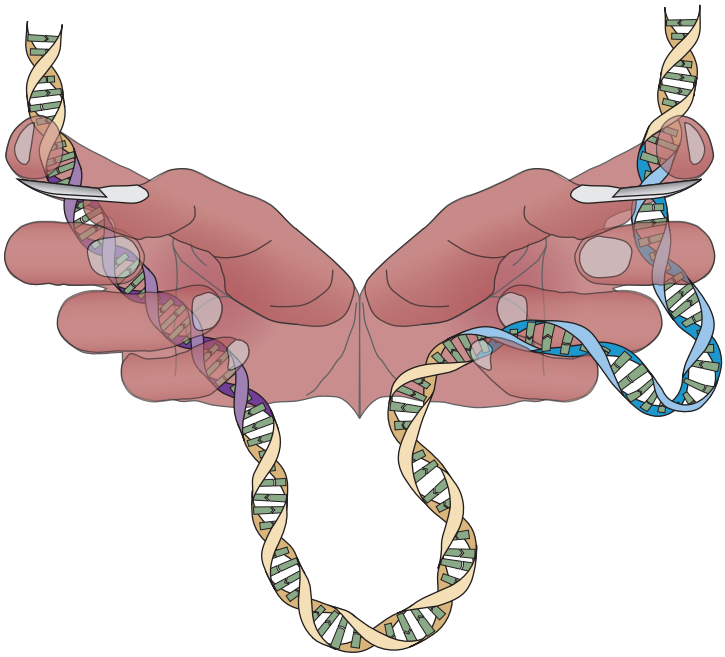
[75] X. Yang et al., "Cellular bioenergetics: Energy fluxes, budgets, and constraints in cells", Proc Natl Acad Sci 118 (2021), e2026786118.

[76] A. M. Tayar, M. F. Hagan, and Z. Dogic, "Active liquid crystals powered by force-sensing DNA-motor clusters", Proc Natl Acad Sci 118 (2021), e210873118.

[77] R. A. Banks, V. Galstyan, H. J. Lee, S. Hirokawa, A. Ierokomos, T. Ross, Z. Bryant, M. Thomson, and R. Phillips, "Motor processivity and speed determine structure and dynamics of motor-microtubule assemblies", bioRxiv (under review at eLife) (2021).

Chapter 2

SEQUENCE-DEPENDENT DYNAMICS OF SYNTHETIC AND ENDOGENOUS RSSS IN V(D)J RECOMBINATION



[1] S. Hirokawa, G. Chure, N. M. Belliveau, G. A. Lovely, M. Anaya, D. G. Schatz, D. Baltimore, and R. Phillips, "Sequence-dependent dynamics of synthetic and endogenous RSSs in V(D)J recombination", Nucleic Acids Research 48 (2020), 6726.

2.1 Abstract

Developing lymphocytes of jawed vertebrates cleave and combine distinct gene segments to assemble antigen-receptor genes. This process called V(D)J recombination involves the RAG recombinase binding and cutting recombination signal sequences (RSSs) composed of conserved heptamer and nonamer sequences flanking less well-conserved 12- or 23-bp spacers. Little quantitative information is known about the contributions of individual RSS positions over the course of the RAG-RSS interaction. We employ a single-molecule method known as tethered particle motion to track the formation, lifetime,

and cleavage of individual RAG-12RSS-23RSS paired complexes (PCs) for numerous synthetic and endogenous 12RSSs. We reveal that single-bp changes, including in the 12RSS spacer, can significantly and selectively alter PC formation or the probability of RAG-mediated cleavage in the PC. We find that some rarely-used endogenous gene segments can be mapped directly to poor RAG binding on their adjacent 12RSSs. Finally, we find that while abrogating RSS nicking with Ca^{2+} leads to substantially shorter PC lifetimes, analysis of the complete lifetime distributions of any 12RSS even on this reduced system reveals that the process of exiting the PC involves unidentified molecular details whose involvement in RAG-RSS dynamics are crucial to quantitatively capture kinetics in V(D)J recombination.

2.2 Introduction

Jawed vertebrates call upon developing lymphocytes to undergo a genomic cut-and-paste process known as V(D)J recombination, where disparate gene segments that do not individually code for an antigen-receptor protein are systematically combined to assemble a complete, antigen receptor-encoding gene [1]. V(D)J recombination supports the production of a vast repertoire of antibodies and T-cell receptors that protect the host organism from a broad array of pathogens. However, gene segment combinations are not made in equal proportions; some gene segment combinations are produced more frequently than others [2–5]. Although V(D)J recombination requires careful orchestration of many enzymatic and regulatory processes to ensure functional antigen receptor genes whose products do not harm the host, we strip away these factors and focus on the initial stages of V(D)J recombination. Specifically, we investigate how the dynamics between the enzyme that carries out the cutting process and its corresponding DNA binding sites adjacent to the gene segments influence the initial stages of recombination for an array of synthetic and endogenous binding site sequences.

The process of V(D)J recombination (schematized in Fig. 2.1) is initiated with the interaction between the recombination-activating gene (RAG) protein complex and two short sequences of DNA neighboring the gene segments, one that is 28 bp and another that is 39 bp in length. These recombination signal sequences (RSSs) are composed of a well-conserved heptamer region immediately adjacent to the gene segment, a more variable 12- (for the 12RSS) or 23-bp (for the 23RSS) spacer sequence and a well-conserved nonamer region.

For gene rearrangement to begin, RAG must bind to both the 12- and the 23RSS to form the paired complex (PC) state (Fig. 2.1B). Throughout the binding interaction between RAG and either RSS, RAG has an opportunity to nick the DNA (enlargement in Fig. 2.1B) [6]. RAG must nick both RSSs before it cleaves the DNA adjacent to the heptamers to expose the gene segments and to create DNA hairpin ends (Fig. 2.1C). DNA repair proteins complete the reaction by joining the gene segments to each other and the RSSs to one another (Fig. 2.1D).

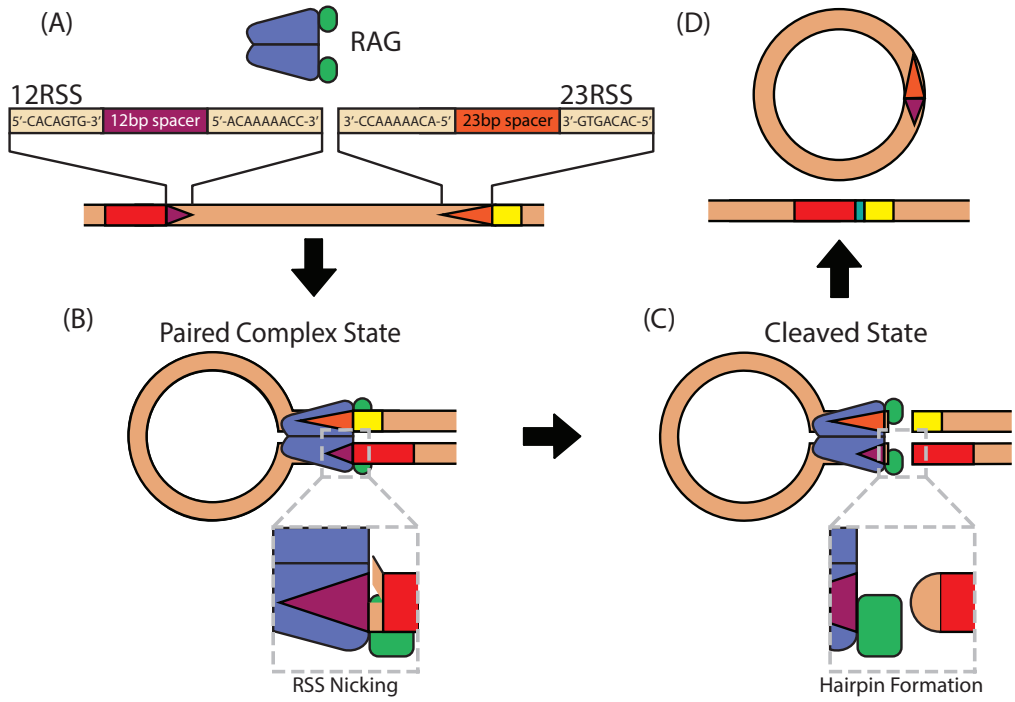


Figure 2.1: **Schematic focusing on the initial steps of V(D)J recombination.** (A) The RAG complex composed of RAG1 (purple) and RAG2 (green) binds to the 12- and 23RSSs (dark purple and orange triangles, respectively) neighboring gene segments (shown as red and yellow boxes on the DNA), (B) forming the paired complex (PC). At any point when it is bound to an RSS, RAG can introduce a nick in the DNA between the heptamer and gene segment (shown with the magnified 12RSS) and must do so to both sites before (C) it cleaves the DNA to expose the gene segments. As indicated by the magnified gene segment end, the exposed DNA strands of the gene segment are connected to form a DNA hairpin. (D) Additional proteins join these segments together. In this work, the stages subsequent to DNA cleavage are not monitored.

RSS sequence-conservation studies across many organisms have shown a vast diversity of 12- and 23RSS sequences, mainly found through heterogeneity in the spacer region [7]. Bulk assays reveal that changing an RSS sequence can significantly influence the RAG-RSS interaction and ultimately the success rate of completing recombination [8–12]. Recent structural results provide evidence that RAG binding is sensitive to both base-specific contacts and the local flexibility or rigidity of the 12- and 23RSS [13–15]. Despite this extensive characterization on the interaction, little is known about how a given RSS sequence affects each step of the RAG-RSS reaction. In this work, we provide one of the most comprehensive studies of how RSS sequences govern the initial steps of V(D)J recombination and provide a quantitative measure of their effects on the formation frequency, lifetime, and cleavage probability of the PC.

We employ a single-molecule technique known as tethered particle motion (TPM) in which an engineered strand of DNA containing a 12RSS and 23RSS is attached to a glass coverslip at one end and to a polystyrene bead at the other (Fig. 2.2A). Using brightfield microscopy, we collect the root mean squared displacement (RMSD) of the bead over time to identify the state of the RAG-RSS interaction. As illustrated in Fig. 2.2B, when RAG forms the PC with the RSSs, the shortened DNA tether constrains the motion of the bead, reducing the RMSD. When RAG cleaves the PC, the bead is released and diffuses away from the tether site (Fig. 2.2C). TPM has been applied to track the dynamic behavior of various protein-DNA systems, including RAG and RSS [16–21]. It is with the temporal resolution provided by TPM that we can track the full progression of individual RAG-RSS interactions from PC formation to cleavage.

We were interested in using TPM to determine the extent to which endogenous RSSs dictate the usage frequency of their neighboring gene segments and, for those RSS positions that do seem to influence gene segment usage, identify the steps in the RAG-RSS reaction when the RSSs help or hurt the selection of their gene segment by extracting kinetic rates. We first examine single bp changes to a designated reference 12RSS, thereby establishing a mechanistic understanding of the contribution of individual nucleotide positions to RAG-RSS dynamics. With the synthetic 12RSSs providing context, we study a set of endogenous 12RSSs, each of whose sequences can be directly related

to the reference sequence and a subset of the characterized synthetic 12RSSs. This selection of 12RSSs was also chosen from repertoires where the usage frequencies of their gene segments are known. Finally, due to the depth of insight offered by waiting time distributions generated by the TPM assay, in an attempt to provide some of the first measurements of various RAG-RSS kinetics, we show through our analysis of the PC lifetime distributions that regardless of choice of 12RSS or divalent cation, our TPM data consistently disagree with a single-rate model. We discuss the consequences of our finding in the context of our understanding of the molecular details of the RAG-RSS reaction. As this study resulted in a wealth of data on a large number of RSS sequences, we have developed an interactive online resource for visualizing the dataset in its entirety.

2.3 Results

2.3.1 Synthetic RSSs

We chose a 12RSS flanking the immunoglobulin κ variable ($Ig\kappa$V) gene segment, V4-57-1, as the reference sequence due to its use in a previous TPM study on RAG-RSS interactions [20]. This sequence has also been used in structural studies of RAG-RSS complexes [13, 15], allowing us to compare our results with known information on the RAG-RSS structure. To explore how RAG-RSS interactions are affected by single bp changes, we examined 40 synthetic RSSs consisting of single bp changes across 21 positions of the V4-57-1 12RSS, with a particular focus on altering the 12 bp spacer which is the least well-understood element in the RSS. We also studied changes made to positions 3-7 of the heptamer and various positions of the nonamer. The first three positions of the heptamer are perfectly conserved [7], likely to support DNA distortions needed for both nicking and base-specific interactions with the cleavage domain on RAG1 after nicking [13–15], while heptamer positions 4-7 also mediate base-specific interactions with RAG [13]. The nonamer is bound by a nonamer-specific binding domain on RAG1 [13, 22]. Throughout our synthetic and endogenous RSS study, we used the same concentration of the two RAG components (RAG1 and RAG2) that were co-expressed and co-purified; and the same concentration of the high mobility group box 1 (HMGB1) protein, which binds nonspecifically to DNA and helps facilitate RAG binding to the RSSs [12]. We also fixed the distance between the two binding sites to be 1200 bp, thereby constraining our study to the influence

of binding site sequence on RAG-RSS dynamics alone. In addition, all of the 12RSSs in this study are partnered with a well-characterized 23RSS [13, 15, 20] adjacent to the frequently-used Jκ1 gene segment from the mouse *Ig*κ locus on chromosome 6 [5]. The sequence of this RSS is provided in Table A.1 of Appendix A. All primer sequences and bead, loop, and cut counts for each synthetic 12RSS are provided in Table A.2 of Appendix A.

We pooled the relevant data across experimental replicates to characterize synthetic RSSs by three empirical properties, namely the frequency of entering the PC (looping frequency), the quartiles of the PC lifetime (dwell time) distribution, and the probability of exiting the PC through DNA cleavage

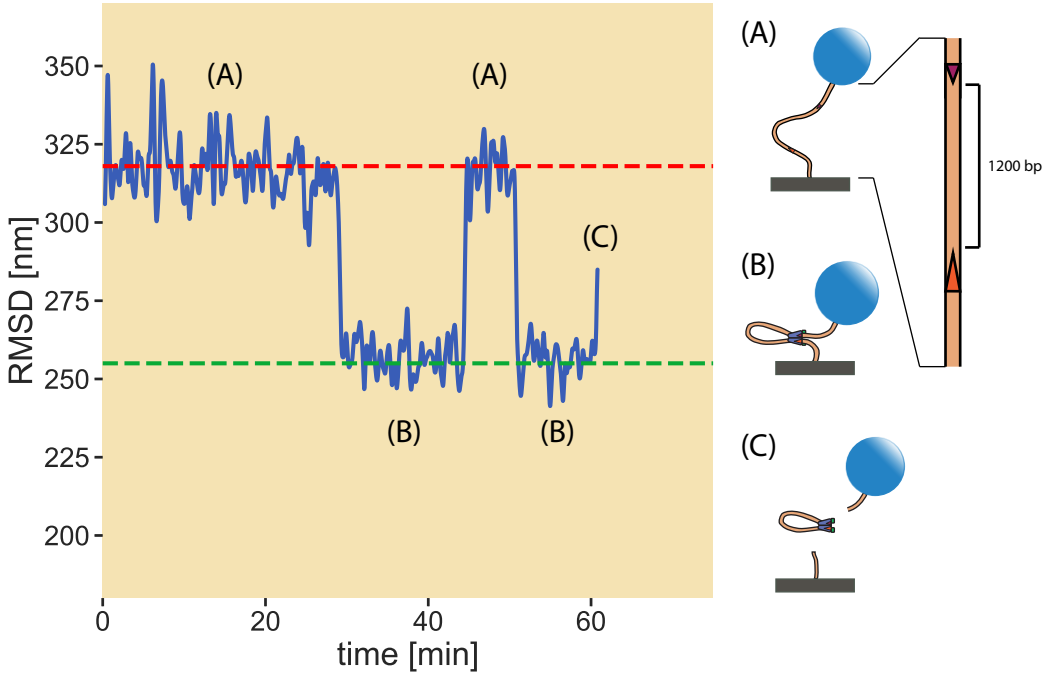


Figure 2.2: **Sample data output of TPM.** By tracking the root mean square displacement (RMSD) of the tethered bead position undergoing restrained Brownian motion, we discern when the DNA tether is (A) in the unlooped state, (B) in the PC (looped) state, and (C) cleaved. The dashed horizontal lines distinguish the unlooped (red) and looped (green) states of the DNA, and are drawn before examining the bead trajectories. The RMSD values of these lines are based on the length of the DNA tether; the distance between the RSSs along the strand; the extent to which HMGB1, a protein that binds nonspecifically to DNA and helps facilitate RAG binding, kinks the DNA; and a set of calibration experiments relating the range of motion of the bead to the length of its tether. As depicted with the magnified DNA strand in (A), the 12RSS and 23RSS are positioned 1200 bp away throughout the study.

(cutting probability). We define the looping frequency as the ratio of distinct PCs observed to the total number of beads monitored over the course of the experiment. Because a single DNA tether can loop and unloop multiple times over the course of the experiment, the looping frequency can in principle range from 0 to ∞. The measured looping frequency and the 95% confidence intervals from bootstrapping the looping frequency as demonstrated in Fig. A.3) are shown for all of the synthetic RSSs in Fig. 2.3.

As demonstrated in Fig. 2.4A, the dwell times were obtained from measuring the lifetimes of each PC state, irrespective of whether the PC was cleaved or reverted to an unlooped state. For each synthetic RSS, all of the PC lifetimes are pooled to generate a histogram of dwell time distributions such as that in Fig. 2.4B, from which the mean, shown as a white circle with an N for nucleotide, and the first and third quartiles, shown as the furthest extents of the blue error bar, are used to compare the synthetic RSSs in Fig. 2.4C.

Finally, to compute the cutting probability, we considered the fate of each PC

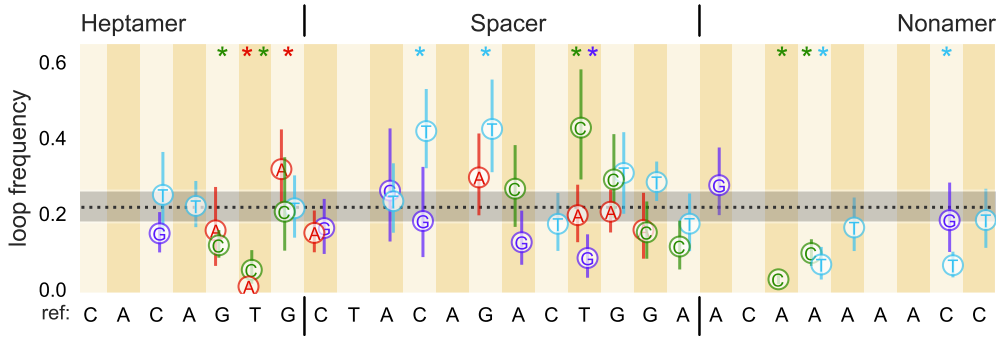


Figure 2.3: **Looping frequency for single bp changes introduced at various positions of the reference 12RSS.** Loop frequency with 95% confidence interval of the distribution of possible looping frequencies from 10^6 bootstrap replicates. The dotted black line is set at the reference loop frequency, 0.22, with shaded area denoting the extent of the 95% confidence interval for the reference. Alternating vertical stripe colors and the reference sequence written along the x-axis demarcate the position where the change was made and the original nucleotide. The introduced nucleotide is provided in the figure with the letter and color-coded (red for A, green for C, light blue for T and purple for G). Heptamer, spacer, and nonamer regions are also separated by vertical lines in the sequences. Asterisks at the top of certain positions are color-coded to specify the nucleotide whose resultant looping frequency differs from the reference sequence with p-value ≤ 0.05. All p-values for each 12RSS used are reported in Fig. A.4.

as a Bernoulli trial with cleavage probability p_{cut}. This treatment allows us to construct the full probability distribution of p_{cut} defined explicitly in Fig. 2.5A and fully detailed in Section 2.5 and in Section A.1 of Appendix A for a PC containing the RSS of interest. The measured number of loops n_{loops} and cuts n_{cuts} collected from experiments (in the case of Fig. 2.5A, 152 loops and 70 cuts) are parameters inserted into the equation to yield a distribution such as

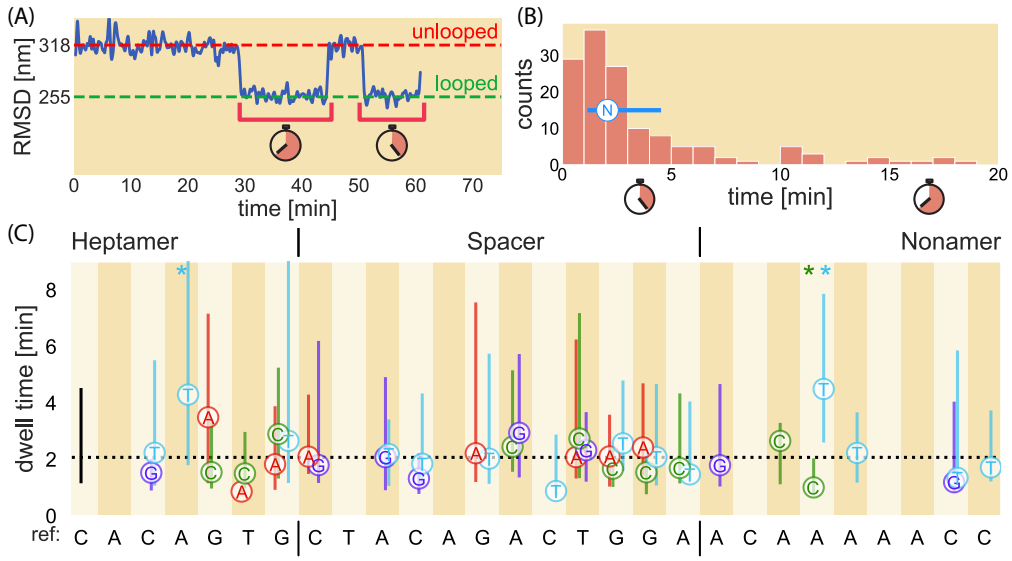


Figure 2.4: **Dwell time quartiles for single bp changes introduced at various positions of the reference 12RSS.** (A) Example bead trajectory data (blue) and the dwell times of the two loops that are formed (brackets). As in Fig. 2.2, the red dashed line corresponds to the unlooped DNA tether state while the green dashed line denotes the predicted looped state. (B) Histogram of all dwell times collected for a given RSS. Note that all loops involving the RSS of interest are included in the histogram, regardless of whether the loop precedes cutting or a return to the unbound state. The median is shown as the circle containing N (for nucleotide) with lines extending to the first and third quartiles. The method for obtaining the circle and error bars as shown in (B) are then applied to each synthetic 12RSS dataset and presented in (C) where the letters denote the replacement nucleotide. The dotted black line in (C) denotes the reference 12RSS median dwell time, 2.1 minutes, with the black bar at the left denoting the first and third quartiles of the distribution. Vertical stripes; x-axis labeling; heptamer, spacer, and nonamer distinction; and color-coding of nucleotide changes (red for A, green for C, light blue for T and purple for G) are the same as in Fig. 2.3. Asterisks at the top of certain positions are color-coded to specify the nucleotide whose resultant dwell time differs from the reference sequence with p-value ≤ 0.05. All p-values for each 12RSS used are reported in Fig. A.4.

in Fig. 2.5B. We computed the most likely p_{cut} and one standard deviation, as demonstrated in Fig. 2.5B by the white circle with the N and blue error bars, respectively, for each synthetic RSS, and compiled them in Fig. 2.5C. The Cutting Probability Model Explorer interactive figure provides a visualization for how the probability distribution is sensitive to the empirically-collected number of loops and cuts. Detailed discussions of the choice of metrics and the corresponding error estimates are provided in Section 2.5 and in Sections A.2.2-A.2.4. We also show in Fig. A.5 and Section A.2.5 of Appendix A that our definitions of the looping frequency and cutting probability decouple the PC forming and cleavage steps in the RAG-RSS reaction, thereby clarifying which step is the limiting factor in completing the cleavage phase of V(D)J recombination. We complement the condensed synthetic RSS results presented here with an interactive figure that provides a more complete visualization of each RSS studied on the website. The Synthetic RSS Explorer interactive figure includes heatmaps to qualitatively illustrate how the synthetic RSSs differ in the three defined metrics. By clicking on a particular cell in any of the heatmaps, the interactive displays the measured looping frequency of the synthetic RSS containing the corresponding bp change with several confidence interval percentages from the bootstrapping. Hovering over a cell also brings up a window showing the number of beads, loops, and cuts observed for the mutant. In addition, the webpage shows empirical cumulative distribution functions (ECDFs) of PC lifetimes in three groups: PCs that are cleaved, PCs that are unlooped, and both together. This webpage includes the complete posterior probability distribution of p_{cut} for each synthetic RSS.

Figs. 2.3, 2.4C, and 2.5C illustrate the substantial effect that a single bp change to an RSS can have on the formation, stability, and cleavage of the PC, respectively, reaffirming that RSS sequence plays a role in regulating the initial steps of recombination. Of interest is the observed difference in phenomena between changes made to the third position and those made to the last four bases of the heptamer region. Bulk assays have shown that deviating from the consensus C at heptamer position 3 essentially eliminates recombination [8, 10], yet we found that changing from the C to G or T did not inhibit PC formation (Fig. 2.4C). In fact, these alterations show similar looping frequencies and PC lifetimes (Fig. 2.4C) as found for the reference sequence. Instead, both the C-to-G and C-to-T alterations to heptamer position 3 almost completely suppress cleavage (Fig. 2.5C). We provide the full probability

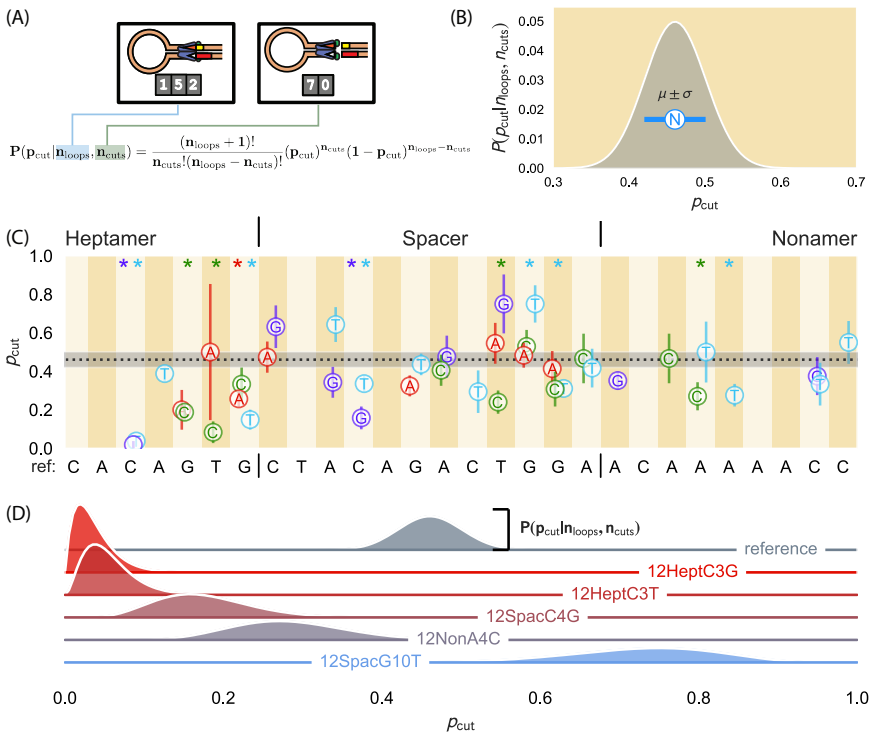


Figure 2.5: **Cutting probabilities for single bp changes introduced at various positions of the reference 12RSS.** (A) For a given RSS, the total number of distinct loops in the assay n_{loops} (in this case, 152 loops) and the subset of those loops that RAG cleaves n_{cuts} (70) are applied to the function shown that identifies the full distribution of the cutting probability p_{cut} of the PC for a 12RSS of interest. (B) Example distribution for a particular RSS, with the most likely cutting probability μ with N for 'nucleotide' and standard deviation σ shown as a circle with blue error bars, respectively. (C) μ and σ are shown for each synthetic RSS with the dotted black line denoting the most probable p_{cut} for the reference sequence, roughly 0.46, with the grey shaded region setting one standard deviation. Vertical stripes; x-axis labeling; heptamer, spacer, and nonamer distinction; and color-coding of nucleotide changes (red for A, green for C, light blue for T and purple for G) are the same as in Fig. 2.3. (D) Ridgeline plot of posterior distributions of the cutting probability, given the number of loops observed and loops that cut (see SI) for a subset of the synthetic RSSs (labeled and colored along the zero-line of the respective ridgeline plot). Height of the distribution to the horizontal line of the same color corresponds to the posterior distribution. See the Cutting Probability Model Explorer interactive webpage to see how the posterior distribution depends on the number of loops and cuts observed. Asterisks at the top of certain positions are color-coded to specify the nucleotide whose resultant cutting probability differs from the reference sequence with p-value ≤ 0.05. All p-values for each 12RSS used are reported in Fig. A.4.

distribution for the estimate of p_{cut} for these two RSSs in Fig. 2.5D. Nearly all of the probability density is concentrated below 10%, showing that cutting the PC is exceedingly rare. Thus, although deviating from a C at heptamer position 3 does not prevent RAG from forming the PC, the alteration impedes DNA cleavage.

Among the changes made to the last four bases of the heptamer from the reference sequence, the fifth and sixth positions showed the most striking reductions on PC formation (Fig. 2.3). Of more than 240 DNA tethers with the 12RSS containing a T-to-A change at heptamer position 6, only two PCs formed, one of which subsequently led to cleavage. This result is consistent with recent findings that the consensus TG dinucleotide at the last two positions of the heptamer supports a kink in the DNA and may be critical for RAG binding [14]. We notice that some changes such as the one at heptamer position 4 (A to T) increase the median time spent in the PC (Fig. 2.4C). This RSS also had one of the widest dwell time distributions of all of the synthetic RSSs studied. While some alterations to the last four heptamer positions yielded little change in cleavage propensity compared to the reference, others showed a reduction in p_{cut}. The single bp change that had the greatest effect, located at heptamer position 6 (T to C) showed that only 2 out of 24 PCs led to cleavage.

Although we observed only modest differences in the median dwell times when we altered the reference sequence in the spacer region, some alterations substantially affected the looping frequency and cutting probability. The C-to-T change at spacer position 4 doubled the frequency of observing the PC while a T-to-G change at the ninth position reduced PC formation nearly as much as changes made at heptamer position 6 (Fig. 2.3). These two changes made in the spacer reflect the observed extremes of spacer sequence effects on the looping frequency. While many of the changes in the spacer region do not alter the cutting probability, we can still find spacer-altered RSSs that improve or inhibit cleavage. Fig. 2.5D shows that changing the fourth position from C to G shifts the probability distribution of p_{cut} to lower values, while altering the tenth position of the spacer from G to T shifts the distribution toward an increased cleavage probability. RAG1 makes contacts along the entire length of the 12RSS spacer [14], helping to explain our finding that changes to the spacer can substantially alter the probability of PC formation and cutting, thereby playing more of a role than simply separating the heptamer and non-

amer sequences.

Similar to spacer changes, most nonamer changes show strongly overlapping dwell time distributions, with median PC dwell times differing from the reference sequence by less than 1 minute (Fig. 2.4C). However, unlike spacer-modified RSSs, most nonamer-altered RSSs reduced the frequency of PC formation. Disruptions to the poly-A sequence in the center of the nonamer cause a substantial reduction in looping frequency, most notably the near complete inhibition of PC formation with the A-to-C change at nonamer position 3 (Fig. 2.3). This detrimental effect of deviating from the poly-A tract agrees with previous work demonstrating numerous protein-DNA interactions in this region and with the proposal that the rigidity produced from the string of A nucleotides is a critical feature for RAG1 to bind the nonamer [14, 22]. Furthermore, this reduction in looping frequency can extend to changes made toward the end of the nonamer, depending upon the nucleotide, as shown with the significant reduction for the C-to-T mutation at nonamer position 8 (Fig. 2.3). The sequence deviations in the nonamer region, however, do not significantly affect cleavage once the PC has formed, as evidenced by the overlap in the posterior distributions of the reference sequence and its nonamer variant showing the greatest reduction in cleavage probability (position 4, A to C), in Fig. 2.5D. Overall, nonamer deviations from the reference RSS have negative effects on PC formation with minimal effects on subsequent DNA cleavage, consistent with extensive biochemical and structural evidence that the primary function of the nonamer is to facilitate RAG-DNA binding [22].

2.3.2 Endogenous RSSs

To build on our study of single bp effects on RAG-RSS dynamics, we selected a subset of endogenous RSSs from the mouse Vκ locus on chromosome 6 based on existing gene usage frequency data collected by Aoki-Ota *et al.* [5] and because the sequence differences between these RSSs and the reference RSS are individually examined in the synthetic RSS results. We studied a variety of frequently-used ($> 5\%$ frequency of usage) gene segments (V1-135, V9-120, V10-96, V19-93, V6-15, and V6-17), two moderately-used ($> 1\%$ and $< 3\%$ frequency) gene segments (V4-55 and V5-43) and two rarely-used ($< 0.5\%$ frequency) gene segments (V4-57-1 and V8-18) [5]. We note that the V4-57-1 12RSS is identical to the reference 12RSS in the synthetic study. Furthermore, we use the same Jκ1 23RSS in the endogenous RSS study as in

the synthetic study. In addition, we examined DFL16.1, the most frequently used D gene segment from the murine immunoglobulin heavy chain (*Igh*) locus on chromosome 12 [4, 23]. Unlike the Vκ gene segments, which only need to combine with one gene segment, D gene segments must combine with two other gene segments to encode a complete protein. As a result, DFL16.1 is flanked on both its 5' and 3' sides by distinct 12RSS sequences, denoted DFL16.1-5' and DFL16.1-3', respectively, both of which are examined in this study. The sequences of all endogenous RSSs studied here as well as the number of beads, loops, and cuts observed are provided in Tables S1 and S3. We apply TPM on these sequences to determine whether their involvement in the RAG-RSS reaction could both provide insight into the usage frequency of their flanking gene segments and be predicted based on the activity profile of the synthetic RSSs.

To develop a better sense for how RAG interacts with these RSSs in their endogenous context, the 6 bp coding flank sequence adjacent to the heptamer of all but the V4-57-1 RSS was chosen to be the natural flank provided by the endogenous gene segment. RAG interacts with the coding flank during DNA binding and PC formation [13–15] and coding flank sequence can influence recombination efficiency, particularly the two bp immediately adjacent to the heptamer [24–26]. Two T nucleotides and in many cases even a single T immediately 5' of the heptamer inhibit the nicking step of cleavage and thus reduce recombination efficiency [24–26]. We did not extensively analyze the contribution of coding flank sequence in this study, and only V6-15 RSS among the studied RSSs would be predicted to interact poorly with RAG due to the T flanking the heptamer; all other coding flanks have combinations of A and C as the two terminal coding flank bases. We kept the same coding flank for the V4-57-1 RSS as in a previous study [20] to facilitate closer comparison of the results of the synthetic RSSs. We do not expect much difference between the endogenous coding flank sequence (5'-CACTCA, where the two nucleotides closest to the heptamer are underlined) and the coding flank used here (5'-GTCGAC) because the two terminal coding flank bases are similar to those of all but the V6-15 RSS and for reasons discussed in the Discussion and in Section A.5 of Appendix A. The coding flank sequences for all studied endogenous RSSs are included in Table A.1. We present the results of the RAG-endogenous RSS interaction in Fig. 2.6 and provide an interactive tool for exploring these data on the paper website. The Endogenous RSS Explorer

includes an interactive feature where the looping frequency, ECDFs of looping lifetimes, and posterior distributions of the cleavage probability of any two endogenous RSSs can be directly compared.

The variable nature of all three metrics [looping frequency (Fig. 2.6A), dwell time (B), and cutting probability (C; full posterior distributions for all endogenous 12RSSs studied here are shown in Fig. A.6)] across RSSs highlights how, similar to the synthetic RSSs, endogenous sequences influence formation, stability, and cleavage of the PC differently. Of particular interest is the behavior of DFL16.1-3' which shows the highest propensity for PC formation but some of the shortest PC lifetimes. Despite this short median dwell time, the probability of the PC successfully proceeding to DNA cleavage is high, approximately 0.5. Notably, the frequency of PC formation and the probability of cleavage are both greatly reduced for DFL16.1-5' as compared to DFL16.1-3', although their median PC dwell time and the width of the dwell time distributions are approximately equal. Reduced function of DFL16.1-5' relative to DFL16.1-3' is consistent with prior studies [23, 27, 28] and is addressed further in the Discussion.

The endogenous RSSs of the Vκ gene segments show varying efficiencies of PC formation and cleavage. Many of the endogenous RSSs studied here, including those of gene segments used frequently *in vivo* (V1-135, V9-120, V10-96, V19-93, V6-17, and V6-15), demonstrate looping frequencies between 15 and 30 events per 100 beads. Gene segments V4-57-1 and V4-55 are used with almost 0% and roughly 2.5% frequency, respectively [5], yet in our experiments, they enter the PC with comparable frequency (approximately 20 to 30 loops per 100 beads). In general, we find these two sequences to behave almost identically in our experimental system, illustrating that other biological phenomena, such as higher-order DNA structure, govern the segment usage *in vivo* [4, 29]. The endogenous V8-18 12RSS exhibits infrequent PC formation and cleavage and short median PC lifetimes, much like the DFL16.1-5' 12RSS. Using the V8-18 12RSS, only 5 looping events were detected from 146 DNA tethers and cleavage was never observed. Despite the similarities in reaction parameters for the V8-18 and DFL16.1-5' RSSs, DFL16.1 is the most frequently used D gene segment in the repertoire [4] while V8-18 is never used [5]. A likely explanation for the exclusion of V8-18 in the repertoire is the 'A' at heptamer position 6 of the 12RSS (see Discussion). In contrast, the DFL16.1 is substantially

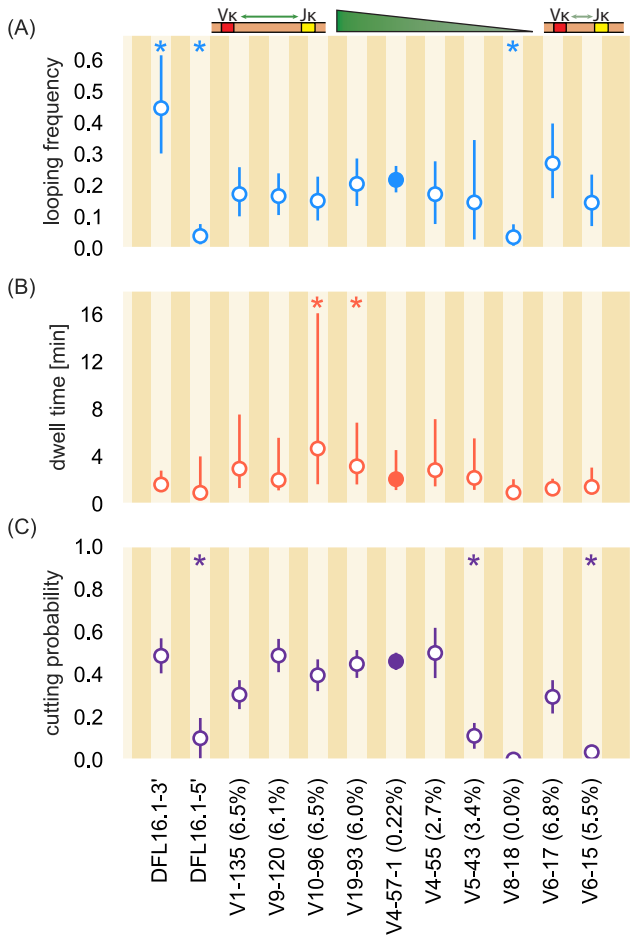


Figure 2.6: **Observed dynamics between RAG and endogenous RSS sequences.** (A) Frequency of PC formation (looping frequency) with 95% confidence interval. (B) Median PC lifetime with the lower error bar extending to the first quartile and the upper error bar extending to the third quartile. (C) Probability of DNA cleavage (cutting probability) of RAG with error bars showing one standard deviation. For discussion of the errors in Fig. 2.6A and 2.6C, see Section A.2.5 of Appendix A. DFL16.1-3' and DFL16.1-5' flank the same gene segment but in different orientations on the Igh chromosome. As shown in the graphic above Fig. 2.6A, Vκ gene segments listed are ordered by their position along the chromosome, with linear distance from the Jκ gene segments decreasing from left to right. Numbers in parentheses next to Vκ gene segment denote percentage of usage in repertoire [5]. The V4-57-1 12RSS has a filled in circle to denote that it is the reference sequence in the synthetic RSS study. Asterisks at the top of subfigures denote endogenous RSSs whose measured quantity differs from the V4-57-1 (reference) 12RSS with p-value ≤ 0.05. All p-values for each 12RSS used are reported in Fig. A.4.

utilized in the *Igh* repertoire despite the poor contribution in PC formation and cleavage of its 5' 12RSS most likely because this RSS does not participate in recombination until after its gene segment has undergone D-to-J recombination with its more efficient 3' 12RSS, thus moving the gene segment into the RAG-rich environment of the "recombination center." This relocation is thought to facilitate RAG binding to the 5' RSS of the committed D gene segment [30, 31].

Fig. 2.6B demonstrates that, with the exception of the V10-96 RSS, PC lifetimes are similarly distributed across the endogenous RSSs examined in this work. Most RSSs have median dwell times between 1 to 3 minutes with the V8-18 12RSS displaying the shortest-lived median dwell time of roughly 40-50 seconds. While most endogenous RSSs here have a similar range between the first and third quartiles (see interactive figure on the paper website), the V10-96 12RSS distribution is noticeably wider, with the first quartile of the distribution being a longer lifetime than the median lifetime for most endogenous RSS distributions and the third quartile of this RSS extending out to over 16 minutes. These observations suggest a similar stability of the PC for all but the V10-96 RSS once RAG manages to bind simultaneously to both 12- and 23RSSs.

Fig. 2.6C indicates that six endogenous RSS sequences from V1-135 to V4-55 have comparable cutting probabilities ranging from 0.4 to 0.5. Considering that the less-frequently used V4-57-1 and V4-55 gene segments have 12RSSs that show similar cutting probabilities and looping frequencies to the 12RSSs of more frequently-selected gene segments, other factors appear to prevent their efficient use. The low probability of cutting (0.05; Fig. 2.6C) with the V6-15 12RSS is particularly noteworthy, indicating that RAG tends to easily break the looped state rather than commit to cleavage. However, this low cutting probability might be attributed to the T in the coding flank immediately adjacent to the heptamer. Other features of the system must dictate the high-frequency usage of V6-15 *in vivo* [5].

2.3.3 Kinetic Modeling of the PC Lifetime Distribution

Figs. 2.4C and 2.6B show that the vast majority of median looping lifetimes ranged between 1 to 3 minutes with rare exceptions, suggesting similar dwell time distributions for many of the RSS variants. However, many of these

synthetic and endogenous RSSs have different probabilities of DNA cleavage, suggesting that at the very least the rate of cutting changes. These similarities in the lifetime distributions but differences in outcomes invited a thorough dissection of the data to extract key quantitative insights into the changes in the kinetics between 12RSS constructs. As TPM has been used to extract kinetic parameters for various other protein-DNA systems [17, 18, 32, 33], we used the distributions of the PC lifetimes in an attempt to establish the rates of unlooping and cutting for each RSS and discern a deeper connection between RSS sequence and fate of the PC. We developed a simple model in which a PC state can have two possible fates: either simple unlooping of the DNA tether or cleavage of the DNA by RAG. We characterized each of these outcomes as independent yet competing processes with rates k_{unloop} and k_{cut} for unlooping and DNA cleavage, respectively. If the waiting time distribution t_{unloop} or t_{cut} for each process could be measured independently where only one of the two outcomes was permitted to occur, one would expect the probability densities of these waiting times given the appropriate rate to be single exponential distributions of the form

$$P(t_{\text{unloop}} \,|\, k_{\text{unloop}}) = k_{\text{unloop}} e^{-k_{\text{unloop}}\, t_{\text{unloop}}} \tag{2.1}$$

for the unlooping process and

$$P(t_{\text{cut}} \,|\, k_{\text{cut}}) = k_{\text{cut}} e^{-k_{\text{cut}}\, t_{\text{cut}}} \tag{2.2}$$

for DNA cleavage. However, as these two Poisson processes are competing, we cannot estimate k_{cut} solely from the waiting time distribution of paired complex states that led to DNA cleavage nor k_{unloop} using the states which simply unlooped. As each individual cutting or unlooping event is assumed to be independent of all other cutting and unlooping events, the distribution of the dwell time t before the PC either unloops or undergoes cleavage can be modeled as an exponential distribution parameterized by the sum of the two rates,

$$P(t \,|\, k_{\text{leave}}) = k_{\text{leave}}\, e^{-k_{\text{leave}}\, t}, \tag{2.3}$$

where $k_{\text{leave}} = k_{\text{unloop}} + k_{\text{cut}}$.

Given the collection of waiting time distributions measured for each RSS, we estimated the values of k_{leave} which best describe the data. We find that the observed dwell times are not exponentially distributed for any 12RSS sequence

analyzed, either endogenous or synthetic. Examples of these waiting time distributions along with an exponential distribution parameterized by the 95% credible region for k_{leave} can be seen for twelve of the RSS variants in Fig. 2.7. In general, the observed dwell times are underdispersed relative to a simple exponential distribution with an overabundance of short-lived PCs. We also find that the observed dwell time distributions are heavily tailed with exceptionally long dwell times occurring more frequently than expected for an exponential distribution.

The ubiquity of this disagreement between the simplest kinetic model and the observed data across all of the examined RSSs indicates that leaving the PC state either by reverting to the unlooped state or committing to the cleaved state is not a one-step process, suggesting that at least one of the two fates for the PC state on its own is not single-exponentially distributed as assumed in our null model of the dynamics.

One hypothesis for the disagreement between the model given in Eq. 2.3 and the data is that other processes, such as nicking of the DNA by RAG, create effects in the tethered bead trajectories that are too subtle to be detected in the TPM assays. Nicking creates a more stable RAG-single RSS complex (though this effect on PC stability had not been previously quantified) [13, 34] and can occur at any time after RAG binds to the RSS [6], making it exceedingly difficult to determine whether a given PC has one, both or neither of the RSSs nicked. As a result, we may not be able to model the combined kinetics of unlooping and cleavage without also identifying when RAG nicks the RSSs to which it is bound.

Substitution of Ca^{2+} in place of Mg^{2+} in the reaction buffer allows RAG to bind the RSSs but blocks both nicking and cleavage [35], leaving unlooping as the only possible fate of a PC. To determine if unlooping could be modeled as a simple Poisson process, we measured the PC dwell time distribution for a subset of the RSSs in a reaction buffer containing Ca^{2+}.

While we observe no cleavage of PCs in the Ca^{2+}-based buffer, looping is as frequent, if not more frequent, than in the Mg^{2+}-based buffer (see Fig. A.7). However, even in the absence of nicking, the dwell times of PC events are still not in agreement with an exponential distribution (left panels of Fig. 2.8A-C). For the dwell time distribution to defy a single-exponential form, the process of unlooping itself cannot be a Poisson process with only one kinetic rate.

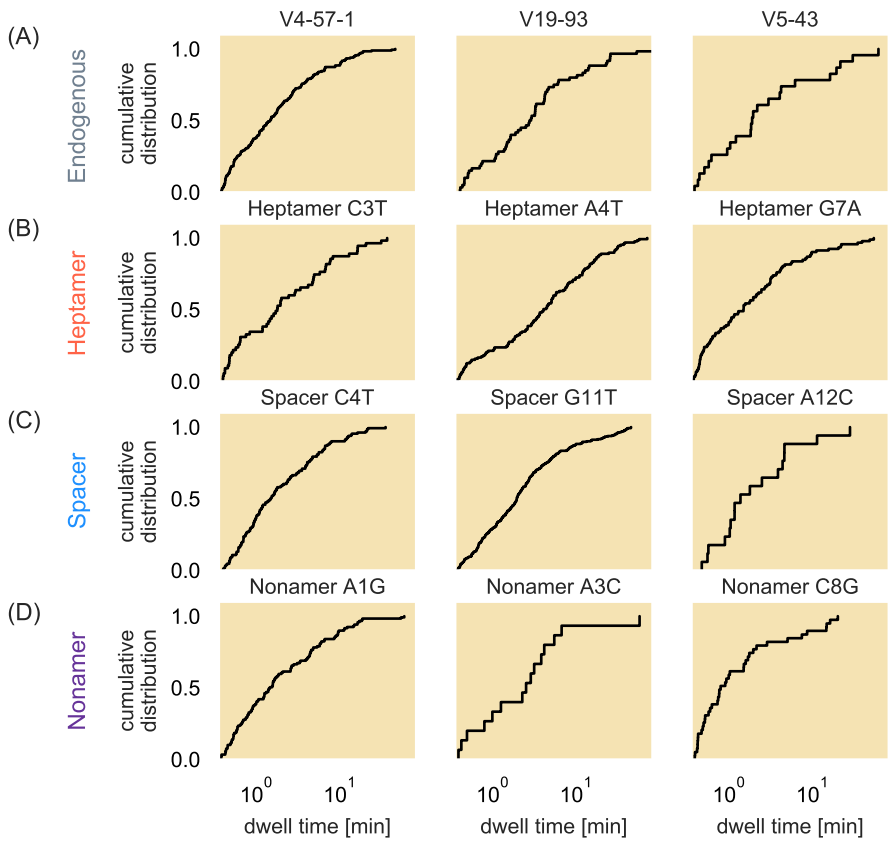


Figure 2.7: **Non-exponential waiting time distributions for endogenous and synthetic 12RSSs.** The empirical cumulative distribution of the measured PC lifetimes (black lines) are shown for representative endogenous sequences (A) as well as for the synthetic RSSs with single point alterations made in the heptamer (B), spacer (C), or nonamer (D) regions. The shaded area corresponds to the 95% credible region of a true exponential distribution parameterized in Eq. 2.3 given a posterior distribution for k_{leave}, the rate of the arrival of either an unlooping or cleavage event. All dwell time axes are plotted on a logarithmic scale.

Extracting kinetic rates of exit from the PC state is not possible without also observing a critical, yet currently indiscernible by TPM, biochemical process between RAG and RSS. We also note that for each of the RSS variants the observed PC lifetimes are short lived compared to those in the Mg^{2+}-based buffer, as can be seen in the bottom plots of Fig. 2.8. Because Ca^{2+} does not significantly alter DNA flexibility compared to Mg^{2+} [36], our data argue that nicking itself results in a longer-lasting PC. This is notable in light of recent structural evidence showing that nicking and the associated "flipping out" of two bases at the RSS-gene segment junction away from their complementary

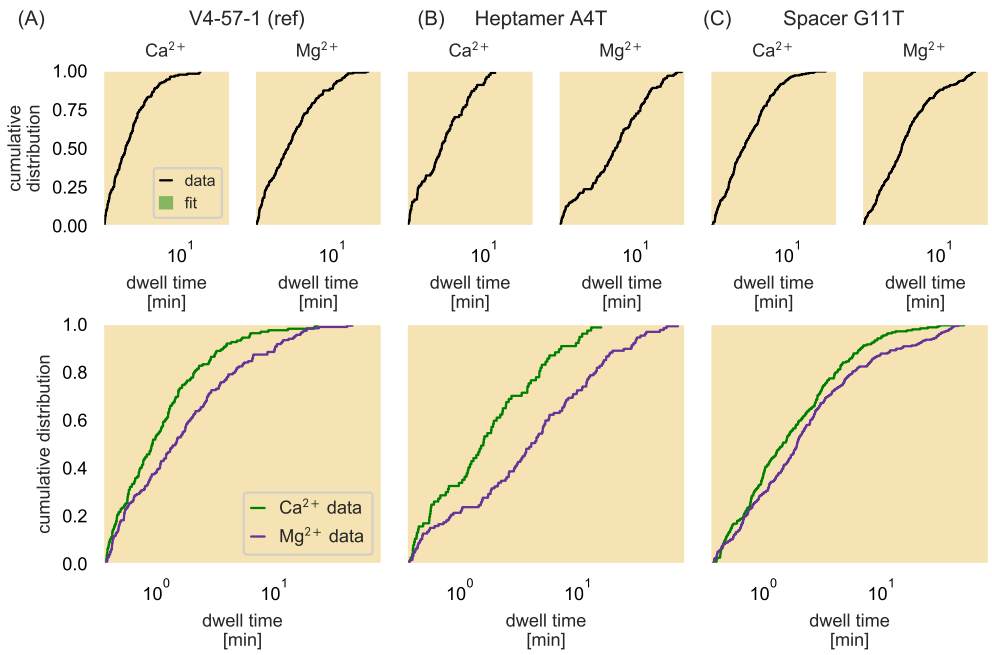


Figure 2.8: **Empirical cumulative distributions of PC lifetimes with different divalent cations.** The empirical cumulative dwell time distributions are plotted in black over the 95% credible region of the fit to an exponential distribution (top row) for the reference sequence (A), a base pair change in the heptamer region of the 12RSS (B), and a base pair change in the spacer region (C). The bottom plots show direct comparisons of the empirical cumulative dwell time distributions collected in either Ca^{2+}- (green) or Mg^{2+}- (purple) supplemented reaction buffer for each RSS. The dwell time axis on all plots are logarithmically scaled.

bases create a fully "closed" RAG-RSS binding conformation that would be predicted to improve stability of the complex [13, 14]. With the more stable conformation from nicking one or both RSSs, the PC state persists for longer than if RAG could not nick either RSS, which is reflected in the longer dwell time distributions when using Mg^{2+}.

2.4 Discussion

Through the temporal resolution provided by TPM, we have discerned how RAG forms and cleaves the PC for a series of synthetic and endogenous RSSs. We find that the RSSs of frequently-used gene segments typically do not support more efficient PC formation or cleavage than those neighboring gene segments of more modest usage. This observation is consistent with recent findings that RSS strength, as assessed by the RSS information content (RIC)

algorithm [11, 37–39], is only one of multiple parameters needed to be able to predict gene segment usage frequency [29, 40]. Furthermore, we found from analyzing single bp variations of the V4-57-1 RSS that the efficiencies of PC formation and cleavage are sensitive to single bp changes depending upon the conservation level at the respective position. We see that altering the perfectly-conserved third position of the heptamer almost completely blocked cleavage by RAG without significantly altering PC formation frequency or dwell time distribution. In contrast, certain deviations from the consensus nucleotide at the last four positions of the heptamer or in the nonamer decreased the frequency of PC formation. Finally, even though few positions of the spacer have a clear consensus nucleotide [7], formation and cleavage of the PC can still be strongly affected by a single bp change in the spacer. In fact, sequence-context effects might explain why some of these synthetic RSSs in less conserved positions of the spacer have such a strong influence on PC formation and cleavage on their own.

We asked to what extent we could account for the behavior of an endogenous RSS based on its constituent nucleotides as revealed by our synthetic RSS study. The Synthetic-Endogenous RSS Comparison interactive tool on the paper website allows one to select an endogenous RSS to reveal not only its data on PC formation, PC lifetime distributions, and cleavage probability distributions, but also data for each nucleotide difference between it and the reference 12RSS through the relevant synthetic RSSs. For ease of comparing the endogenous RSS with a synthetic RSS relevant to the sequence difference, hovering the computer mouse over the nucleotide of interest in the sequence changes the color of all other relevant synthetic RSSs to grey. Although our finding that hidden molecular details in the RAG-12RSS-23RSS interaction prevent us from constructing a quantitative model that directly relates endogenous RSS behavior to the effects measured for each individual sequence deviation, these results provide several insights into the relation between RSS function and its constituent nucleotides. In particular, the data reveal a subset of RSS positions, including some in the spacer, that appear to strongly influence RAG-RSS interactions.

The synthetic RSS with the G-to-T change at spacer position 10 strongly increases the cleavage probability and also enhances PC formation (Fig. 2.3, 2.5C, 2.5D). These improvements might be due to the 5'-TG-3' dinucleotide

created by this change at spacer positions 10 and 11. Such a pyrimidine-purine (YR) pairing is inherently deformable [41] and a substantial 60° bend in the 12RSS is seen at this location in the spacer in RAG-RSS complexes [14]. Hence, as noted previously [14], a YR combination at the 3' end of the spacer in the 12RSS is favorable for DNA binding, consistent with our data. The DFL16.1-5' RSS contains a T at spacer position 10 (Table A.1), as well as several other nucleotides in the spacer that each individually increase PC formation (see the paper website), but this RSS exhibits inefficient PC formation (Fig. 2.6A). Because spacer position 11 is also a T in the DFL16.1-5' RSS, the T at position 10 does not create a YR pair and instead, the last seven bp of the spacer are all pyrimidines. A spacer with such a sequence might be particularly poor at supporting the DNA distortions needed for RAG-12RSS binding. This example of the importance of sequence context in determining how a particular bp will influence RSS function supports a concept borne out of the development of the RIC algorithm [11, 37, 39].

The contributions that coding flanks make to RAG-RSS dynamics [13] are important considerations to quantitatively model the RAG-DNA interactions, as each endogenous RSS neighbors a different coding flank. We attributed the low cleavage probability of the V6-15 RSS to the T immediately adjacent to the RSS in the coding flank, which has been shown to be detrimental to recombination efficiency [24–26]. Because the other endogenous RSSs studied are rich in C and A nucleotides in the two bp adjacent to the heptamer, we compared data for two pairs of DNA constructs that differed only in coding flank sequence. One comparison involves the substrate containing the coding flank sequence used on the V4-57-1 RSS (5'-GTCGAC) and a substrate with a C-to-A change adjacent to the heptamer (5'-GTCGAA). The other pair is the V4-55 endogenous RSS substrate and the synthetic RSS substrate containing a C-to-A alteration at spacer position 1, where, fortuitously, the RSSs are identical and the coding flanks differ by five base pairs (5'-CACCCA for V4-55 and 5'-GTCGAC for the synthetic RSS). In both cases, the looping frequencies, PC lifetime distributions, and cutting probability distributions are similar for the respective pairs, arguing that these coding flank differences contribute little to the overall RAG-RSS reaction (see Fig. A.8 and A.9). Hence, coding flank differences present in all of the endogenous RSS substrates analyzed here, with the exception of the V6-15 RSS, are unlikely to have a strong influence on RAG-RSS dynamics. However, a more extensive examination of coding flank,

particularly for G- and T-rich sequences, in a dynamic experimental method such as TPM will help to shed light on the extent to which these RSS-adjacent sequences influence the various steps of V(D)J recombination.

The V5-43 12RSS has a low level of PC formation, likely because of its C-to-T change at nonamer position 8, while its poor cutting probability can be attributed to a collection of sequence changes that reduce cleavage probability. The low frequency of PC formation with the V9-120 and V6-15 RSSs is likely driven primarily by the A-to-T change at nonamer position 4, with additional negative contributions coming from altering the reference spacer. And the DFL16.1-3' RSS, which supported the highest frequency of PC formation across all RSSs studied, differs from the reference RSS at the fourth and sixth positions of the spacer that each in their own synthetic RSSs strongly stimulated PC formation. These findings support the important conclusion that spacer sequence can influence RSS synapsis by RAG.

We find that the DFL16.1-5' RSS is much less competent for PC formation and cleavage than the DFL16.1-3' RSS. Weaker activity of the 5' RSS compared to the 3' RSS is consistent with the results of recombination assays performed using plasmid substrates in cells [27, 28] and for chromosomal recombination when DFL16.1 was placed approximately 700 bp from its *Igh* J gene segment partner, J_H1 [23]. However, when assayed in their natural location over 50 kb from the J_H gene segments, the two RSSs support roughly equal levels of recombination as long as they are in the same orientation relative to the J_H 23RSSs [23]. The existing data argue that the DFL16.1-5' RSS is intrinsically less active for recombination than the DFL16.1-3' RSS, but this difference can be minimized over large chromosomal distances when chromatin "scanning" by RAG is the dominant mechanism for bringing RSSs together to form the PC [23, 42]. Such scanning requires that a RAG-single RSS complex be able to either bind and then release, or else entirely skip over, proximal partner RSSs to be able to recombine with more distal RSSs. Our findings relating RSS sequence to the efficiencies of PC formation and cleavage within the PC provide a valuable resource for considering how RSS sequence might influence the scanning process.

Our study of both synthetic and endogenous RSSs explains the low usage of the V8-18 gene segment in the *Igκ* repertoire and further highlights the strong impact that can be exerted from a single nucleotide change to an RSS. The

V8-18 RSS contributes to inefficient PC formation and further interrogating each sequence mismatch between the V8-18 and reference RSSs revealed that its T-to-A alteration at heptamer position 6 is sufficient to virtually abrogate PC formation. This effect on gene segment selection may not be unique to the V8-18 gene segment: both the V1-131 and V8-26 gene segments have the same T-to-A deviation from the consensus heptamer position 6, and neither is used in recombination [5]. This deviation from consensus further provides a mechanistic explanation for why the VκA2b gene segment is underutilized in the antibody repertoire of Navajos, which in turn has been proposed to account for the high susceptibility of Navajos and several genetically-related Native American groups to *Haemophilus influenza* type b infection [43]. The VκA2b RSS differs in sequence from the more common and efficiently recombined VκA2a RSS by a single T-to-A change at heptamer position 6 [43–45]. We conclude that the inefficient recombination caused by this alteration is due to a defect in PC formation and suggest that any gene segment whose RSS contains an A at the sixth position of the heptamer will recombine poorly. Consistent with this, A is almost never observed at the sixth position of the heptamer in either the 12- or 23RSS [7].

From the length control of cytoskeletal filaments [46] to the partitioning of molecules in cellular division [47], distributions offer a mechanistic window into constraints on the class of permissible quantitative models. In the case of our efforts to calculate kinetic rates of RAG-RSS dynamics using the PC lifetime distribution, we discerned two interesting findings on the nature of the interaction. Upon first applying our fitting procedure to determine the rates of unbinding and cleavage, we learned that at least one of these two processes did not behave as a simple Poisson process. Even though the exceptionally long dwell times that contribute to the extended tail in the ECDFs may be accounted for in part by the occasional "dead-end" PC where the purified RAG loses function but remains bound to the PC, the different concavities between the ECDFs and theoretical exponential distribution at the earlier lifetimes as shown in Fig. 2.7 suggest a biochemical process involving RAG and RSS that violates a Poisson process. Thinking that our inability to detect nicking was the culprit, we examined the rate of unlooping in the absence of nicking by using Ca^{2+} instead of Mg^{2+} in our reactions. Here, our finding that the PC lifetimes were not exponential for any of the studied RSSs told us that the dwell time distribution convolves other time-sensitive processes with unlooping

and cleavage. These Ca^{2+} results suggest that the PC state may have multiple conformations like the *lac* repressor [48] in that the two RAG1/2 dimers may have multiple states, or that binding to the heptamer and to the nonamer on each RSS are actually separate sequential processes. One possible source of distinct conformations is the dramatic 180° rotation of the DNA that must occur prior to nicking. Rotated and unrotated configurations of un-nicked RSSs have been identified in recent structural studies [14, 15, 49], but would be indistinguishable in the TPM assay. Despite these challenges to obtaining a quantitative description, our data demonstrate that nicking of an RSS is not a prerequisite for RAG to form the PC state, consistent with previous gel shift analyses performed either in Ca^{2+} or with RAG mutants lacking catalytic activity [50–53]. In addition, our findings demonstrate that PCs formed in the presence of Mg^{2+}, which allows for RSS nicking, are longer lived than those formed in the presence of Ca^{2+}, extending previous findings made with RAG bound to single RSSs [34]. While we cannot rule out the possibility that Ca^{2+} perturbs features of the PC other than the ability to undergo nicking, the inference that PCs containing nicked RSS(s) are more stable than those containing intact RSSs is consistent with the formation of a fully "closed" PC conformation after nicking in which additional protein-protein and protein-DNA interactions are observed that should further stabilize the complex [13, 14, 49]. To our knowledge, this study is the first attempt to obtain kinetic rates of unlooping from and cutting of the PC and reveals that there are still key details in the reaction whose temporal behavior has not been observed but ultimately disqualify the PC lifetime distribution from obeying a Poisson process with a single kinetic rate.

The work presented here leaves open several questions about RAG-RSS dynamics. Although our TPM assay detects PC formation and cleavage, it does not detect nicking, preventing us from determining how the RSSs studied influence the rate of nicking or when nicking occurs relative to PC formation. Even without nicking, we see that the unlooping dynamics behave differently from a simple Poisson process. This result suggests a need for an experimental method such as single-molecule FRET [54] that can detect such subtle conformational changes that occur between RAG and the RSS. Finally, we have left the 23RSS unchanged in this study, but it is possible that the trends that we see for our synthetic or endogenous 12RSSs may change with a different partner RSS and shed more light on the "beyond 12/23 rule" [11, 55,

56]. Ultimately, these finer details in the RAG-RSS interaction can provide a more complete kinetic description of the initial phases of V(D)J recombination. While we changed the 12RSS sequence in this work, the TPM assay in principle allows us to titrate other parameters, such as the distance between RSSs, or introduce more biochemical players to better contextualize our work in the bigger picture of recombination *in vivo*.

2.5 Materials and methods

Protein purification

The two RAG components, core RAG1 and core RAG2 (RAG1/2), are purified together as outlined in Ref. [20]. Maltose binding protein-tagged murine core RAG1/core RAG2 were co-expressed by transfection in HEK293-6E suspension cells in a 9:11 w/w ratio for 48 hours before purifying using amylose resin. HMGB1 is purified as outlined in Ref. [20]. His-tagged HMGB1 was expressed in isopropyl-β-D-1-thiogalactopyranoside-induced BL21 cells for 4 hours at 30°C before purification. For more details, see Section A.9 of Appendix A.

Flow cell assembly

TPM flow cells were assembled by drilling four holes along each length of a glass slide before cleaning the slides and cover slips. The slides and cover slips were functionalized with an epoxidizing solution for at least an hour and a half so that anti-digoxigenin, to which the digoxigenin ends of the DNA tethers attach, could adhere to the glass. Upon completion of the treatment, flow cells are assembled by cutting four channels into double-sided tape to connect the drilled holes at opposite ends of the glass slide before adhering to the cover slip on one side and the glass slide on the other. Short connective tubes are inserted into each of the holes to serve as inputs and outputs for fluids and sealed using 5-minute epoxidizing solution. The constructed flow cells are baked on the hot plate to allow the epoxy and double-sided tape to set.

Tethered bead assembly

Tethered beads are assembled as in Fig. A.1. Flow cell channels are incubated with anti-digoxigenin for two hours to allow for adhering DNA to the glass surfaces. After washing away excess anti-digoxigenin in a buffer solution

containing Tris-HCl, KCl, $MgCl_2$, DTT, EDTA, acetylated BSA and casein, engineered strands of 2900 bp-long DNA containing a 12RSS and a 23RSS located 1200 bp apart and tagged with digoxigenin on one end and biotin at the other end are injected into the flow cells to attach the digoxigenin end of the DNA to the anti-digoxigenin-scattered surfaces. After excess DNA is washed out, streptavidin-coated polystyrene beads 490 nm in diameter are added to the channels and incubated for no more than 3 minutes to bind the biotin-labeled end of the DNA. Excess beads are washed away and the TPM assembly buffer is replaced with a RAG reaction buffer containing Tris-HCl, KCl, glycerol, DTT, potassium acetate, $MgCl_2$, DMSO and acetylated BSA. For Ca^{2+} studies, $CaCl_2$ is used in place of $MgCl_2$ in the RAG reaction buffer and in the same concentration. See Section A.1.2 of Appendix A for a schematic of the TPM assembly process.

TPM experiment

TPM experiments involve the simultaneous acquisition of bead trajectories from two different channels on separate microscopes. One of the channels contains tethered DNA with a 12RSS and a 23RSS oriented toward each other (nonamer regions on both RSSs closest to each other). Properly tethered beads are filtered using various methods to ensure proper spacing from neighboring beads and that individual beads are tethered by a single strand of DNA. The trajectories of the selected beads are then examined in the absence of RAG and HMGB1 for ten minutes before flowing in 9.6 nM murine core RAG1/core RAG2 and 80 nM full-length HMGB1 and acquiring bead trajectories for at least one hour. Root-mean-squared displacements (RMSDs) of the bead trajectories as shown in Fig. 2.2 are calculated by Gaussian filtering with an 8-second standard deviation. Bead selection criteria, corrections and smoothing of trajectories, and identification of PCs are provided in the Sections A.1.3-A.2.1 of Appendix A. Example dataset of all analyzed bead trajectories from one replicate is presented in Fig. A.2.

Statistical inference

We used Bayesian and Frequentist methods in this work to calculate parametric and nonparametric quantities, respectively. The PC formation frequencies were assigned confidence intervals via bootstrapping. Briefly, the observed beads and their reported PC formation counts were sampled with replacement

to generate a simulated data set of the same length as the number of observations. The looping frequency was then calculated as the total loops formed among the generated dataset divided by the number of beads and the distribution was resampled again. This procedure was performed 10^6 times and we report various percentiles of these bootstrap replicates, as shown both in the main text and on the paper website. A more detailed explanation is provided in Section A.2.2 of Appendix A.

To compute the cleavage probability and PC leaving rate k_{leave}, we used a Bayesian definition of probability and constructed a posterior distribution for each as is explicitly laid out in Sections A.2.3 and A.2.6 of Appendix A. The displayed posterior distributions for the cleavage probability were generated by numerically evaluating the posterior distribution over a range of cleavage probabilities bounded from 0 to 1. The reported values for the cleavage probability and uncertainty were computed analytically and is derived in Section A.2.3.

To estimate k_{leave} we again constructed a posterior distribution. Here, we chose an exponential form for the likelihood and assumed an inverse Gamma distribution as a prior on the leaving rate. This posterior was then sampled using Markov chain Monte Carlo as is implemented in the Stan probabilistic programming language [57]. A more detailed derivation of the posterior distribution is provided in the Section A.2.6 of Appendix A. All models and code for this inference are available on the paper website.

Significance testing was performed for the looping frequency, median PC lifetime, and fraction of cutting events. Our null hypothesis for each metric was that the measured value for the altered 12RSS was drawn from the same distribution as the V4-57-1 (reference) 12RSS with p-values≤ 0.05 determined to be statistically significant. All p-values for each of these metrics and details about their calculation are provided in Section A.2.4 of Appendix A.

Interactive figures

All results presented in this manuscript are visually complemented with interactive figures on the paper website at `https://www.rpgroup.caltech.edu/vdj_recombination/`. The Cutting Probability Model Explorer shows how the posterior distribution for the cutting probability changes depending upon the number of loops and number of cuts observed, both of which can be ad-

justed with their respective scroll bars. The Synthetic RSS Explorer page displays data for synthetic RSSs. Clicking on individual cells in the paired complex formation frequency, paired complex dwell time, or paired complex cleavage probability heatmaps reveals plots of the looping frequencies with different confidence interval percentages from 10^6 bootstrap replicates; empirical cumulative distribution functions (ECDFs) of the PC lifetimes that revert to an unlooped configuration, are cut, or a combination of the two fates; and full posterior distributions of the probability of cutting for the synthetic RSS in blue and the reference RSS in grey. Number of beads, loops, and cuts observed for the synthetic RSS are displayed by hovering over the cells of the heatmaps. The Endogenous RSS Explorer page displays these same plots but allows for comparison between any two endogenous RSSs studied through dropdown menus, with the data for one RSS displayed in grey, including observation counts, and those for the other RSS shown in blue. The Synthetic-Endogenous RSS Comparison tool provides a means for selecting a particular endogenous RSS by a dropdown menu and directly comparing data for the endogenous RSS (grey) and the individual synthetic RSS that constitutes the sequence difference between the endogenous RSS and the V4-57-1 (reference) RSS, as revealed in the endogenous sequence with highlighted letters where the endogenous and reference RSSs differ.

Data and code availability

All data and code are publicly available. Due to their large volume, raw image files can be obtained upon request. Preprocessed image data can be downloaded from CaltechDATA research data repository under the DOI:10.22002/D1.1288. Processed data files, Matlab, and Python code used in this work can be downloaded either from the paper website or on the dedicated GitHub repository (DOI:10.5281/zenodo.346571).

BIBLIOGRAPHY

[1] S. Tonegawa, "Somatic generation of antibody diversity", Nature 302 (1983), 575.

[2] B. Nadel et al., "Sequence of the spacer in the recombination signal sequence affects V(D)J rearrangement frequency and correlates with nonrandom Vκ usage in vivo", J Exp Med 187 (1998), 1495.

[3] J. A. Weinstein et al., "High-throughput sequencing of the zebrafish antibody repertoire", Science 324 (2009), 804.

[4] N. M. Choi et al., "Deep sequencing of the murine Igh repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies", J Immunol 191 (2013), 2393.

[5] M. Aoki-Ota et al., "Skewed primary Igκ repertoire and V-J joining in C57BL/6 mice: Implications for recombination accessibility and receptor editing", J Immunol 188 (2012), 2305.

[6] D. G. Schatz and P. C. Swanson, "V(D)J recombination: Mechanisms of initiation", Annu Rev Genet 45 (2011), 167.

[7] D. A. Ramsden, K. Baetz, and G. E. Wu, "Conservation of sequence in recombination signal sequence spacers", Nucleic Acids Res 22 (1994), 1785.

[8] J. E. Hesse et al., "V(D)J recombination: A functional definition of the joining signals", Genes Dev 3 (1989), 1053.

[9] D. A. Ramsden and G. E. Wu, "Mouse κ light-chain recombination signal sequences mediate recombination more frequently than do those of λ light chain", Proc Natl Acad Sci 88 (1991), 10721.

[10] Y. Akamatsu et al., "Essential residues in V(D)J recombination signals", J Immunol 153 (1994), 4520.

[11] A. I. Lee et al., "A functional analysis of the spacer of V(D)J recombination signal sequences", PLoS Biol 1 (2003), 56.

[12] P. C. Swanson, "The bounty of RAGs: Recombination signal complexes and reaction outcomes", Immunol Rev 200 (2004), 90.

[13] H. Ru et al., "Molecular mechanism of V(D)J recombination from synaptic RAG1-RAG2 complex structures", Cell 163 (2015), 1138.

[14] M.-S. Kim et al., "Cracking the DNA code for V(D)J recombination", Mol Cell 70 (2018), 1.

[15] H. Ru et al., "DNA melting initiates the RAG catalytic pathway", Nat Struct Mol Biol 25 (2018), 732.

[16] D. A. Schafer et al., "Transcription by single molecules of RNA polymerase observed by light microscopy", Nature 352 (1991), 444.

[17] L. Han et al., "Concentration and length dependence of DNA looping in transcriptional regulation", PLoS ONE 4 (2009), e5621.

[18] S. Johnson, M. Linden, and R. Phillips, "Sequence dependence of transcription factor-mediated DNA looping", Nucleic Acids Res 40 (2012), 7728.

[19] T. Plénat et al., "High-throughput single-molecule analysis of DNA-protein interactions by tethered particle motion", Nucleic Acids Res 40 (2012).

[20] G. A. Lovely et al., "Single-molecule analysis of RAG-mediated V(D)J DNA cleavage", Proc Natl Acad Sci 112 (2015), E1715.

[21] A. Brunet et al., "How does temperature impact the conformation of single DNA molecules below melting temperature?", Nucleic Acids Res 46 (2017), 2074.

[22] F. F. Yin et al., "Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis", Nat Struct Mol Biol 16 (2009), 499.

[23] Y. Zhang et al., "The fundamental role of chromatin loop extrusion in physiological V(D)J recombination", Nature 573 (2019), 600.

[24] R. M. Gerstein and M. R. Lieber, "Coding end sequence can markedly affect the initiation of V(D)J", Genes Dev 7 (1993), 1459.

[25] U. R. Ezekiel et al., "The composition of coding joints formed in V(D)J recombination is strongly affected by the nucleotide sequence of the coding ends and their relationship to the recombination signal sequences", Mol Cell Biol 17 (1997), 4191.

[26] K. Yu and M. R. Lieber, "Mechanistic basis for coding end sequence effects in the initiation of V(D)J recombination", Mol Cell Biol 19 (1999), 8094.

[27] G. H. Gauss and M. R. Lieber, "The basis for the mechanistic bias for deletional over inversional V(D)J recombination", Genes Dev 6 (1992), 1553.

[28] P.-Y. Pan, M. R. Lieber, and J. M. Teale, "The role of recombination signal sequences in the preferential joining by deletion in DH-JH recombination and in the ordered rearrangement of the IgH locus", Int Immunol 9 (1997), 515.

[29] S. Gopalakrishnan et al., "Unifying model for molecular determinants of the preselection Vβ repertoire", Proc Natl Acad Sci 110 (2013), E3206.

[30] R. Subrahmanyam et al., "Localized epigenetic changes induced by DH recombination restricts recombinase to DJH junctions", Nat Immunol 13 (2012), 1205.

[31] D. G. Schatz and Y. Ji, "Recombination centres and the orchestration of V(D)J recombination", Nat Rev Immunol 11 (2011), 251.

[32] O. K. Wong et al., "Interconvertible lac repressor-DNA loops revealed by single-molecule experiments", PLoS Biol 6 (2008), e232.

[33] S. Kumar et al., "Enhanced tethered-particle motion analysis reveals viscous effects", Biophys J 106 (2014), 399.

[34] U. Grawunder and M. R. Lieber, "A complex of RAG-1 and RAG-2 proteins persists on DNA after single-strand cleavage at V(D)J recombination signal sequences", Nucleic Acids Res 25 (1997), 1375.

[35] K. Hiom and M. Gellert, "A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage", Cell 88 (1997), 65.

[36] S. Guilbaud et al., "Dependence of DNA persistence length on ionic strength and ion type", Phys Rev Lett 122 (2019), 028102.

[37] L. G. Cowell et al., "Identification and utilization of arbitrary correlations in models of recombination signal sequences", Genome Biol 3 (2002), 1.

[38] L. G. Cowell et al., "Prospective estimation of recombination signal efficiency and identification of functional cryptic signals in the genome by statistical modeling", J Exp Med 197 (2003), 207.

[39] L. G. Cowell et al., "Computational tools for understanding sequence variability in recombination signals", Immunol Rev 200 (2004), 57.

[40] D. J. Bolland et al., "Two mutually exclusive local chromatin states drive efficient V(D)J recombination", Cell Reports 15 (2016), 2475.

[41] F. Lankaš et al., "DNA basepair step deformability inferred from molecular dynamics simulations", Biophys J 85 (2003), 2872.

[42] S. Jain et al., "CTCF-binding elements mediate accessibility of RAG substrates during chromatin scanning", Cell 174 (2018), 102.

[43] A. J. Feeney, P. Goebel, and C. R. Espinoza, "Many levels of control of V gene rearrangement frequency", Immunol Rev 200 (2004), 44.

[44] A. J. Feeney et al., "A defective Vkappa A2 allele in Navajos which may play a role in increased susceptibility to Haemophilus influenzae type b disease", J Clin Invest 97 (1996), 2277.

[45] B. Nadel et al., "Decreased frequency of rearrangement due to the synergistic effect of nucleotide changes in the heptamer and nonamer of the recombination signal sequence of the V kappa gene A2b, which is associated with increased susceptibility of Navajos to Haemophilus influenzae type b disease", J Immunol 161 (1998), 6068.

[46] L. Mohapatra et al., "The limiting-pool mechanism fails to control the size of multiple organelles", Cell Syst 4 (2017), 559.

[47] N. Rosenfeld et al., "Gene regulation at the single-cell level", Science 307 (2005), 1962.

[48] S. Johnson et al., "Multiple LacI-mediated loops revealed by Bayesian statistics and tethered particle motion", Nucleic Acids Res 42 (2014), 10265.

[49] X. Chen et al., "Cutting antiparallel DNA strands in a single active site", Nat Struct Mol Biol 27 (2020), 119.

[50] K. Hiom and M. Gellert, "Assembly of a 12/23 paired signal complex: A critical control point in V(D)J recombination", Mol Cell 1 (1998), 1011.

[51] S. D. Fugmann et al., "Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex", Mol Cell 5 (2000), 97.

[52] P. C. Swanson, "The DDE motif in RAG-1 is contributed in trans to a single active site that catalyzes the nicking and transesterification steps of V(D)J recombination", Mol Cell Biol 21 (2001), 449.

[53] P. C. Swanson, "Fine structure and activity of discrete RAG-HMG complexes on V(D)J recombination signals", Mol Cell Biol 22 (2002), 1340.

[54] J. Zagelbaum et al., "Real-time analysis of RAG complex activity in V(D)J recombination", Proc Natl Acad Sci 113 (2016), 11853.

[55] A. H. Drejer-Teel, S. D. Fugmann, and D. G. Schatz, "The beyond 12/23 restriction is imposed at the nicking and pairing steps of DNA cleavage during V(D)J recombination", Mol Cell Biol 27 (2007), 6288.

[56] J. K. Banerjee and D. G. Schatz, "Synapsis alters RAG-mediated nicking at Tcrb recombination signal sequences: Implications for the "beyond 12/23" rule", Mol Cell Biol 34 (2014), 2566.

[57] B. Carpenter et al., "Stan: A probabilistic programming language", J Stat Softw 76 (2017), 1.

Chapter 3

GRID FRAP PATTERNING REVEALS A DISPERSIVE EFFECT IN THE BULK OF A LINEARLY CONTRACTING MICROTUBULE NETWORK

[1] S. Hirokawa, H. J. Lee, R. A. Banks, and R. Phillips, "Grid FRAP patterning reveals a dispersive effect in the bulk of a linearly contracting microtubule network", in preparation (2022).

3.1 Abstract

Deliberate photobleaching has provided a means for quantitatively identifying the diffusive, reactive, and convective effects of fluorescent molecules in biological systems, but have been sparsely applied to active systems. We photobleach a grid pattern onto a highly-connected fluorescent microtubule network to observe the reorganization of the filaments in the bulk of the network as they contract to form an aster. Analysis of the fluorescent squares shows that while there is a linear contraction of the microtubules toward the center of the network, the microtubules that make up the fluorescent squares also exhibit a local dispersive effect. Use of different motor speeds and an advection-diffusion equation where diffusion is treated as the null hypothesis

reveals that the dispersion comes largely due to microtubules sliding relative to each other from the motors processing along them.

3.2 Introduction

From the population level where animals move as a collective to the subcellular with the formation of such structures as the mitotic spindle, coordinated movement and emergent patterning can be observed across all biological scales. The question of how such self-organization emerges has led to growing theoretical [1–4] and experimental efforts over the past quarter of a century. While the experimental system of choice varies from the granular particles [5, 6] to the cellular [7, 8], *in vitro* active matter systems using elements of the cytoskeleton offer a powerful means to study the diverse array of structures found in biology [9]. By mixing multimerized motors with filaments, a broad range of ordered patterns emerge [10, 11]. Such systems have become increasingly sophisticated of late with the ability to spatially select where motors couple and drive filament redistribution and thus affect local ordering [12–14]. A common observation from these assays is that the geometry of the light-activating region that induces such selective network formation influences the behavior of the network to produce local order. However, the filament rearrangement that occurs in the bulk of these activated networks is not well known. Recent efforts in the last quarter century have led to several models that quantitatively describe the redistribution of microtubules due to the interactions with motors [15–19]. However, these quantitative descriptions have not been coupled with experimental evidence to reveal the extent of the reorganization in the bulk of the network.

One of the most important tools in microscopy for the imaging of biological systems is the ability to fluorescently label different molecules and image them. This is especially powerful when imaging multiple distinct molecules and using different wavelength-dependent filters to independently measure their position with the same sample. While it is common to avoid exposing fluorescent particles to too much light and permanently turn off their fluorescence, a phenomenon called photobleaching, this can also become a powerful tool. Fig. 3.1 shows this benefit in more detail. A laser can be used to shine an intense beam of light that photobleaches molecules in a select region, leaving the molecules outside of the beam path to remain fluorescent and unaffected (Fig. 3.1B). After some amount of time, the fluorescent molecules can begin to

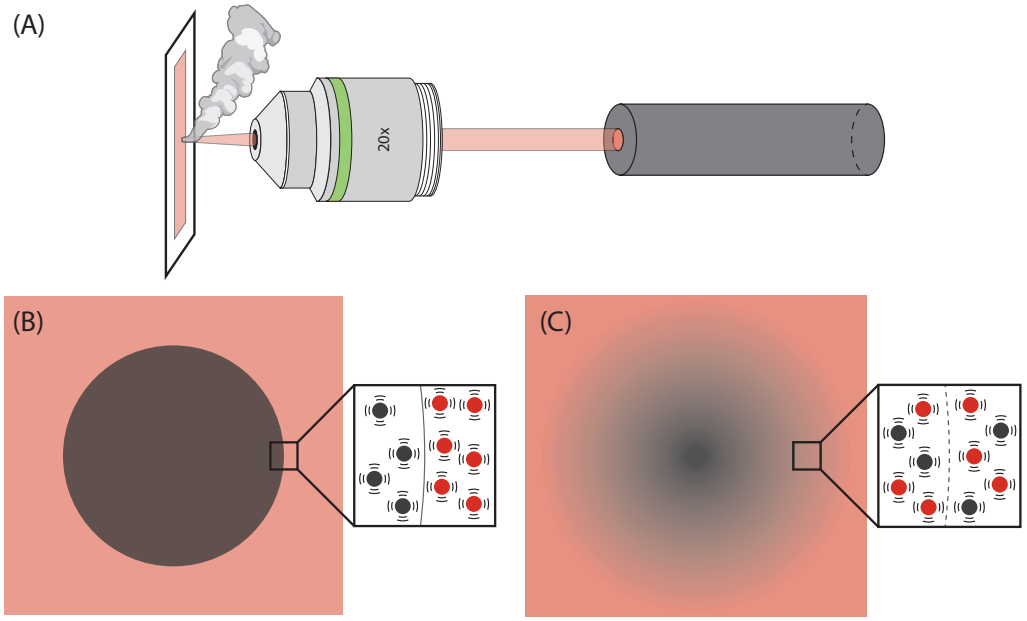


Figure 3.1: **Concept of fluorescence recovery after photobleaching (FRAP).** (A) A laser shines on a sample of red fluorescent molecules in order to bleach the molecules within its path, (B) leaving behind a circular pattern within which all of the particles are dim and outside of which the molecules are still fluorescent. (C) After some amount of time, the fluorescent molecules will fill in the bleached region, leading to a recovery of fluorescence in the area.

populate the bleached region again, leading to recovery of fluorescence (Fig. 3.1C). This technique is known as fluorescence recovery after photobleaching (FRAP).

FRAP has been applied extensively on a range of *in vitro* and *in vivo* systems to examine the movement of fluorescent particles that enter the darkened region [20–24]. These fluorescence recovery studies have typically been tied to various extensions of diffusion equations, whether they further involve convective flow, reaction of molecules, or transport [16, 21–23, 25]. Recent work has demonstrated the effectiveness of FRAP assays in *in vitro* active matter systems, but have been specifically applied to systems of high filament densities [26–28]. Even so, use of photobleaching to examine movement of the key players in active systems has still been applied sparingly in the field and are less common in systems where the filament network contracts to organize over a smaller volume and increase density by orders of magnitude. In such cases of network compaction, bleaching regions within the bulk of the network can

similarly aid in determining the length scale over which filaments are redistributing or elastically contracting.

In the work presented here, we incorporate photobleaching into the light-activated motor dimerization system [13] and examine movement of microtubules in the bulk of the network during the contraction stage on the way to forming an aster. By photobleaching the microtubule network with a grid pattern so as to observe both angular and radial dependencies, we reveal a local dispersal of microtubules amidst a global contraction of the filament network. Use of different motors shows that motor speed changes the contraction rate but maintains the size of this dispersal effect. We further examine this phenomenon using finite element method (FEM) to numerically solve an advection-diffusion model on our photobleach data to validate that the dispersal effect can occur merely due to passive diffusion and that the dispersal is likely generated by the sliding of filaments from the crosslinking motors.

3.3 Results

3.3.1 Photobleaching a grid pattern

In addition to the projector required to induce dimerization of kinesin through the iLid-micro system [29] and image the microtubules, we include a laser for photobleaching with a peak intensity of ≈645 nm and in whose beam path we include a cylindrical lens array mounted to a motorized rotation mount to produce parallel lines of light. A gimbal-mounted mirror allows us to sweep the parallel lines to create thicker lines while the rotation mount allows us to rotate the lens array by 90° before sweeping parallel lines in the orthogonal direction to generate the grid pattern. Further details of the custom microscope build are available in the Materials and methods section and in Sec. B.1.6 of Appendix B.

Fig. 3.2 shows three examples of the grid photobleaching pattern onto a microtubule network at different time points in its life history and the subsequent deformations of the bleached lines and fluorescent squares. These three instances roughly correspond to (A) the initial stages of contraction where the coupled microtubules begin to pull away from the microtubules that fall outside of the activated light region (outlined with a beige dashed line in the pre-photobleaching frame); (B) in the middle of contraction at least two minutes after the network has contracted away from the reservoir of uncoupled

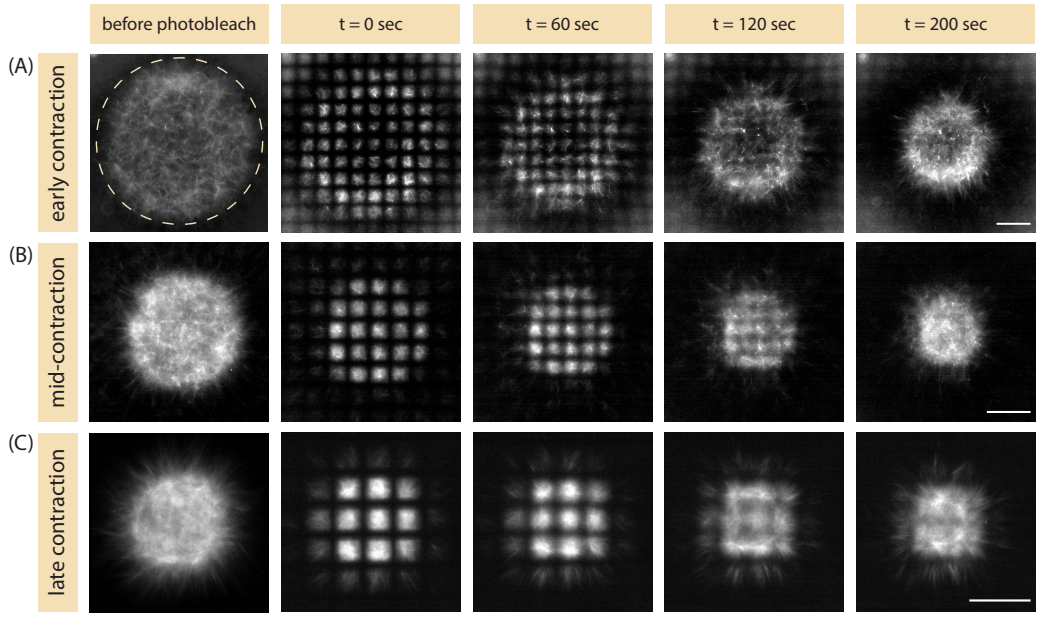


Figure 3.2: **Photobleaching a grid pattern onto the contracting microtubule network.** Here, three different time points during microtubule network contraction are shown where photobleaching is applied: (A) early in contraction as the network pulls away from the reservoir of uncoupled microtubules; (B) in the middle of the contracting phase; and (C) toward the end of contraction before the network no longer appears to shrink. In addition to the image of the network before photobleaching (leftmost column) and the image taken immediately after photobleaching (0 seconds), images of the network 60 seconds, 120 seconds, and 200 seconds after photobleaching are also shown. Dashed line in the pre-bleached early contracting microtubule network outlines the circular activation pattern used to dimerize iLid- and micro-tagged motors to each other. Scale bars in the $t = 200$ sec column apply to their respective rows of images.

filaments; and (C) toward the end of contraction when the size of the network no longer appears to decrease. As the second column in Fig. 3.2 shows, upon photobleaching the grid pattern, individual fluorescent squares, which we will call unit cells, are produced. Over a minute after photobleaching (middle column), unit cells contract toward the center of the network while the photobleached lines appear to thin. By two minutes after photobleaching, neighboring unit cells appear to blend into each other and at later times any remnants of the photobleached pattern disappear.

3.3.2 Tracking of fluorescent squares shows global contraction but local dispersion

To better understand the global network contraction toward the center and the local spreading of fluorescent filaments before filaments of neighboring unit cells begin to interact, we segmented individual unit cells and measured their centroids and areas over successive frames while preserving total fluorescence intensity. This image processing is detailed more closely in Section B.2 of Appendix B. Fig. 3.3A shows that individual unit cells in each experiment are assigned an identification number as labeled at the upper left of each unit cell. Fig. 3.3B shows line plots of the distance of each unit cell centroid as labeled in Fig. 3.3A from the center of the network over successive frames. We see that while each centroid moves roughly linearly in time toward the origin, those centroids further away from the origin appear to move toward the origin faster than those closer to the network center. We computed the slope of the distance-to-time relationship of each unit cell across all of the experiments and compiled them in Fig. 3.3C to find that the contraction speed linearly increases with distance from the center, indicating a general linear contraction of the microtubule network. Upon fitting a line through the origin, we find that the contraction rate $\alpha = 1.6 \times 10^{-3}$ s^{-1} with the 95% credible region falling between 1.4×10^{-3} s^{-1} and 1.8×10^{-3} s^{-1}.

Despite the linear global contraction observed for the centroids, a more macroscopic examination of the unit cells does not show a pure contraction of their constitutive microtubules. In fact, instead of each unit cell contracting into smaller squares, which would be expected for a strictly contracting network subject to the linear contraction speed scaling computed before, we observe a dispersion of the filaments can be seen in successive frames of the photobleached network after contraction. Fig. 3.3D shows the normalized area of each unit cell as a function of time in gray while the mean normalized area is shown in blue. A purely contracting field subject to the linear contraction rate as measured by tracking the movement of the centroids would mean that the normalized areas would fall on the dashed red line, which scales as $(1 - \alpha t)^2$ (see Sec. B.5 of Appendix B), yet each unit cell maintains an area greater than the pure contraction bound. In fact, on average, the unit cells appear to conserve their area. This area-preserving effect illustrates that despite the global contraction of the network, filaments are locally spreading counter to this contraction and reorganizing in the bulk.

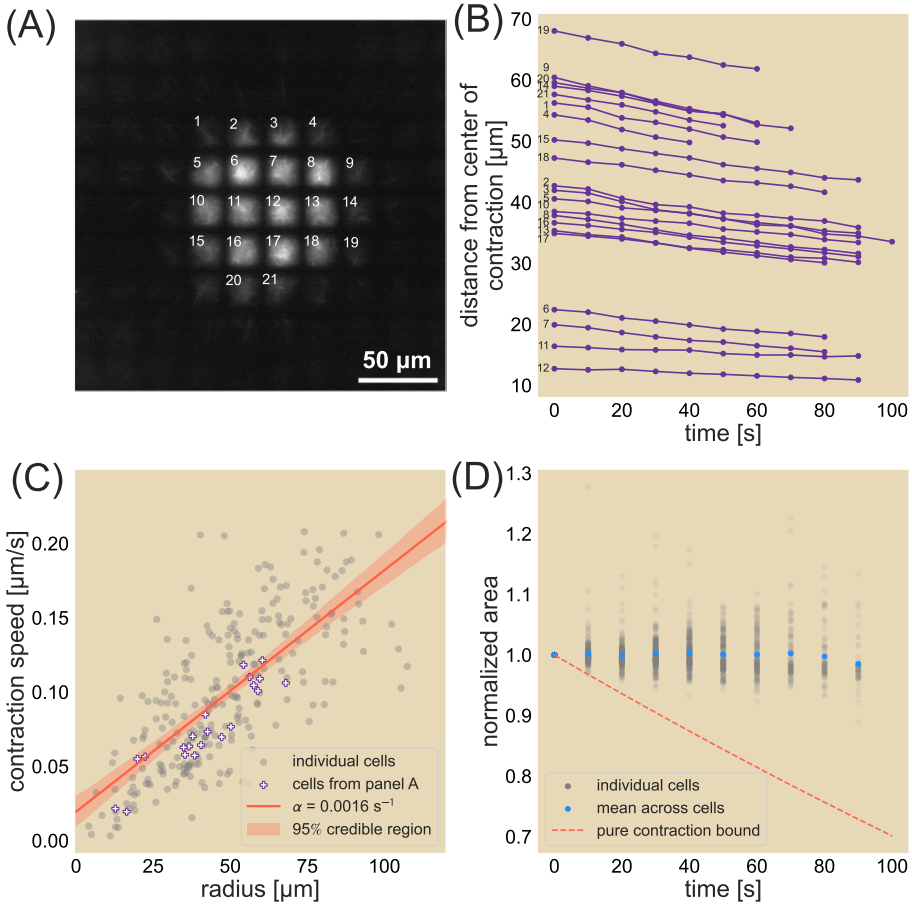


Figure 3.3: **Incorporation of photobleaching into a contracting microtubule network.** (A) Image of the microtubule network after photobleaching with a grid pattern. Numbers in the upper left of each unit cell denote an identification tag for tracking. Data for unit cells toward the outer edges of the network are truncated when cells appear to merge with each other. (B) Centroids of each unit cell are tracked for up to 90 seconds after photobleaching and plotted as distance from the center of the network against time. Numbers to the left of each $t = 0$ sec point denote the unit cell ID as indicated in (A). (C) Plot of unit cell contraction speed as a function of their average distances from the center of the network. The contraction speed of each unit cell is obtained by fitting the distance vs time data found in (B) to a line. The median contraction rate is $\alpha = 1.6 \times 10^{-3}$ s^{-1}. (D) The area of each unit cell is normalized against their initial area as obtained by the unit cell segmentation scheme and plotted as a function of time. The mean normalized area is plotted in blue among individual unit cells (gray). Dashed red line corresponds with the normalized area if the unit cells purely contracted according to the median contraction rate obtained in (C). Number of unit cells decrease at later times as they begin to merge with neighboring cells.

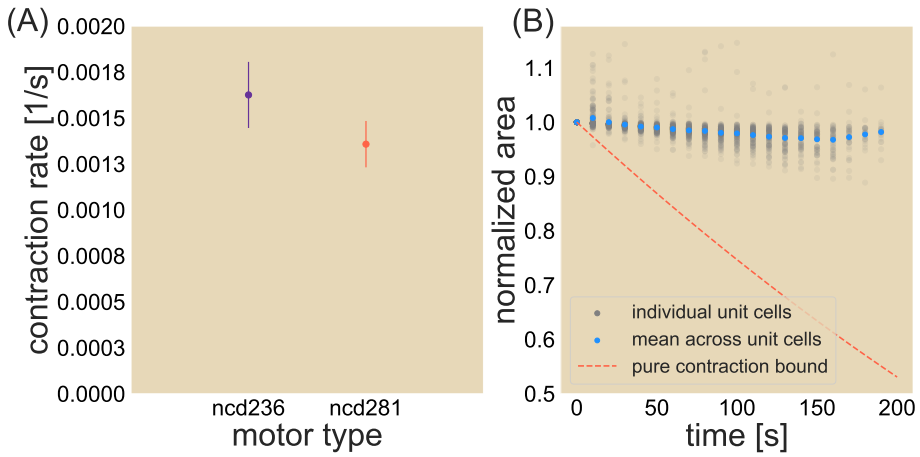


Figure 3.4: **Contraction rate and normalized area of Ncd281.** (A) Comparison of contraction rates of the microtubule network between Ncd236 and Ncd281. Error bars denote range of 95% credible region (B) Normalized area of unit cells over time. Red dashed line denotes expected trajectory of unit cell areas subject purely to contraction rate. Normalized areas of individual unit cells noted in grey with their averages at each time point after photobleaching noted in blue.

3.3.3 Dispersion of microtubules is unaffected by motor speed

The unchanging unit cell areas suggest two possible dispersive effects. One possibility is that some microtubules in the bulk may be disconnected from the network and thus allowed to diffuse until recaptured by a bound motor. If diffusion plays a clear role, then we hypothesize that decreasing the contraction rate would cause the normalized areas to become larger in time while increasing the contraction would drive unit cells to contract faster than diffusion can disperse the microtubules, thus causing unit cell areas to decrease in time. Another possibility is that the motors are driving large reorientation and repositioning of microtubules in the network rather than locking them in relative to their coupled filaments. Such an effect would lead some microtubules to extend outward from the center of unit cells and disperse, particularly if the network is not locally ordered. To test whether motors play a role in the dispersion, we applied the same photobleaching experiments using a variant of the Ncd236 used that is slower without a notable change in ATP hydrolysis rate. This Ncd variant, Ncd281, was shown to be roughly two-thirds the speed of Ncd236 and is otherwise assumed to have roughly the same kinetic properties [30].

Fig. 3.4A shows that upon using the slower Ncd281, the contraction rate of the microtubule network decreases to 1.4×10^{-3} s^{-1} with a 95% credible region between 1.2×10^{-3} s^{-1} and 1.5×10^{-3} s^{-1}. When examining the normalized area of the unit cells generated from the photobleach pattern, unit cells on average remain roughly the same area in the network, decreasing on average to only 0.97 of its original size. Had the dispersion been due to diffusion of the microtubules, the dispersive effect should have made unit cells expand faster than the global contraction would have shrunk them, making the unit cells grow larger than their initial area. However, as the unit cells maintain their size with the slower motor, this preservation of unit cell size is suggestive that the dispersive pattern comes largely due to the continual reorganization of microtubules by the motors. Indeed, we find that even while using faster motors, the unit cells are still roughly the same size. Photobleaching experiments using K401 expressed in bacteria and K401 expressed in insects, which have different motor speeds, show that unit areas remain well preserved.

3.3.4 A grid-like photobleach pattern simultaneously shows the effects of diffusion and advection

Due to the advection of the microtubules toward the center of the network and a seeming diffusive effect that causes the unit cells to disperse, we model the contraction process using an advection-diffusion model of the tubulin concentration $c(r,t)$. At its most generic, such a model has a material flux $\mathbf{J}$ of the form

$$\mathbf{J} = -D\nabla c + \mathbf{v}(\mathbf{r})c, \tag{3.1}$$

where D is the diffusion constant and $\mathbf{v}(\mathbf{r})$ is the velocity profile of the advective flow as a function of distance from the center of contraction r. Here, as motivated by results shown in Sec. 3.3.2, the velocity function is linear with r and negative in the radial direction:

$$\mathbf{v}(\mathbf{r}) = -\frac{v_{\mathrm{m}}}{R}\mathbf{r}. \tag{3.2}$$

Here, we elect to make the constant coefficient a fraction where v_{m} the velocity at $r = R$. As a result, the advection-diffusion model takes the form of

$$\frac{\partial c}{\partial t} = D\nabla^2 c + \nabla \cdot \left[\frac{v_{\mathrm{m}}}{R}\mathbf{r}c\right]. \tag{3.3}$$

To better understand the behavior of the concentration profile subject to this advection-diffusion process, we derived a general solution when Eq. 3.3 is sub-

ject to a no-flux boundary condition to represent our experimental efforts to fully disconnect the coupled microtubules in the iLid-activation region from the reservoir of filaments outside the activation circle. This derivation is available

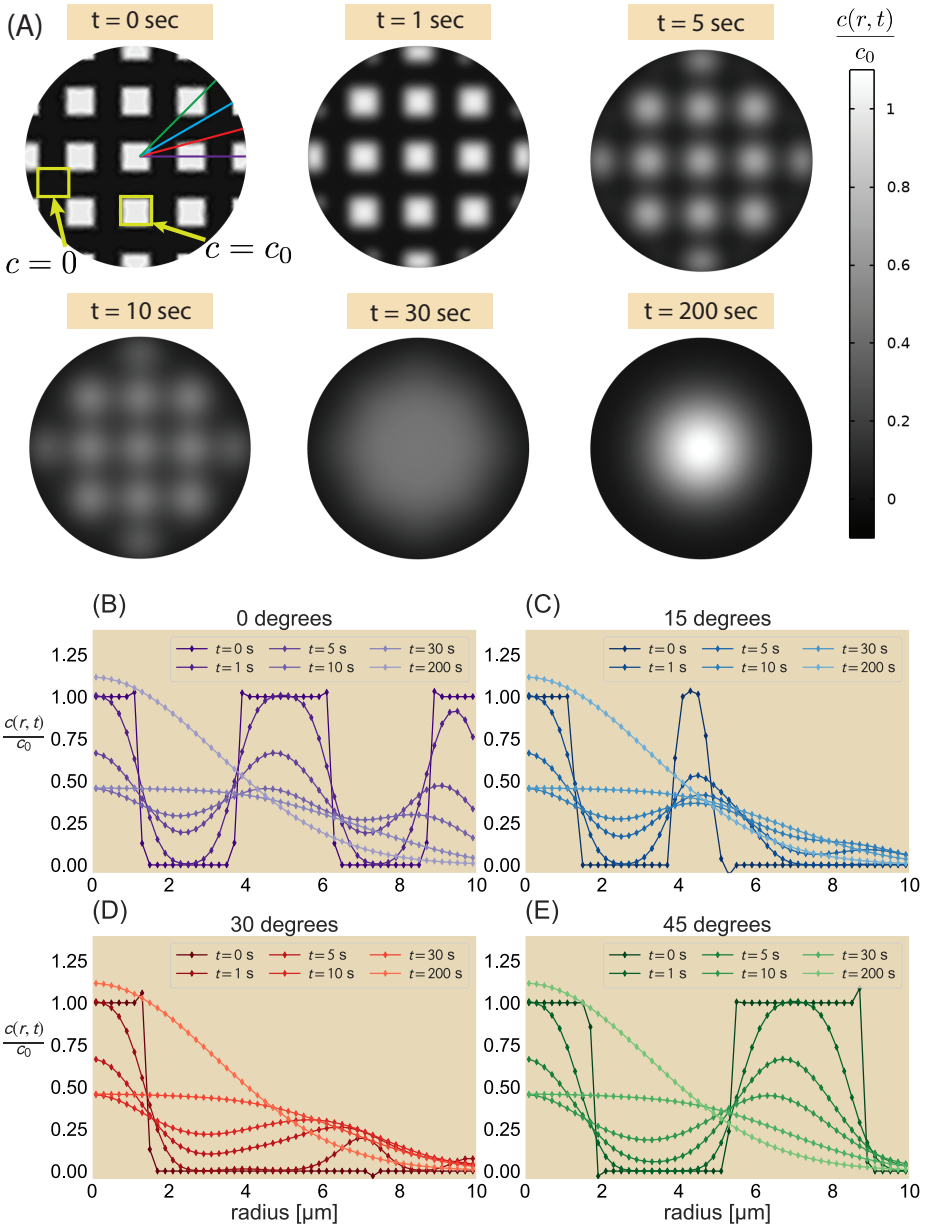


Figure 3.5: **Concentration profile for a grid pattern initial condition at six time points.** (A) Heatmap of the concentration profile in the circle. A line is traced at 0 degrees (purple), 15 degrees (blue), 30 degrees (red), and 45 degrees (green) from the origin to the outer radius of the defined geometry and in correspondence with (B-E), respectively. Later times in the line plots are denoted by increasingly lighter color tones. Here, $D = 0.1\,\frac{\mu m^2}{s}$ and $v_m = 0.1\,\frac{\mu m}{s}$.

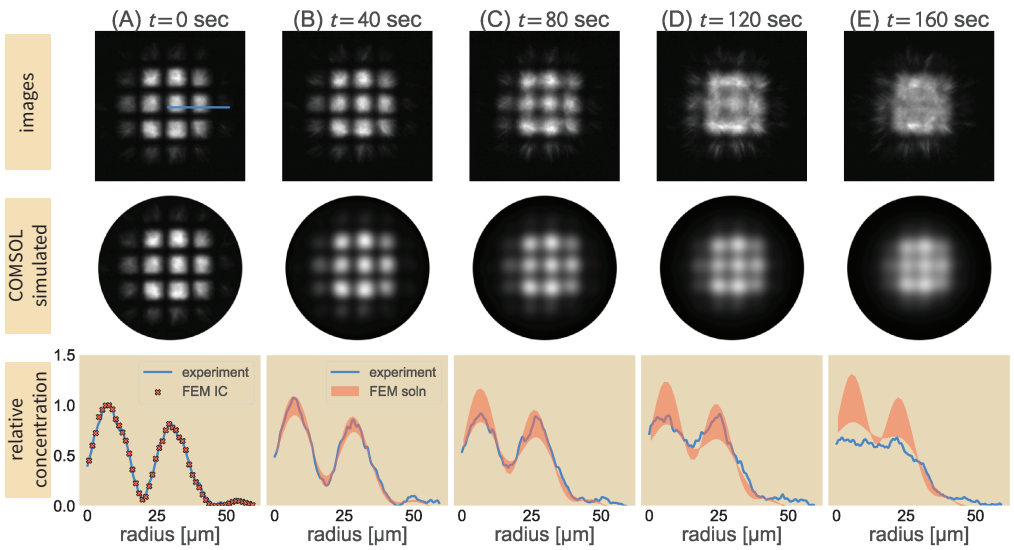


Figure 3.6: **Comparison of advection-diffusion FEM solution to experimental results.** Comparison of experimental image (top row), FEM solution (middle row), and Line profiles of relative tubulin concentration at (A) $t = 0$ sec, (B) $t = 40$ sec, (C) $t = 80sec$, (D) $t = 120$ sec, and (E) $t = 160$ sec. Concentration profiles are from concentration along blue line shown in the $t = 0$ image. Red 'X' in $t = 0$ concentration profile denotes the FEM initial condition implemented in COMSOL. Red shaded regions are set by concentration profiles where 1.8×10^{-3} s^{-1} $\leq \alpha \leq 2 \times 10^{-3}$ s^{-1} and $0.05\,\frac{\mu m^2}{s} \leq D \leq 0.15\,\frac{\mu m^2}{s}$. COMSOL simulated concentration heatmaps are generated using $D = 0.15\,\frac{\mu m^2}{s}$ and $\alpha = 1.9 \times 10^{-3}$ s^{-1}.

in Sections B.10 and B.12 of Appendix B for the 1D and 2D cases, respectively. We further derive three specific solutions each with different initial conditions: (1) a uniform concentration; (2) a uniform concentration but with tubulin removed in a smaller region $r < R_0$ as if performing a circular FRAP assay; and (3) a Gaussian concentration profile with a circular FRAP pattern. This deeper exploration into the model allowed us to determine that the concentration tends toward a Gaussian profile with $\sigma^2 = \frac{D\,R}{v_m}$. These three initial conditions also suggest that the role of D and v_m would be difficult to distinguish with a circular FRAP pattern as the removal of molecules in the inner region produces a concentration gradient that drives diffusion in the same direction as advection, further validating the choice of using a grid pattern in the experiment.

As a result of the agreement found between numerical solutions obtained by finite element methods (FEM) using COMSOL Multiphysics® and the analytical solutions under each of the three initial conditions, we turn fully to the

use of FEM simulations on Eq. 3.3 for the grid photobleaching pattern. We start with the initial condition of a gridlike pattern where the concentration in the gridlines is 0 and those in the individual unit cells are set to c_0. Individual unit cells have side lengths of 2.5 µm and have center-to-center distances of 5 µm from each other. This initial condition is reflected in Fig. 3.5A, where the white squares denote initial concentrations of c_0 while the surrounding black regions mark no molecules. We note that this initial condition has an angular dependence. By subjecting molecules to the same diffusion constant of $D = 0.1\,\frac{\mu m^2}{s}$ and maximum contraction velocity of $v_m = 0.1\,\frac{\mu m}{s}$, we see that the concentration localized to the unit cells disperse faster than the advection pushes the materials toward the origin. However, at longer times such as at $t = 200$ seconds shown in Fig. 3.5A, we observe an accumulation of molecules toward the origin. Fig. 3.5B-E show line traces of the concentration at different angles from the origin to the boundary of the system in 15° increments (and as defined by the four lines in Fig. 3.5A). As each of the line traces illustrate, the concentration profiles spread out at a faster rate than the advective current flows materials toward the origin over the short times, but in the long time limit the concentration accumulates toward the center to generate the familiar Gaussian distribution. We explore the implications of tuning the advection and diffusion parameters as well as the derivation of the Péclet number that reveals this interplay in the Section B.15 of Appendix B

We next examine how well this advection-diffusion model matches that of the experimental data. To do this, rather than manually designing a simulation environment that mimics a typical photobleached MT network, we imported an image of a contracting network in the first frame after photobleaching the filament network and supplied it into COMSOL as our initial condition (implementation in COMSOL is available in Section B.13 of Appendix B). From there, we allowed the system to involve subject to our advection-diffusion model for a range of contraction rate $\alpha \equiv \frac{v_m}{R}$ and diffusion constant D parameters. Fig. 3.6 shows a comparison of the advection-diffusion model to the experimental results at various time points using an imageset where the microtubule network is photobleached toward the end of contraction. Here, we can see that while the advection-diffusion model at first shows good agreement between experiments and numerical results, by about the 160 second mark the experimental results appear to show a more uniform concentration profile than the advection-diffusion results suggests it should have (Fig. 3.6).

An examination of concentration profiles further highlights the limits to the model. We drew a line along both the image and the FEM results, shown as the blue line in the image under Fig. 3.6A. We mark the range of concentration profiles at later timepoints as the red shaded region with a contraction rate $\frac{v_m}{R} = 1.8 \times 10^{-3}$ s^{-1} and diffusion constant $D = 0.15\frac{\mu m^2}{s}$ setting one bound of the shaded region and $\frac{v_m}{R} = 2.0 \times 10^{-3}$ s^{-1} and $D = 0.05\frac{\mu m^2}{s}$ setting the other bound. Here, we can see that the concentration profile of the experiments agrees well with the FEM solution over shorter time scales. At about 2 minutes (Fig. 3.6D), the concentration profile appears to show an increase in tubulin concentration in the local minimum located ≈15 µm from the network center. By $t = 160$ sec, the advection-diffusion model overestimates the concentration of tubulin and does not recapitulate the flattening concentration profile found in the experiments. Thus, the advection-diffusion model, while able to capture the competing effects of the advection toward the network center and the dispersion that has been observed in experiments over shorter time scales begins to diverge from experimental results at later times. This finding further backs up the finding that passive diffusion alone cannot account for the dispersive effect found in the experiments. We further elaborate on the implications of the disagreement between model and observations in the Discussion section.

3.4 Discussion

By imposing a grid photobleaching pattern onto fluorescent squares, we observe that despite a global contraction that scales linearly with the distance from the network center, the microtubules are constantly redistributing and interacting with new microtubules at the local level due largely to the motors moving filaments relative to one another. The dispersive effect observed during the contracting process of the network comes in contrast to other active contraction systems where an elastic contractile effect is the dominant effect [12, 19]. However, the work presented here differs in the choice of motors used (kinesin motors in this work as opposed to dynein for Foster *et al.* and myosin in Schuppler *et al.*) and in the case of Schuppler *et al.*, flexible actin filaments are heavily crosslinked by both active and inactive myosin motors and remain connected to the region outside of the activation zone.

Our results show that regardless of how much the microtubule network has already contracted and by extension independent of the density of microtubules

from its initially uniform distribution to when it is nearing the end of contraction, the area that the fluorescent microtubules from the same unit cell occupy is on average time-independent. From using motors with different speeds and finding that the area of coverage by microtubules belonging to the same fluorescent unit cell is roughly preserved during the contraction process, the dispersion may come from coupled filaments sliding due to polarity sorting, as similarly observed in the dense microtubule limit [26]. We note that previous uses of photobleaching of a microtubule network revealed that the area of the bleached region remains the same, much like the area preservation of the fluorescent squares observed here, and suggested that the filaments did not slide much relative to each other [27]. In contrast, due to the global contraction of the filament network, for the unit cells to maintain their area of fluorescence rather than contract by the contraction rate of the system suggests that the filaments must be constantly exchange with their nearest neighbors throughout the global contraction, leading to a local expansion of individual filaments and thus preserving the area that these fluorescent microtubules sweep out. While work by Tayar *et al.* is performed in the high density limit and thus leads to alignment of filaments locally, our work and the seemingly isotropic dispersion of filaments suggests lower alignment among filaments [27].

When we used the advection-diffusion model (Eq. 3.3) as our null hypothesis and implemented our experimental data as initial conditions, we found that even when setting a range of parameters for the contraction rate and diffusion constant the model only recapitulated the behavior over short time scales before the experimental data showed concentration profiles that would not naturally arise from this model. And although this disagreement between data and the model further supports the unlikeliness that dispersion of microtubules is heavily driven by passive diffusion in the network, it also offers some thoughts as to how the model can be refined.

In many models that describe FRAP results and are used to extract key parameters such as binding rates or diffusion rates from experimental data, the fluorescent and photobleached particles are assumed not to interact with each other beyond their Brownian collisions which are treated no differently than their collisions with the surrounding medium. They may undergo chemical reaction and convert to another species, but oftentimes the implementation of FRAP is employed to study a limited set of extensions of the diffusion

equation, including diffusion alone [31, 32], reaction-diffusion [33], convection-diffusion [34], and transport-diffusion [23]. This assumption means that the time evolution of the distribution of fluorescent molecules is mathematically independent of its photobleached counterparts in the quantitative model. In contrast, because the microtubules are crosslinked by motors, their interactions with each other must be taken into account. This feature of the model also suggests that when trying to quantitatively recapitulate the photobleach data, the photobleached microtubule density must also be factored into the equation. This finding further suggests that rather than assuming a constant diffusion D, one may need to model the diffusion as a function of concentration $D(c)$ in Eq. 3.1. And although we coarse-grained the effects of the motors with a linear advection profile as motivated by the unit cell centroid movement, a natural consideration would be to include time-evolving motor densities and creating a system of coupled time-evolving equations [15, 17, 35]. Such an endeavor would suggest the need to photobleach fluorescent motors to examine their dynamics within the network.

Our use of photobleaching to better understand the bulk reorganization of the network during contraction revealed that microtubules are constantly moving relative to one another and interacting with new partners. While our work has provided deeper insights into the extent of filament redistribution during contraction, much is still not known about the origins of the network formation from the initially randomly oriented, uniformly distributed arrangement of filaments prior to contraction. Specifically, the critical states of the formed network that drives the contraction process is still unclear. Photobleaching as applied in the work presented here may prove useful in examining the extent of filament reorganization early on to better understand what sets the conditions for the network to begin to contract. Our findings also offer motivation for examining redistribution of molecules in other actively driven contexts, such as in systems of opposing motors or subject to more complex iLid-micro activation geometries [13, 36].

3.4.1 Materials and methods

Microscopy set-up.

The microscopy elements used to activate the iLid-micro dimerization and image the different fluorescence channels are similar to those found in Ross *et al.* [13]. Briefly, a digital light processing projector from Texas Instruments was

used to activate the motor dimerization and image the microtubule channels. An excitation filter wheel was placed in front of the projector to filter out the different channels. Photobleaching was performed using a diode laser with a center wavelength of 645 nm. A piezoelectric mirror gimbal mount from Thorlabs was placed downstream of the laser to deflect the beam path over a small range before the laser light passes through a cylindrical lens array inserted into a direct-drive rotation mount. The gimbal mount can then sweep the projected lines laterally to thicken the photobleaching lines before the rotation mount is rotated 90° and the gimbal mount changes the deflecting angle of the beampath in the orthogonal direction. Imaging is performed using a 20x objective. More details are available in the Sec. B.1.6 of Appendix B.

Microtubule network assay

The microtubule network formation and contraction assay is set up similarly as in Ross *et al.* [13]. Micro- and iLid-tagged motors are mixed in equal motor monomer ratios with GMPCPP-stabilized microtubules labeled with Alexa 647 in a reaction mix containing among other components ATP, ATP recycling reagents, and crowding agents. While elements of the oxygen scavenging are kept in the reaction, the glucose oxidase and catalase are removed from the reaction to ensure photobleaching. Removal of these oxygen scavengers minimally affects fluorescence intensity during imaging from using the projector, as shown in Sec. B.3 of Appendix B. In cases of examining the photobleached microtubule field while also knowing the density of the entire microtubule field, Alexa 647-labeled microtubules and Alexa 488-labeled microtubules are mixed in ratios that provide roughly the same signal and used in the reaction.

Image acquisition arrangement

Control of the light-dimerizing activation, photobleach laser activation, and imaging are performed through the Micro-Manager (MM) software [37, 38] while photobleaching is synchronized using a series of in-house compiled executable files that control the movement of the gimbal and rotation mounts. During acquisition, a customized Beanshell script in MM changes the projection pattern on the DLP to create a circular light pattern for the iLid activation and full field for the imaging channels. When the desired state of the microtubule network is reached for performing photobleaching, the script completes the image acquisition cycle before turning on the photobleaching laser and calling to the executables to create the grid before the next acquisition cycle.

Motor purification

Kinesin motors are expressed using the pBiex-1 vector transfected in Sf9 suspension cells. Cells are transfected at 5-7 µg for every 15×10^6 cells at a starting concentration 10^6 cells per mL of Sf900-III media using a liposome-based transfection regent (Escort IV Transfection Reagent). Cells are harvested $\backsim$60-72 hours after transfection and purified using the FLAG affinity tag and anti-FLAG antibody resins. Proteins are stored in 50% glycerol by volume with 1.5 mM DTT, 50 µM EDTA, 50 µM EGTA, and 15 µM ATP and stored at -20°C. Full storage buffers and final concentrations of components are available in Sec. B.1.1 of Appendix B.

BIBLIOGRAPHY

[1] T. Vicsek et al., "Novel type of phase transition in a system of self-driven particles", Phys Rev Lett 75 (1995), 1226.

[2] J. Toner and Y. Tu, "Long-range order in a two-dimensional dynamical XY model: How birds fly together", Phys Rev Lett 75 (1995), 4326.

[3] S. Ramaswamy, "The mechanics and statistics of active matter", Annu Rev Condens Matter Phys 1 (2010), 323.

[4] J. Denk and E. Frey, "Pattern-induced local symmetry breaking in active-matter systems", Proc Natl Acad Sci 117 (2020), 31623.

[5] N. Kumar et al., "Flocking at a distance in active granular matter", Nat Commun 5 (2014), 4688.

[6] A. Kudrolli et al., "Swarming and swirling in self-propelled polar granular rods", Phys Rev Lett 100 (2008), 058001.

[7] Y. Maroudas-Sacks et al., "Topological defects in the nematic order of actin fibres as organization centres of *Hydra* morphogenesis", Nat Phys 17 (2021), 251.

[8] K. Copenhagen et al., "Topological defects promote layer formation in *Myxococcus xanthus* colonies", Nat Phys 17 (2021), 211.

[9] D. Needleman and Z. Dogic, "Active matter at the interface between materials science and cell biology", Nat Rev Mater 2 (2017), 1.

[10] F. J. Nédélec et al., "Self-organization of microtubules and motors", Nature 389 (1997), 305.

[11] T. Surrey et al., "Physical properties determining self-organization of motors and microtubules", Science 292 (2001), 1167.

[12] M. Schuppler et al., "Boundaries steer the contraction of active gels", Nat Commun 7 (2016), 1.

[13] T. D. Ross et al., "Controlling organization and forces in active matter through optically defined boundaries", Nature 572 (2019), 224.

[14] R. Zhang et al., "Spatiotemporal control of liquid crystal structure and dynamics through active patterning", Nat Mater 20 (2021), 875.

[15] H. Y. Lee and M. Kardar, "Macroscopic equations for pattern formation in mixtures of microtubules and molecular motors", Phys Rev E 64 (2001), 056113.

[16] F. Nédélec, T. Surrey, and A. Maggs, "Dynamic concentration of motors in microtubule arrays", Phys Rev Lett 86 (2001), 3192.

[17] S. Sankararaman and G. I. Menon, "Self-organized pattern formation in motor-microtubule mixtures", Phys Rev E 70 (2004), 031905.

[18] J. M. Belmonte, M. Leptin, and F. Nédélec, "A theory that predicts behaviors of disordered cytoskeletal networks", Mol Syst Biol 13 (2017), 1.

[19] P. J. Foster et al., "Active contraction of microtubule networks", eLife 4 (2015), 1.

[20] J. Lippincott-Schwartz, E. Snapp, and A. Kenworthy, "Studying protein dynamics in living cells", Nat Rev Mol Cel Biol 2 (2001), 444.

[21] J. Braga, J. G. McNally, and M. Carmo-Fonseca, "A reaction-diffusion model to study RNA motion by quantitative fluorescence recovery after photobleaching", Biophys J 92 (2007), 2694.

[22] "Mobility of cytoplasmic, membrane, and DNA-binding proteins in *Escherichia* coli", ().

[23] M.-V. Ciocanel et al., "Analysis of active transport by fluorescence recovery after photobleaching", Biophys J 112 (2017), 1714.

[24] J. Lippincott-Schwartz, E. L. Snapp, and R. D. Phair, "The development and enhancement of FRAP as a key tool for investigating protein dynamics", Biophys J 115 (2018), 1146.

[25] D. Axelrod et al., "Mobility measurement by analysis of fluorescence photobleaching recovery kinetics", Biophys J 16 (1976), 1055.

[26] S. Fürthauer et al., "Self-straining of actively crosslinked microtubule networks", Nat Phys 15 (2019), 1295.

[27] A. M. Tayar, M. F. Hagan, and Z. Dogic, "Active liquid crystals powered by force-sensing DNA-motor clusters", Proc Natl Acad Sci 118 (2021), e210873118.

[28] L. M. Lemma et al., "Multiscale microtubule dynamics in active nematics", Phys Rev Lett 127 (2021), 148001.

[29] G. Guntas et al., "Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins", Proc Natl Acad Sci 112 (2015), 112.

[30] N. F. Endres et al., "A lever-arm rotation drives motility of the minus-end-directed kinesin Ncd", Nature 439 (2006), 875.

[31] N. B. Cole et al., "Diffusional mobility of Golgi proteins in membrances of living cells", Science 273 (1996), 797.

[32] R. Swaminathan, C. P. Hoang, and A. S. Verkman, "Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: Cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion", Biophys J 72 (1997), 1900.

[33] M. Kang et al., "A quantitative approach to analyze binding diffusion kinetics by confocal FRAP", Biophys J 99 (2010), 2737.

[34] K. D. Sullivan et al., "Improved model of fluorescence recovery expands the application of multiphoton fluorescence recovery after photobleaching *in vivo*", Biophys J 96 (2009), 5082.

[35] K. Husain and M. Rao, "Emergent structures in an active polar fluid: Dynamics of shape, scattering, and merger", Phys Rev Lett 118 (2017), 078104.

[36] Z. Qu et al., "Persistent fluid flows defined by active matter boundaries", Comms Phys 4 (2021), 1.

[37] A. D. Edelstein et al., "Computer control of microscopes using μManager", Curr Protoc Mol Biol (2010), 14.20.1.

[38] A. D. Edelstein et al., "Advanced methods of microscope control using μManager software", J Biol Methods 1 (2014), e11.

Appendix A

SUPPLEMENTAL INFORMATION FOR CHAPTER 2: SEQUENCE-DEPENDENT DYNAMICS OF SYNTHETIC AND ENDOGENOUS RSSS IN V(D)J RECOMBINATION

A.1 Experimental methods

A.1.1 Microscopy components and configuration

All TPM experiments were performed using two Olympus IX71 inverted microscopes with brightfield illumination. Experiments were run in parallel where one microscope imaged a flow cell containing DNA without any RSSs while the other microscope collected data on DNA strands containing the fixed 23RSS sequence and the 12RSS under consideration. Initially, one microscope (Olympus IX73) was outfitted with a 100x objective while another (Olympus IX73) had a 60x objective with a 1.6x magnifier. Both microscopes used Basler A602f-2 cameras. Partway through the study, each microscope was upgraded to larger fields of view for more data-collection by outfitting the hardware with a 60x objective (Olympus) and a 1920-pixel×1200-pixel monochromatic camera with a global shutter (Basler acA1920-155um). The camera is configured in an in-house Matlab image acquisition script to acquire images at a frame-rate of 30 Hz. Each optical set-up is calibrated to relate DNA of lengths ranging from 300 bp to 3000 bp to the root mean squared distance (RMSD) of their tethered beads.

A.1.2 TPM preparation

A schematic of the tethered bead assembly process as discussed in the Materials and Methods of the manuscript is shown in Fig. A.1. All buffers and assembly components are added to the flow cells by gravity flow. After anti-digoxigenin has coated the coverslip surface, flow cell chambers are washed twice with TPM assembly buffer containing 20 mM Tris-HCl (pH 8.0), 130 mM KCl, 2 mM $MgCl_2$, 0.1 mM EDTA, 0.1 mM DTT, 20 µg/mL acetylated bovine serum albumin (BSA), and 3 mg/mL casein. Once washed, DNA tethers are added and diluted in the TPM assembly buffer to a concentration of roughly 2.5 pM. The tethers are allowed to incubate within the cell for 15 minutes, allowing for the digoxigenin-functionalized ends of tethers to attach to

the anti-digoxigenin-coated coverslip. Unbound excess DNA is then removed from the flow cell and custom-ordered streptavidin-coated beads (Bangs Labs) are added to the flow cells, binding the DNA at the biotin ends, and left to incubate for three minutes before flushing excess beads from system. The prepared flow cell chamber is then equilibrated with RAG reaction buffer containing 25 mM Tris-HCl (pH 7.6), 75 mM KCl, 0.05% glyercol, 1 mM DTT, 30 mM potassium acetate, 2.5 mM $MgCl_2$, 5% DMSO and 100 µg/mL acetylated BSA for TPM experiments involving nicking or else the same buffer except with $CaCl_2$ in place of and at the same concentration as $MgCl_2$ for RAG-RSS interactions in the absence of DNA nicking.

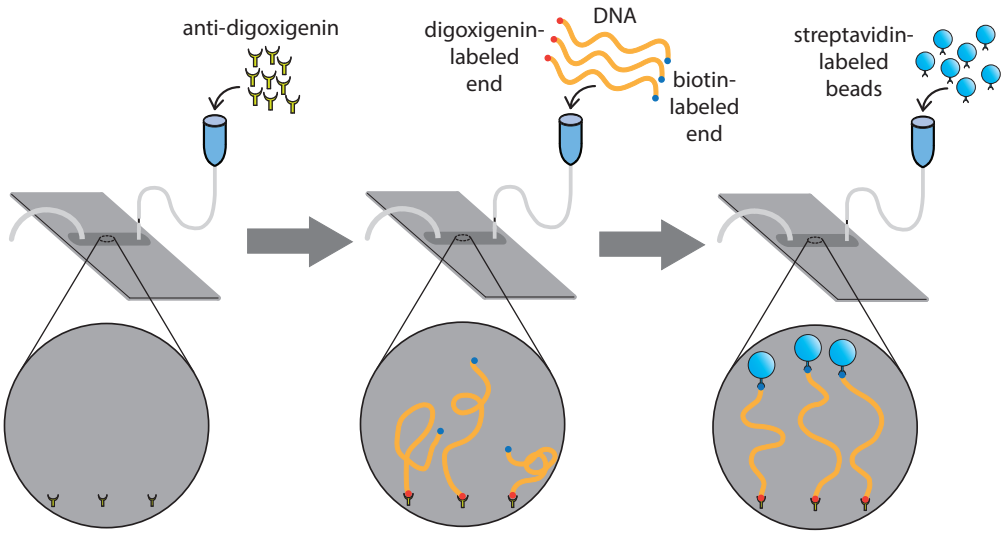


Figure A.1: **Tethered bead preparation process.** Tethered beads are first assembled by adding anti-digoxigenin from Sigma-Aldrich into the flow cell chamber by gravity flow and left to incubate for at least two hours. The fluid is then displaced from the chamber by washing in TPM assembly buffer and introducing DNA tethers containing the desired 12RSS and a constant 23RSS. Unbound DNA tethers are then flushed out and streptavidin-coated beads are introduced to the flow cell. Once the tethered beads have been assembled, chambers are equilibrated with buffer used to study RAG-RSS reaction.

A.1.3 Image processing

Image processing is performed on a field of view in the same manner established by Han *et al.* [1, 2]. After acquiring 60 images over two seconds, beads are identified by setting an intensity threshold before filtering over object sizes. Smaller regions of interest (ROIs) are drawn around each marker identified as

a bead. After initial processing, an additional 120 images over four seconds are acquired and processed by determining intensity-weighted center of masses of beads. The radial root mean squared displacement (RMSD) of the bead position is then determined using the 120 images and compared to the calibration curve based on the expected length of the DNA. Beads are accepted if their RMS values correspond to DNA lengths within 100 bp of their actual lengths for the paired complex assays ($l_{DNA} \approx 2900$ bp). Beads are then further processed to examine their symmetry of motion. After the correlation matrix for the bead position over the 120 frames is obtained, the eigenvalues of the matrix are then obtained, which yield the lengths of the major and minor axes of the range of motion of the bead. If the square root of the ratios of the maximum eigenvalue over the minimum eigenvalue is greater than 1.1, then the asymmetry of the motion is considered to be due to the bead being tethered to multiple DNA strands and is therefore rejected. The remaining beads are kept for data acquisition.

Feedback of the RMSD values of the bead center are obtained during experimentation using a Gaussian filtered by applying an 8-second (240 frame) standard deviation, as done for the post-acquisition processing. To correct for drift in the bead position, often due to the slow unidirectional motion of the microscope stage, the raw data are filtered through a first-order Butterworth filter with a cutoff frequency of 0.05 Hz. All ROI-binned image files can be downloaded from the CaltechDATA research repository under the DOI:10.22002/D1.1288. All code used to analyze these images can be found on the paper website or the paper GitHub repository (DOI:10.5281/zenodo.346571).

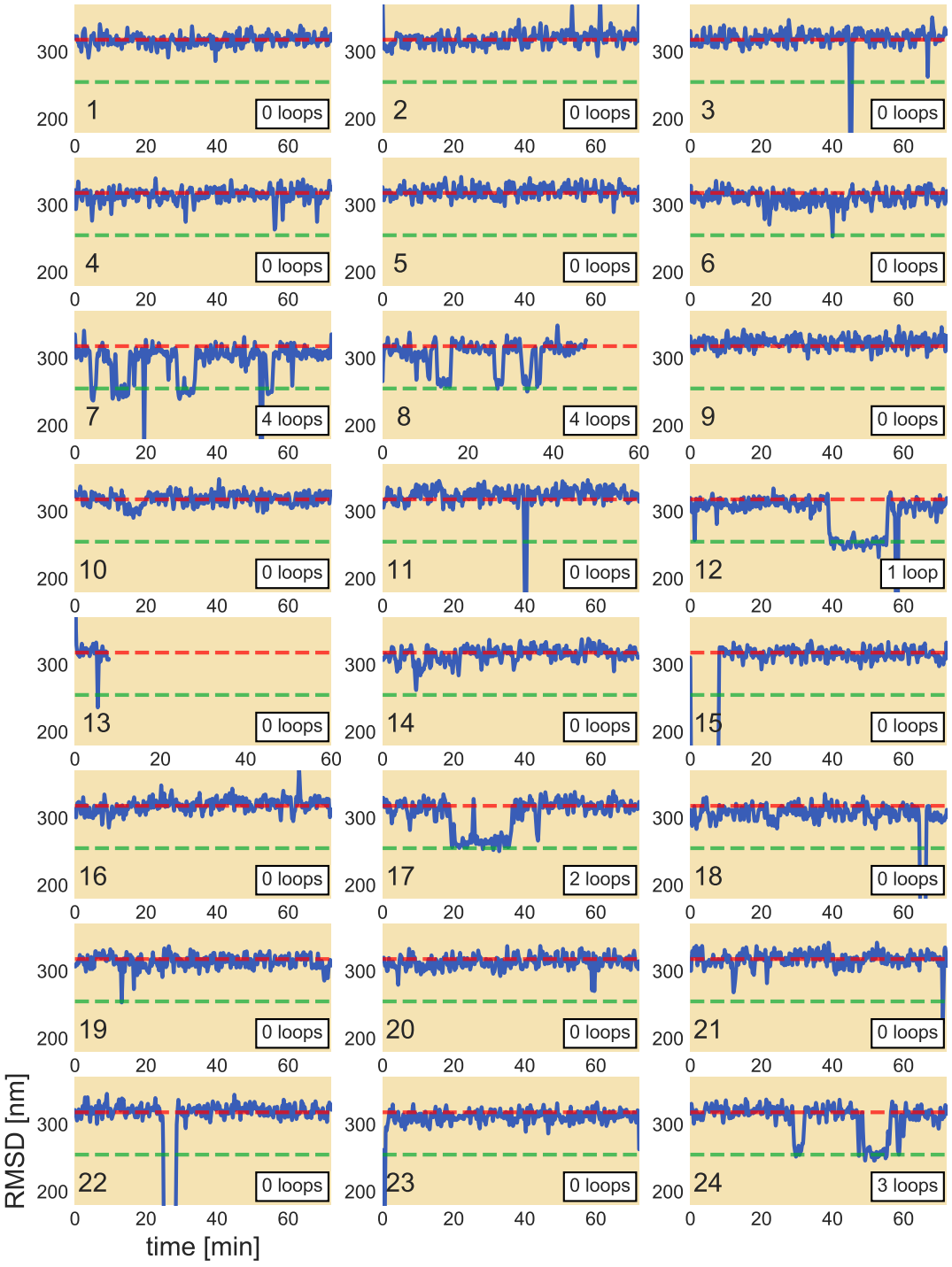
A.2 Data analysis: Extracting all relevant information from bead traces

All of the data reported and used in our results come solely from analyzing the RMSD as a function of time for each individual bead, hereafter called the "bead traces." This source must be further filtered in order to remove beads that passed through the initial image processing steps but still exhibit spurious behaviors, such as sticking to the glass surface or multiple beads falling into the same ROI and confounding the image processing. For each bead, the number of loops formed, the dwell time of each looped state, and whether the loop reverted to the unlooped state or was cleaved by RAG are then extracted and further analyzed through the bootstrapping method for the looping frequency

confidence interval, the Bayesian analysis to obtain our posterior distributions of the cutting probability and the dwell time distributions for our analysis on kinetics of leaving the paired complex state.

A.2.1 Selecting beads for further analysis

Bead selection criteria after preprocessing is applied in the same manner as described elsewhere [1–4]. After correcting for various systematic errors of the experiment, such as slow stage drift, we further smooth the RMSD values of the bead at each instance by applying a Gaussian filter with a -3 dB frequency of 0.0166 Hz corresponding to an 8-second standard deviation. Beads are then manually filtered based upon their RMSD trajectories both before and after introducing RAG and HMGB1 accompanied by 4-second movies of the motion of the bead. Tethers that show multiple attached beads are removed due to a larger variance in the RMSD trajectories for a given state. These beads can also be viewed through a software that shows the raw images at a defined time of the experiment. Furthermore, beads whose traces in the absence of protein lie below the expected RMSD value are considered to be a shorter DNA length than expected or an improperly tethered DNA strand and are also rejected. All other bead trajectories are tracked, as shown with the example set of trajectories from one replicate involving a DNA construct containing the C-to-T deviation at heptamer position 3 of the V4-57-1 (reference) 12RSS in Fig. A.2, until one of four outcomes occurs: 1) RAG cleaves the DNA, causing a sharp increase in RMSD past the tether point and can be observed with the bead diffusing from the ROI (shown for beads 26 and 39). 2) the bead sticks to the glass slide for longer than a few minutes or 3) another bead enters the cropped region enclosing the studied bead due to stage drift that has not been correct, with one of the two outcomes resulting in the truncation of trajectories as in beads 8, 13, 30, or 37. Or, as is also common, 4) the experiment ends, which typically runs for at least one hour of acquisition, without any of these outcomes. In cases where at least one bead reports the looped state at the hour mark without reporting a fate (not shown in this dataset), data acquisition continues until those beads report either unlooping, are untethered, or do not report a fate after roughly 15 minutes of the PC state persisting. The results of one TPM assay, performed with the C-to-T mutation at heptamer position 3 with 39 beads, are shown in Fig. A.2.


RMSD [nm]
time [min]
1 0 loops
2 0 loops
3 0 loops
4 0 loops
5 0 loops
6 0 loops
7 4 loops
8 4 loops
9 0 loops
10 0 loops
11 0 loops
12 1 loop
13 0 loops
14 0 loops
15 0 loops
16 0 loops
17 2 loops
18 0 loops
19 0 loops
20 0 loops
21 0 loops
22 0 loops
23 0 loops
24 3 loops

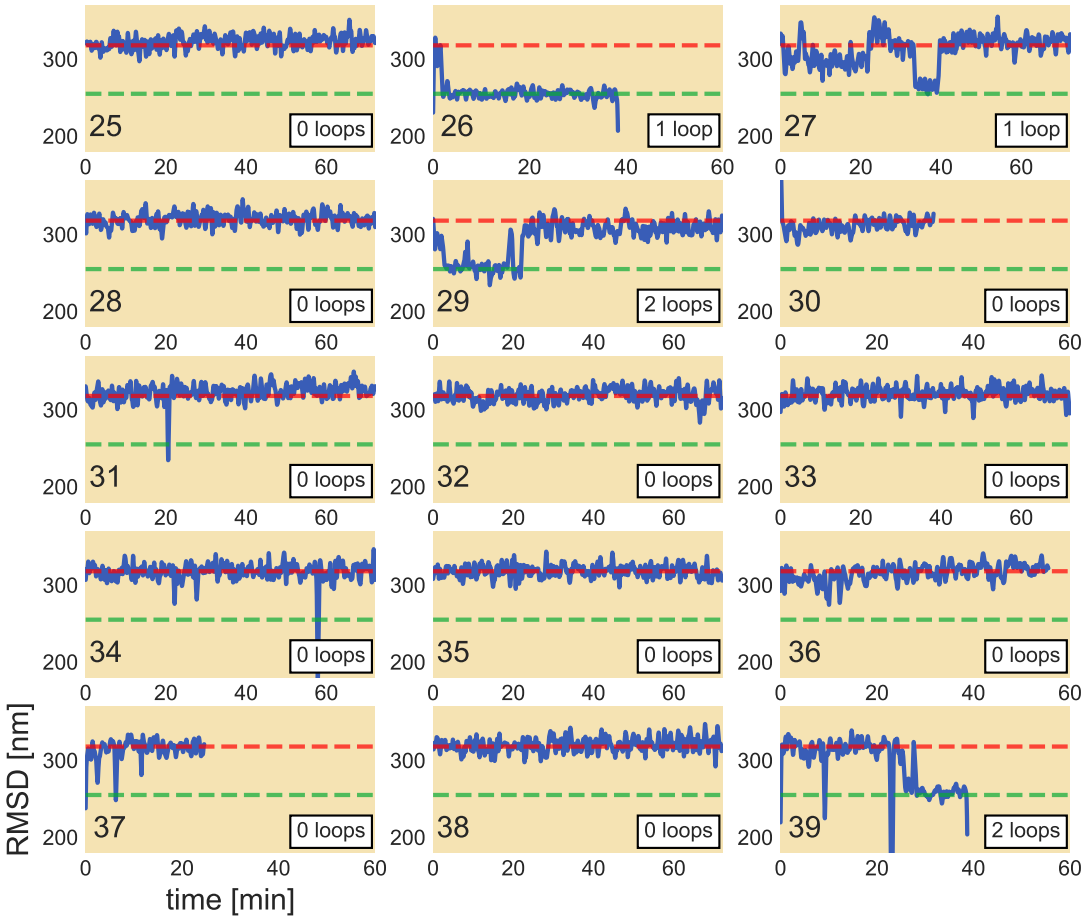


Figure A.2: **Sample bead trajectories for beads that have passed all filters in one replicate.** DNA construct contained the C-to-T alteration at heptamer position 3 of the V4-57-1 (reference) 12RSS. Number in lower left of each bead trajectory denotes bead number. Number of loops denoted in the white box to the lower right of each plot denotes number of loops that the TPM analysis software identifies. Red dashed line shows the empirically-measured root-mean-squared displacement (RMSD) for unlooped DNA length while the green dashed line shows the expected RMSD upon paired complex formation based on the empirically-measured unlooped DNA RMSD. Trajectories where beads reporting a paired complex state stop reporting trajectories, as in beads 26 and 39 are identified as cleaved DNA tethers. Bead trajectories that are truncated before the experiment is terminated but do not show the looped state at the end, such as beads 8, 13, 30, or 37, are not examined past the truncation point because the bead is passively lost from the DNA untethering from the anti-digoxigenin Fab molecule on the coverslip, another bead floating into the field of view and distorting the RMSD analysis, or the bead sticking to the coverslip.

Once the beads have been selected, they are entered into an analysis pipeline that identifies whether a bead is in the unlooped or paired complex state using three thresholding RMSD values at every given instance of data acquisition, as performed in [2]. Looped states are subject to the same 21-second deadfilter as in [2] to be considered as a bona fide paired complex state. In instances where a bead trajectory drops below the minimum RMSD threshold, which is often an indication of temporary adhesion of the bead to the glass slide, or above the maximum RMSD threshold, set due to other temporary aberrations in bead motion, the time that the bead trace spent outside of these bounds are split evenly between the state that the bead was in immediately before and after. With the states of the bead defined at each time point, we can coarse-grain the bead trajectory into the amount of time spent in the paired complex or unlooped states. This allows us not only to determine the lifetime of each paired complex formed but also the number of loops that were formed for a given bead reporter. In addition, all looped states are assigned a binary number based on whether they subsequently lead to unlooping (0) or to the bead untethering (1), the latter of which indicates DNA cleavage by RAG. Data on all beads kept by the TPM data acquisition code, including those that were manually filtered out during post-processing, are available on the CaltechDATA research data repository under the DOI:10.22002/D1.1288.

A.2.2 Bootstrapping looping frequency

As described in Chapter 2, we defined the looping frequency as the total number of observed PC events divided by the total number of beads observed over the experiment. It is tempting to simply repeat this calculation for each experimental replicate, average the results, and report a mean and standard error. However, the number of beads observed can vary greatly from one replicate to another. For example, one replicate may have 20 observed looping events among 100 observed beads, bringing the looping frequency to 0.2. Another replicate of the same RSS may have 0 observed looping events, but among only 10 beads in total, bringing the looping frequency to 0. We would want to apply a penalty to the second measurement as we observed far fewer beads than in the first replicate, however assigning that penalty is also not obvious. To further complicate this calculation, as shown in Fig. A.2, some beads in an experiment will never undergo a looping event while others will show multiple events, making a bead-by-bead calculation of the looping frequency more

challenging.

To address these challenges, we elect to compute and report the looping frequency as the total number of loops observed across all beads and experimental replicates, divided by the number of beads that were studied in total for that particular 12RSS. This metric, being bounded from 0 to ∞, accounts for the fact that for a given 12RSS, looping may occur many times. Furthermore, pooling the beads across replicates results in an appreciably large bead sample size, with the lowest sample size being greater than 80 beads and many RSSs having bead sample sizes in the hundreds.

In order to report a measure of the range of possible looping frequency values that could have been observed for a given RSS, we use the method of bootstrapping on our experimental dataset. In bootstrapping as applied here, we assume that the experimentally-obtained loop count distribution provides the best representation of the population distribution. We can then determine all possible ways we could have obtained the looping frequency by sampling from this empirical distribution. With this bootstrap-generated distribution of possible looping frequency values, we then calculate percentiles to provide confidence intervals on the looping frequency for comparison against the measured looping frequency. To see this in action, suppose our dataset on a particular RSS and salt condition contains N tracked beads across all replicates, with bead i reporting l_i loops. Our measured looping frequency f_{meas} would be $\frac{\sum_i l_i}{N}$. With bootstrapping, we can then determine our confidence interval on the measurement f_{meas} given the bead dataset we obtained with TPM by following the general procedure:

1. Randomly draw N different beads from the dataset of N beads with replacement. This means that the same bead can be drawn multiple times.

2. Sum the total number of loops observed among this collection of N beads and divide by N to get a bootstrap replicate of the looping frequency, $f_{\text{bs},1}$.

3. Repeat this procedure many times. In our case, we obtain 10^6 bootstrap replicates of the looping frequency.

4. For a confidence percentage P, determine the $(50 - \frac{P}{2})^{\text{th}}$ and $(50 + \frac{P}{2})^{\text{th}}$ percentiles from the generated list of 10^6 bootstrapped calculations of the looping frequency.

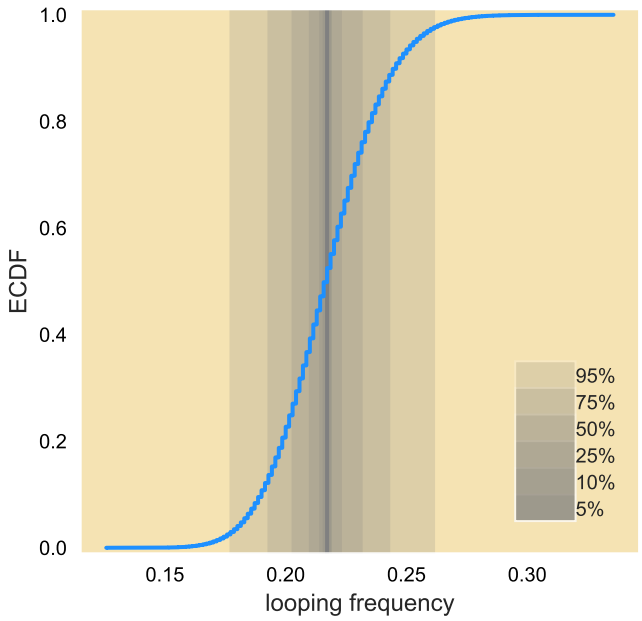


Figure A.3: **Bootstrapped looping frequency and confidence intervals for the V4-57-1 reference sequence.** Empirical CDFs of the bootstrapped looping frequency with 5%, 10%, 25%, 50%, 75% and 95% confidence intervals as represented by the color bar.

As an example, we demonstrate this bootstrap method on the V4-57-1 12RSS, which we also refer to as the reference sequence for our synthetic RSS study. Through TPM, we had tracked 700 beads, each reporting some number of loops l_i. As a result, we draw 700 beads from this dataset with replacement in order to calculate a bootstrap replicate of the looping frequency. We repeat this 10^6 times and obtain the result in Fig. A.3. Although we report the 95% confidence interval in the manuscript, we also offer shades of the 5%, 10%, 25%, 50% and 75% confidence intervals on our website.

A.2.3 Bayesian analysis on probability of cuts

Bayesian analysis on cutting probability is applied in a similar manner to [5]. For a given RSS substrate, to obtain the probability that RAG cuts a paired complex, p_{cut}, we construct a probability density function for p_{cut} conditioned on our data. In this case, our data for each RSS examined is the total number

of loops we observed in TPM, n_{loops}, and the number of loops that were cut, n_{cuts}, so $n_{\text{cuts}} \leq n_{\text{loops}}$. In short, we wish to determine the probability of p_{cut} conditional on n_{loops} and n_{cuts}, or, written concisely, as $P(p_{\text{cut}}|n_{\text{loops}}, n_{\text{cuts}})$. Bayes' Theorem tells us that

$$P(p_{\text{cut}}|n_{\text{loops}}, n_{\text{cuts}})P(n_{\text{loops}}, n_{\text{cuts}}) = P(n_{\text{cuts}}|n_{\text{loops}}, p_{\text{cut}})P(n_{\text{loops}}, p_{\text{cut}}). \quad \text{(A.1)}$$

On the lefthand side Eq. A.1, $P(n_{\text{loops}}, n_{\text{cuts}})$ is the probability of n_{loops} loops and n_{cuts} cut loops, $P(n_{\text{cuts}}|n_{\text{loops}}, p_{\text{cut}})$ is the probability that RAG cuts n_{cuts} loops conditional on the n_{loops} total loops examined and the probability that RAG cuts a given loop p_{cut}. $P(n_{\text{loops}}, p_{\text{cut}})$ is the probability of getting n_{loops} total loops and that RAG has a cut probability p_{cut} for the RSS. A rearrangement of the equation shows that

$$P(p_{\text{cut}}|n_{\text{loops}}, n_{\text{cuts}}) = \frac{P(n_{\text{cuts}}|n_{\text{loops}}, p_{\text{cut}})P(n_{\text{loops}}, p_{\text{cut}})}{P(n_{\text{loops}}, n_{\text{cuts}})}. \quad \text{(A.2)}$$

We can further simplify this equation by noting that the probability of getting n_{loops} loops and a cut probability p_{cut} are independent values. This is evident from the fact that we could have carried out more TPM experiments and in principle p_{cut} should not change from increasing the sample size of loops observed. Thus,

$$P(n_{\text{loops}}, p_{\text{cut}}) = P(n_{\text{loops}})P(p_{\text{cut}}). \quad \text{(A.3)}$$

Furthermore, we can further simplify the probability function in the denominator from noticing that the probability of having n_{loops} total loops and n_{cuts} loops that cut can be broken down into the product of the probability of having n_{cuts} cuts given n_{loops} total loops times the probability of having n_{loops} total loops to begin with, or

$$P(n_{\text{loops}}, n_{\text{cuts}}) = P(n_{\text{cuts}}|n_{\text{loops}})P(n_{\text{loops}}). \quad \text{(A.4)}$$

Inserting equations A.3 and A.4 into equation A.2 gives us

$$\begin{aligned} P(p_{\text{cut}}|n_{\text{loops}}, n_{\text{cuts}}) &= \frac{P(n_{\text{cuts}}|n_{\text{loops}}, p_{\text{cut}})P(n_{\text{loops}})P(p_{\text{cut}})}{P(n_{\text{cuts}}|n_{\text{loops}})P(n_{\text{loops}})}, \\ &= \frac{P(n_{\text{cuts}}|n_{\text{loops}}, p_{\text{cut}})P(p_{\text{cut}})}{P(n_{\text{cuts}}|n_{\text{loops}})}. \end{aligned} \quad \text{(A.5)}$$

We wish to determine the conditional function on the left of Eq. A.5, which we will term our posterior distribution. Here, we construct our posterior distribution from inputting the probabilities on the righthand side of the equation.

We first determine $P(n_{\text{cuts}}|n_{\text{loops}}, p_{\text{cut}})$. This conditional probability function is the probability that we observe n_{cuts} loops cut considering we observe n_{loops} loops forming and if the paired complex has a probability of cutting p_{cut}. Here, we would expect that this is similar to flipping a biased coin n_{loops} times and seeing how many instances heads comes up when the true value of the coin coming up heads is p_{cut}. In this case, we expect this conditional probability to be binomially distributed:

$$P(n_{\text{cuts}}|n_{\text{loops}}, p_{\text{cut}}) = \frac{n_{\text{loops}}!}{n_{\text{cuts}}!(n_{\text{loops}} - n_{\text{cuts}})!}(p_{\text{cut}})^{n_{\text{cuts}}}(1-p_{\text{cut}})^{n_{\text{loops}}-n_{\text{cuts}}}. \quad \text{(A.6)}$$

Next, we would like to determine $P(p_{\text{cut}})$. This is our prior distribution and, because this probability function is not conditioned on any data, this distribution function simply comes from our *a priori* knowledge of p_{cut} independent of the data we have in hand. Here, we choose to say that the only knowledge we have of this parameter is that it, like all probabilities, is bounded between zero and one. We assume that p_{cut} can take any value between zero and one equally. Thus,

$$P(p_{\text{cut}}) = \begin{cases} 1 & 0 \leq p_{\text{cut}} \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad \text{(A.7)}$$

Finally, we need to determine the probability that n_{cuts} loops cut given n_{loops} observed loops. This probability is only conditioned on n_{loops} and not p_{cut}, so we can say that n_{cuts} can take on any integer value between 0 and n_{loops}, inclusive. Thus, we have a discrete uniform distribution:

$$P(n_{\text{cuts}}|n_{\text{loops}}) = \frac{1}{n_{\text{loops}} + 1}. \quad \text{(A.8)}$$

By assembling equations A.6, A.7 and A.8 into equation A.5, we get that

$$P(p_{\text{cut}}|n_{\text{loops}}, n_{\text{cuts}}) = \frac{(n_{\text{loops}} + 1)!}{n_{\text{cuts}}!(n_{\text{loops}} - n_{\text{cuts}})!}(p_{\text{cut}})^{n_{\text{cuts}}}(1-p_{\text{cut}})^{n_{\text{loops}}-n_{\text{cuts}}}. \quad \text{(A.9)}$$

With the posterior distribution in hand, we compute the most probable value of p_{cut} by determining where the derivative of the posterior distribution with

respect to p_{cut} is 0. For ease of calculation, we will take the logarithm of the posterior distribution and derive with respect to p_{cut}:

$$\begin{aligned}\ln[P(p_{\text{cut}}|n_{\text{loops}}, n_{\text{cuts}})] &= \ln\Big[\frac{(n_{\text{loops}}+1)!}{n_{\text{cuts}}!(n_{\text{loops}}-n_{\text{cuts}})!}\Big] + n_{\text{cuts}}\ln(p_{\text{cut}}) \\ &\quad + (n_{\text{loops}}-n_{\text{cuts}})\ln(1-p_{\text{cut}}), \\ \frac{d\ln[P(p_{\text{cut}}|n_{\text{loops}}, n_{\text{cuts}})]}{d\,p_{\text{cut}}}\Big|_{p^*_{\text{cut}}} &= \frac{n_{\text{cuts}}}{p^*_{\text{cut}}} - \frac{n_{\text{loops}}-n_{\text{cuts}}}{1-p^*_{\text{cut}}} = 0. \end{aligned} \tag{A.10}$$

Eq. A.10 then tells us that

$$p^*_{\text{cut}} = \frac{n_{\text{cuts}}}{n_{\text{loops}}}. \tag{A.11}$$

To calculate the variance of p_{cut}, we make the assumption that $n_{\text{loops}} \gg 1$ and look to center about the most probable value, p^*_{cut}. With this assumption, we will approximate the posterior distribution as a Gaussian distribution. In order to see this in action, we will define $x \equiv p - p^*_{\text{cut}}$. Then Eq. A.12 becomes

$$P(p_{\text{cut}}|n_{\text{loops}}, n_{\text{cuts}}) = \frac{(n_{\text{loops}}+1)!}{n_{\text{cuts}}!(n_{\text{loops}}-n_{\text{cuts}})!}(p^*_{\text{cut}}+x)^{n_{\text{cuts}}}(1-p^*_{\text{cut}}-x)^{n_{\text{loops}}-n_{\text{cuts}}}. \tag{A.12}$$

We also invoke the rule that $ln\, n_{\text{cuts}}! \approx n_{\text{cuts}}\ln n_{\text{cuts}} - n_{\text{cuts}} + \frac{1}{2}\ln[2\pi n_{\text{cuts}}]$. We

can then determine the prefactor of the posterior distribution. Specifically,

$$
\begin{aligned}
\frac{(n_{\text{loops}}+1)!}{n_{\text{cuts}}!(n_{\text{loops}}-n_{\text{cuts}})!} &= \exp\{\ln[(n_{\text{loops}}+1)!] - \ln n_{\text{cuts}}! - \ln[(n_{\text{loops}}-n_{\text{cuts}})!]\}, \\
&\approx \exp\{(n_{\text{loops}}+1)\ln(n_{\text{loops}}+1) - (n_{\text{loops}}+1) + \frac{1}{2}\ln[2\pi\,(n_{\text{loops}}+1)] \\
&\qquad - n_{\text{cuts}}\ln n_{\text{cuts}} + n_{\text{cuts}} - \frac{1}{2}\ln(2\pi n_{\text{cuts}}) \\
&\qquad - (n_{\text{loops}}-n_{\text{cuts}})\ln(n_{\text{loops}}-n_{\text{cuts}}) \\
&\qquad + (n_{\text{loops}}-n_{\text{cuts}}) \\
&\qquad - \frac{1}{2}\ln[2\pi(n_{\text{loops}}-n_{\text{cuts}})]\}, \\
&\approx \exp\Big\{(n_{\text{loops}}+1)\Big[\ln n_{\text{loops}} + \ln\Big(1+\frac{1}{n_{\text{loops}}}\Big)\Big] \\
&\qquad - 1 - n_{\text{cuts}}\ln n_{\text{cuts}} \\
&\qquad - (n_{\text{loops}}-n_{\text{cuts}})\Big[\ln n_{\text{loops}} + \ln\Big(1-\frac{n_{\text{cuts}}}{n_{\text{loops}}}\Big)\Big] \\
&\qquad + \frac{1}{2}\ln\Big[\frac{n_{\text{loops}}+1}{2\pi\, n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}\Big]\Big\}, \\
&\approx \exp\Big\{(n_{\text{loops}}+1)(\frac{1}{n_{\text{loops}}} + \frac{1}{2n_{\text{loops}}^2}) - 1 - n_{\text{cuts}}\ln n_{\text{cuts}} + n_{\text{cuts}}\ln n_{\text{loops}} \\
&\qquad - (n_{\text{loops}}-n_{\text{cuts}})\ln(1-p^*_{\text{cut}}) \\
&\qquad + \frac{1}{2}\ln\Big[\frac{n_{\text{loops}}^3}{2\pi\, n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}\Big]\Big\}, \\
&\approx \frac{1}{\sqrt{2\pi\frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}}\exp\Big\{-n_{\text{cuts}}\ln\Big(p^*_{\text{cut}}\Big) - n_{\text{loops}}(1-p^*_{\text{cut}})\ln(1-p^*_{\text{cut}})\Big\}.
\end{aligned}
\tag{A.13}
$$

Here, we make simplifying assumptions, such as that $n_{\text{loops}}+1 \approx n_{\text{loops}}$ and Taylor expansions for $\frac{1}{n_{\text{loops}}}$.

With the prefactor taken care of, we can rework the entire posterior distribution.

$$
\begin{aligned}
P(p_{\text{cut}}|n_{\text{loops}}, n_{\text{cuts}}) \approx & \frac{1}{\sqrt{2\pi\frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}}\exp\Big\{ - n_{\text{cuts}}\ln\Big(p^*_{\text{cut}}\Big) \\
& \qquad - n_{\text{loops}}(1-p^*_{\text{cut}})\ln(1-p^*_{\text{cut}}) \\
& \qquad + n\ln(p^*_{\text{cut}}+x) \\
& \qquad + (n_{\text{loops}}-n_{\text{cuts}})\ln(1-p^*_{\text{cut}}-x)\Big\}, \\
\approx & \frac{1}{\sqrt{2\pi\frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}}\exp\Big\{-n_{\text{cuts}}\ln(p^*_{\text{cut}}) - n_{\text{loops}}(1-p^*_{\text{cut}})\ln(1-p^*_{\text{cut}}) \\
& \qquad + n_{\text{cuts}}\left[\ln(p^*_{\text{cut}}) + \ln(1+\frac{x}{p^*_{\text{cut}}})\right] \\
& \qquad + (n_{\text{loops}}-n_{\text{cuts}})\times \\
& \qquad \left[\ln(1-p^*_{\text{cut}}) + \ln(1-\frac{x}{1-p^*_{\text{cut}}})\right]\Big\}, \\
\approx & \frac{1}{\sqrt{2\pi\frac{n(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}}\exp\Big\{n_{\text{cuts}}\left[\ln(1+\frac{x}{p^*_{\text{cut}}})\right] \\
& \qquad + (n_{\text{loops}}-n_{\text{cuts}})\Big[\ln(1-\frac{x}{1-p^*_{\text{cut}}})\Big]\Big\}, \\
\approx & \frac{1}{\sqrt{2\pi\frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}}\exp\Big\{n_{\text{cuts}}\left[\frac{x}{p^*_{\text{cut}}} - \frac{x^2}{2{p^*_{\text{cut}}}^2}\right] \\
& \qquad + (n_{\text{loops}}-n_{\text{cuts}})\Big[-\frac{x}{1-p^*_{\text{cut}}} - \frac{x^2}{2(1-p^2_{\text{cut}})^2}\Big]\Big\}, \\
\approx & \frac{1}{\sqrt{2\pi\frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}}\exp\Big\{n_{\text{loops}}\,x - n_{\text{cuts}}\frac{x^2}{2{p^*_{\text{cut}}}^2} \\
& \qquad - n_{\text{loops}}\,x - (n_{\text{loops}}-n_{\text{cuts}})\frac{x^2}{2(1-p^*_{\text{cut}})^2}\Big\}, \\
\approx & \frac{1}{\sqrt{2\pi\frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}}\exp\Big\{ - n_{\text{cuts}}\frac{x^2}{2{p^*_{\text{cut}}}^2} - (n_{\text{loops}}-n_{\text{cuts}})\frac{x^2}{2(1-p^*_{\text{cut}})^2}\Big\}, \\
\approx & \frac{1}{\sqrt{2\pi\frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}}\exp\Big\{ - n_{\text{loops}}\frac{x^2}{2p^*_{\text{cut}}} - n_{\text{loops}}\frac{x^2}{2(1-p^*_{\text{cut}})}\Big\},
\end{aligned}
$$

$$\approx \frac{1}{\sqrt{2\pi \frac{n(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}} \exp\Big\{ -\frac{n_{\text{loops}}\, x^2}{2}\Big(\frac{1}{p^*_{\text{cut}}} + \frac{1}{1-p^*_{\text{cut}}}\Big)\Big\},$$

$$\approx \frac{1}{\sqrt{2\pi \frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}} \exp\Big\{ -\frac{n_{\text{loops}}\, x^2}{2}\Big(\frac{1}{p^*_{\text{cut}}(1-p^*_{\text{cut}})}\Big)\Big\},$$

$$\approx \frac{1}{\sqrt{2\pi \frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}}} \exp\Big\{ -\frac{(p-p^*_{\text{cut}})^2}{2\Big[\frac{n_{\text{cuts}}\,(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}\Big]}\Big\}. \tag{A.14}$$

Eq. A.14 tells us that, not only is this Gaussian approximation centered at the most probable value of $p_{\text{cut}} = p^*_{\text{cut}}$, as we would expect, but also that the distribution has a variance of $\sigma^2 = \frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}$. Thus, we report $p^*_{\text{cut}} = \frac{n_{\text{cuts}}}{n_{\text{loops}}}$ and $\sigma^2 = \frac{n_{\text{cuts}}(n_{\text{loops}}-n_{\text{cuts}})}{n_{\text{loops}}^3}$ in Fig. 2.5C and 2.6C of Chapter 2.

A.2.4 Significance testing of looping frequency, median PC lifetime, and cutting fraction

In Chapter 2, we represent particular point mutations and endogenous sequences demonstrating a statistically significant effect with a colored asterisk ($*$). In this section, we elaborate on our definition of what is deemed statistically significant and outline our computational approach.

Defining the null hypothesis

To assess statistical significance of our measurements, we sought to quantify the probability that the observation could be observed under the null hypothesis. For all quantities computed in this work (i.e. looping frequency, PC dwell time, and cutting probability), the null hypothesis was that the observed value of a quantity was drawn from the same distribution as the observed value for the reference (V4-57-1) 12RSS. For each 12RSS and quantity, we computed the probability that an observation generated under the null hypothesis could be at least as extreme as the actual observed quantity. This probability, often reported as a p-value, can be analytically computed using a variety of well-known statistical tests such as the Students' t-test, the Mann-Whitney U-test, and the unequal variance t-test [6]. However, due to the different definitions of the quantities of interest in this work, we used simulation through random number generation as a means to compute this probability.

For all quantities measured, we wished to compute a p-value for the experimental measurement. To do so, we defined a test statistic as the absolute value of the difference in the quantity of interest between a given sequence and that of the reference V4-57-1 12RSS,

$$\delta^{(\text{observation})} = |m_{\text{mutant}}^{(\text{observation})} - m_{\text{reference}}^{(\text{observation})}|. \tag{A.15}$$

Here, $m^{(\text{observation})}$ represents the observed quantity such as looping frequency, median PC dwell time, or the cutting probability computed from the data.

With this test statistic in hand, we reran the experiment *in silico* as follows:

1. Isolate the raw experimental measurements for a given mutant 12RSS and the reference 12RSS and compute the total number of measurements in each dataset, N_{mutant} and $N_{\text{reference}}$.

2. Combine all measurements from both datasets into a single dataset of size $N_{\text{mutant}} + N_{\text{reference}}$ and randomly shuffle the contents.

3. Take the first N_{mutant} entries of the shuffled vector and compute the quantity of interest, $m_{\text{mutant}}^{(\text{simulation})}$. Using the remaining values of the shuffled vector, compute the reference quantity of interest $m_{\text{reference}}^{(\text{simulation})}$.

4. Given these simulated values, compute the value of the test statistic

$$\delta^{(\text{simulation})} = \left| m_{\text{mutant}}^{(\text{simulation})} - m_{\text{reference}}^{(\text{simulation})} \right|. \tag{A.16}$$

5. Store the value of the test statistic in a vector $\vec{\delta}$ and return to step 2. Repeat this process for a total of $N_{\text{simulations}} = 10^7$ times.

6. From the vector of $N_{\text{simulations}}$ test statistic values, compute the p-value by dividing the total number of test values δ_i in the stored vector $\vec{\delta}$ that are greater than or equal to the empirically observed value $\delta^{(\text{observation})}$ and dividing by the 10^7 simulations, or

$$p\text{-value} = \frac{1}{N_{\text{simulations}}} \sum_{i=1}^{N_{\text{simulations}}} k \text{ where } k = \begin{cases} 1, \text{ if } \delta_i \geq \delta^{(\text{observation})} \text{ for } \delta_i \in \vec{\delta} \\ 0, \text{ otherwise.} \end{cases} \tag{A.17}$$

The computed p-values for each sequence and quantity used in this work can be seen in Fig. A.4. In practice, a measurement is determined to be statistically

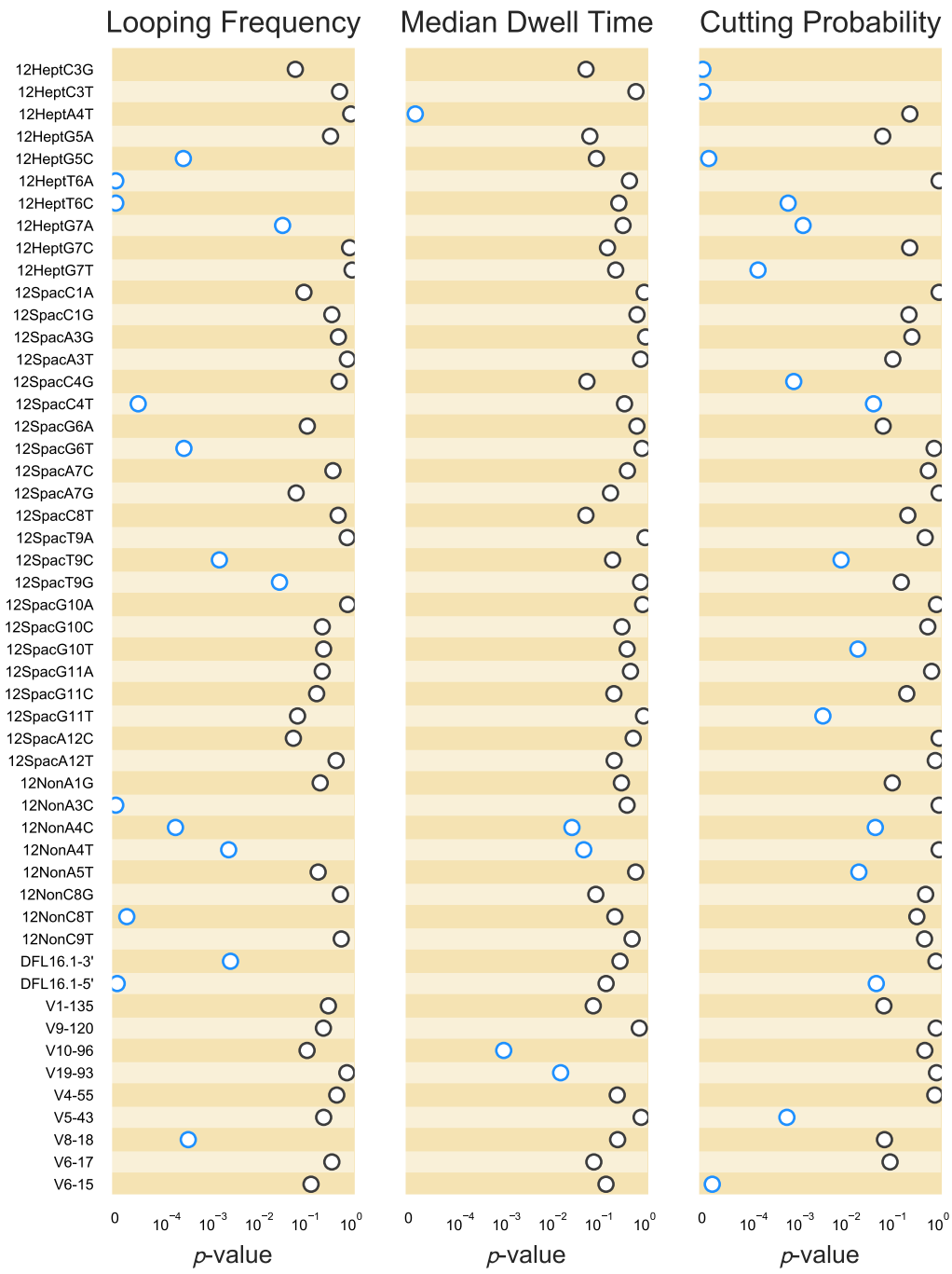


Figure A.4: **Null hypothesis significance testing of looping frequency, median PC lifetime, and cutting fraction for RSSs.** Blue circles denote p-values $p \leq 0.05$.

significant if the p-value is below a given threshold. This threshold is chosen to be the typically chosen 0.05 cutoff value, which indicates that under the null hypothesis, the probability of observing a value at least as extreme as the

experimental measurement is $\leq 5\%$. Measurements with p-value ≤ 0.05 are shown in blue in Fig. A.4.

A.2.5 Relation of looping frequency and cutting probability to bulk *in vitro* cleavage fraction

While we separated different steps of the RAG-RSS reaction through measurements of the looping frequency and cutting probability, we also wanted to know the fraction of DNA tethers that completed the cleavage phase of the reaction. This measurement is applicable to standard bulk *in vitro* cleavage assays where RAG and 12/23RSS-carrying DNA strands are mixed and allowed to carry out the cleavage reaction before running the results on a gel to observe the number of DNA strands cleaved and number that remain intact [7]. Using our TPM data, for each 12RSS construct we calculate the posterior distribution of the fraction of DNA tethers that are cleaved f_{cleaved} based on the total number of tethered beads n_{beads} and the total number of cut tethers n_{cuts},

$$P(f_{\text{cleaved}}|n_{\text{beads}}, n_{\text{cuts}}).$$

By Bayes' theorem,

$$\begin{aligned} P(f_{\text{cleaved}}|n_{\text{beads}}, n_{\text{cuts}})P(n_{\text{beads}}, n_{\text{cuts}}) &= P(n_{\text{cuts}}|n_{\text{beads}}, f_{\text{cleaved}})P(n_{\text{beads}}, f_{\text{cleaved}}), \\ P(f_{\text{cleaved}}|n_{\text{beads}}, n_{\text{cuts}}) &= \frac{P(n_{\text{cuts}}|n_{\text{beads}}, f_{\text{cleaved}})P(n_{\text{beads}}, f_{\text{cleaved}})}{P(n_{\text{beads}}, n_{\text{cuts}})}. \end{aligned} \tag{A.18}$$

Eq. A.18 can be simplified down in the same way as was done with the probability of cutting to yield an equation similar to Eq. A.5.

$$P(f_{\text{cleaved}}|n_{\text{beads}}, n_{\text{cuts}}) = \frac{P(n_{\text{cuts}}|n_{\text{beads}}, f_{\text{cleaved}})P(f_{\text{cleaved}})}{P(n_{\text{cuts}}|n_{\text{beads}})}, \tag{A.19}$$

where p_{cut} is replaced by f_{cleaved} and n_{loops} is replaced by n_{beads}. Each of the functions on the right-hand side of Eq. A.19 shares the same functional form as their counterparts in Eq. A.5: the likelihood function $P(n_{\text{cuts}}|n_{\text{beads}}, f_{\text{cleaved}})$ is also a binomial distribution; we have no prior knowledge of how f_{cleaved} might be distributed, so treat it as uniform from 0 to 1; and n_{cuts} can take on any integer value ranging from 0 (none of the tethers are cleaved) to n_{beads} (all of the tethers are cleaved), so this is also a discrete uniform distribution normalized

by the $n_{\text{beads}}+1$ possible outcomes. Assembling all of these functions into Eq. A.19 yields

$$P(f_{\text{cleaved}}|n_{\text{beads}}, n_{\text{cuts}}) = \frac{(n_{\text{beads}}+1)!}{n_{\text{cuts}}!(n_{\text{beads}}-n_{\text{cuts}})!}(f_{\text{cleaved}})^{n_{\text{cuts}}}(1-f_{\text{cleaved}})^{n_{\text{beads}}-n_{\text{cuts}}}. \tag{A.20}$$

Using the same derivation method as in Section A.2.3, we make a Gaussian approximation to compute the most probable value and standard deviation for f_{cleaved}:

$$f^*_{\text{cleaved}} = \frac{n_{\text{cuts}}}{n_{\text{beads}}}, \tag{A.21}$$

$$\sigma^2_{f_{\text{cleaved}}} = \frac{n_{\text{cuts}}\left(n_{\text{beads}}-n_{\text{cuts}}\right)}{n^3_{\text{beads}}}. \tag{A.22}$$

Fig. A.5 shows, from top to bottom, the looping frequency, cutting probability, and the bead cut fraction. The black dashed line and shaded region for each plot are the point and errorbar equivalent for the V4-57-1 (reference) 12RSS. Specifically, Fig. A.5 shows how the looping frequency and cutting probability can both contribute to limiting the fraction of cleaved DNA tethers. For example, 12NonA3C shows a low looping frequency relative to the reference 12RSS but a similar cutting probability, resulting in a lower bead cut fraction. 12SpacG6T has a higher looping frequency relative to the reference sequence for a comparable cutting probability, yielding a higher bead cut fraction than the reference. We also see that PC cutting probability can limit f_{cleaved}: Even though 12SpacG10T has a similar looping frequency to the reference sequence, the higher cutting probability causes a higher fraction of cleaved tethers. Both changes to heptamer position 3 show that a low cutting probability can abrogate DNA tether cleavage. Both looping frequency and cutting probability as decoupled measurements yield important insights into which processes in the RAG-RSS reaction help or hinder the completion of the cleavage phase in V(D)J recombination.

This observation can also be seen arithmetically. We had defined the looping frequency f_{looped} as the number of loops n_{loops} across all DNA tethers n_{beads}:

$$f_{\text{looped}} = \frac{n_{\text{loops}}}{n_{\text{beads}}}, \tag{A.23}$$

and the most probable value for p_{cut} is the fraction of loops that get cleaved:

$$p^*_{\text{cut}} = \frac{n_{\text{cuts}}}{n_{\text{loops}}}. \tag{A.24}$$

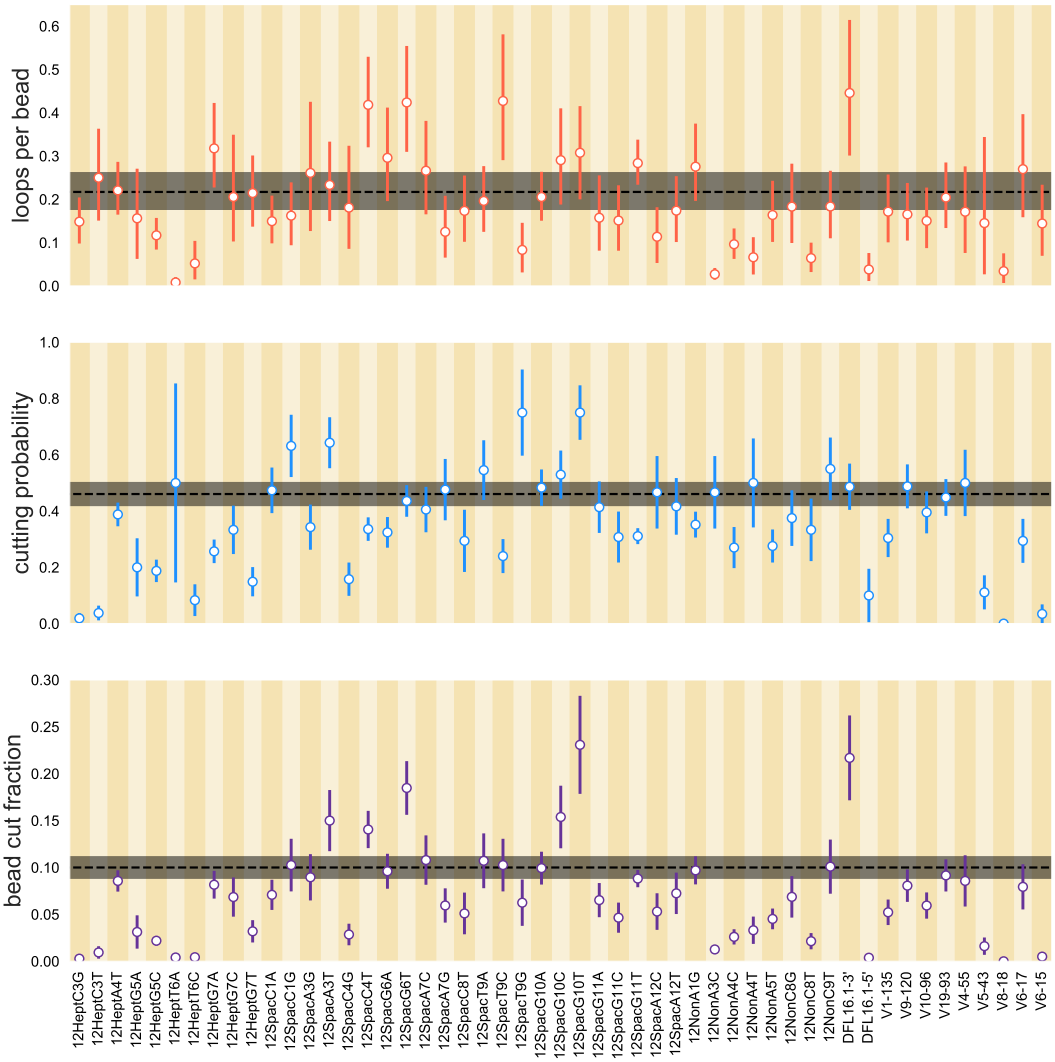


Figure A.5: **Figure stacking of looping frequency (top; red), cutting probability (middle; blue), and bead cut fraction (bottom; purple).** As in the manuscript, looping frequency is shown as the measured value and 95% confidence interval while cutting probability is shown as the most probable p_{cut} and one standard deviation. The bead cut fraction is similarly displayed to the p_{cut} with the most probable f_{cleaved} and one standard deviation. Black dashed line and grey shaded region in each plot corresponds to the measured or most probable value and confidence interval or standard deviation for the V4-57-1 (reference) 12RSS, respectively.

Multiplying both definitions of our metrics allows us to recover Eq. A.21:

$$\begin{aligned} f_{\text{looped}} \times p^*_{\text{cut}} &= \frac{n_{\text{loops}}}{n_{\text{beads}}} \times \frac{n_{\text{cuts}}}{n_{\text{loops}}}, \\ &= \frac{n_{\text{cuts}}}{n_{\text{beads}}}, \\ &= f^*_{\text{cleaved}}. \end{aligned} \tag{A.25}$$

Thus, we recover the relation between the tether cut fraction and the looping frequency and cutting probability, showing that the tether cut fraction will change linearly if one of the metrics is changed due to a change of 12RSS.

A.2.6 Modeling exit from the paired complex as a Poisson process

As discussed in Chapter 2, we attempted to model the kinetics of unlooping and exiting of the paired complex state. In the case of exit, we considered that every paired complex had one of two fates; either the DNA was cleaved and the observed tethered bead was lost or the paired complex dissociated, releasing the bead to its full-length tethered state. We consider these two fates as independent yet competing processes. Under the independence assumption, we can model each process individually as a Poisson process where the time of leaving the paired complex (either through cleavage or unlooping) is exponentially distributed. Mathematically, we can state that the probability of leaving the paired complex at time t_{leave} is defined as

$$P(t_{\text{leave}} \,|\, k_{\text{leave}}) = k_{\text{leave}} e^{-k_{\text{leave}} t_{\text{leave}}}, \tag{A.26}$$

where the leaving rate k_{leave} is defined as the sum of the two independent rates,

$$k_{\text{leave}} = k_{\text{cut}} + k_{\text{unloop}}. \tag{A.27}$$

Therefore, given a collection of paired complex dwell times t_{leave}, we can estimate the most-likely value for k_{leave} providing insight on whether exiting the paired complex can be modeled as a Poisson process.

Rather than reporting a single value, we can determine the probability distribution of the parameter k_{leave}. This distribution, termed the posterior distribution, can be computed by Bayes' theorem as

$$P(k_{\text{leave}} \,|\, t_{\text{leave}}) = \frac{P(t_{\text{leave}} \,|\, k_{\text{leave}}) P(k_{\text{leave}})}{P(t_{\text{leave}})}. \tag{A.28}$$

The posterior distribution $P(k_{\text{leave}} \,|\, t_{\text{leave}})$ defines the probability of a leaving rate *given* a set of measurements t_{leave}. This distribution is dependent on the likelihood of observing the dwell time distribution *given* a leaving rate, represented by $P(t_{\text{leave}} \,|\, k_{\text{leave}})$. All prior information we have about what the leaving rate *could* be is captured by $P(k_{\text{leave}})$ which is entirely independent of the data. The denominator in Eq. A.28 defines the probability distribution of the data marginalized over all values of the leaving rate. For our purposes, this term serves as a normalization constant and will be neglected.

We are now tasked with defining functional forms for the various probabilities enumerated in Eq. A.28. The likelihood already matches the definition in Eq. A.26, so we assign our likelihood as a simple exponential distribution parameterized by the leaving rate. Choosing a functional form for the prior distribution $P(k_{\text{leave}})$ is a much more subjective process. As such, we outline our thinking below.

As written in Eq. A.26, k_{leave} has dimensions of inverse time, meaning that particularly long-lived paired complexes will have $k_{\text{leave}} < 1$ whereas a sequence with unstable paired complexes will have $k_{\text{leave}} > 1$. As we remain ignorant of our data, we consider both of these extremes to be valid values for the leaving rate. However, this parameterization raises technical issues with estimating k_{leave} computationally. We sample the complete posterior using Markov chain Monte Carlo, a computational technique in which the posterior is explored via a biased random walk depending on the gradient of the local probability landscape. With such a widely constrained parameter, effectively sampling very small values of k_{leave} becomes more difficult than larger values. We can avoid this issue by reparameterizing Eq. A.26 in terms of the inverse leaving rate $\tau_{\text{leave}} = \frac{1}{k_{\text{leave}}}$ so that

$$P(t_{\text{leave}} \,|\, \tau_{\text{leave}}) = \frac{1}{\tau_{\text{leave}}} e^{t_{\text{leave}}/\tau_{\text{leave}}}. \tag{A.29}$$

Our parameter of interest now has dimensions of time and can be interpreted as the average life time of a paired complex or, more precisely, the waiting time for the arrival of a Poisson process.

While it is tempting to default to a completely uninformative prior for τ_{leave} to avoid introducing any bias into our parameter estimation, we do have some intuition for what the bounds of the value could be. For example, it is mathematically impossible for the leaving rate to be less than zero. We can also find it unlikely that the leaving rate is *exactly* zero as that would imply irreversible formation of the paired complex. We can therefore say that the value for the leaving rate is positive and can asymptotically approach zero. As we have designed the experiment to actually observe the entry and exit of the paired complex state, we can set a soft upper bound for the leaving rate to be the length of our typical experiment, 60 minutes. With these bounds in place, we can assign some probability distribution between them where it is impossible to equal zero and improbable but not impossible to exceed 60 minutes.

A good choice for such a distribution is an inverse Gamma distribution which has the form

$$P(\tau_{\text{leave}} \,|\, \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \frac{\beta^{\alpha}}{\tau_{\text{leave}}^{(\alpha+1)}} e^{-\beta/\tau_{\text{leave}}}, \tag{A.30}$$

where α and β correspond to the number of arrivals of a Poisson process and their rate of arrival, respectively. Given that only one arrival is necessary to exit a paired complex, we choose α to be approximately equal to 1 and β to be approximately 10. This meets our conditions described previously of asymptotically approaching zero and rarely exceeding 60 minutes.

Combining Eq. A.29 and Eq. A.30 yields the complete posterior distribution. We sampled this distribution for each RSS independently using Markov chain Monte Carlo. Specifically, we used Hamiltonian Markov chain Monte Carlo as is implemented in the Stan probabilistic programming language [8]. The code used to sample this distribution can be accessed on the paper website or GitHub repository.

A.3 Posterior distributions of the endogenous sequences

Fig. A.6 gives the full posterior distributions of the cutting probability for each of the endogenous RSSs. We see clearly that between the two RSSs flanking the DFL16.1 gene segment that RAG is more successful at cleaving the RSS on the 3' side of the gene segment than the RSS on the 5' end. In examining the RSSs adjacent to endogenous Vκ gene segments, we see that the cutting probability is not differentiable across most of the RSSs, but cleavage is dramatically reduced when RAG interacts with the V5-43, V8-18 and V6-15 RSSs. We find that the number of paired complexes formed with the V8-18 12RSS is low to begin with, leading to an uninformative posterior distribution, whereas the V6-15 12RSS may suffer a low cleavage probability due to the T immediately adjacent to the heptamer in the coding flank region, which has been known to broadly reduce recombination efficiency [9–11].

A.4 Ca^{2+}-Mg^{2+} looping frequency comparisons

Although we directly compared the dwell time distributions of three RSS constructs between when the RAG reaction buffer contained Mg^{2+} to allow for nicking and buffer containing Ca^{2+} to prevent nicking, we wanted to know whether the looping frequency would increase when RAG is prohibited from nicking. Our intuition comes from recognizing that without the ability to

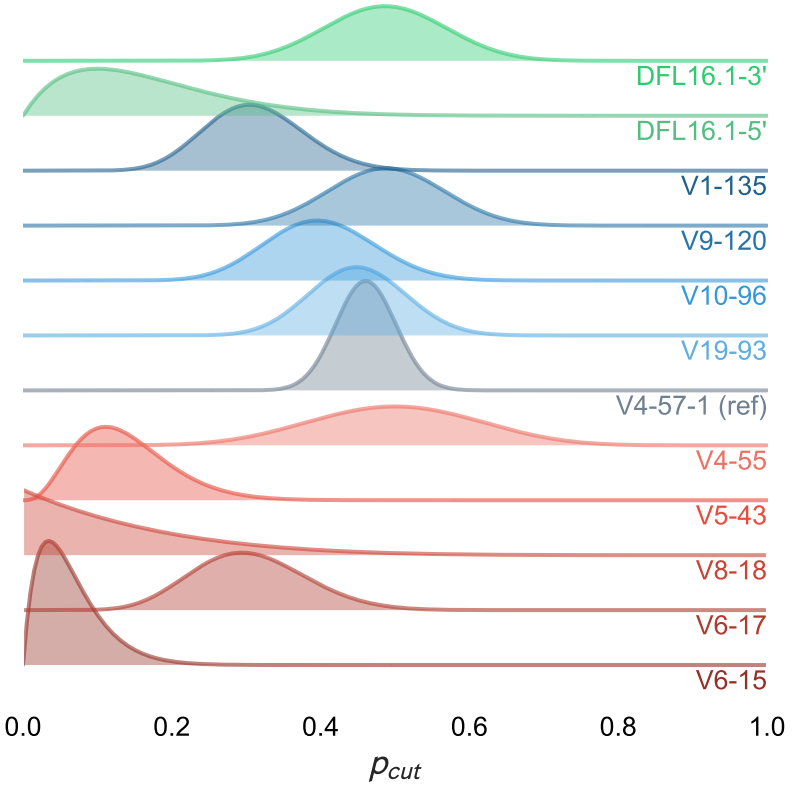


Figure A.6: **Posterior distributions of the cutting probabilities as derived in Section A.2.3 for the endogenous 12RSSs studied.** The top-to-bottom order of the endogenous RSSs is the same as their left-to-right ordering in Fig. 3. Height of the distribution is proportional to the probability of the p_{cut} value.

cleave the DNA, RAG can only release one of the RSSs and leave the paired complex state without cutting the DNA tether. As a result, RAG has an opportunity to form the paired complex with the same DNA tether. We expect that the looping frequency should either increase or remain the same in the Ca^{2+} environment as compared to when Mg^{2+} is used. Fig. A.7 shows that these two outcomes result. Fig. A.7A and A.7C show that RAG forms the paired complex more frequently with the reference sequence and the G-to-T change at the eleventh position of the reference spacer sequence when the reaction occurs in Ca^{2+}. Furthermore, we see that undergoing the reaction with the A-to-T alteration at heptamer position four in Ca^{2+} does not induce much change in the looping frequency as compared to a Mg^{2+} environment (Fig. A.7). Of interest is the fact that the spacer variant, which has a slightly larger measured looping frequency than the reference sequence in Mg^{2+} with over-

lapping 95% confidence intervals, clearly undergoes a more dramatic increase in looping frequency than the reference sequence when the salt is Ca^{2+}. This observation shows that PC formation is more favorable for the spacer variant than the reference sequence. Observed holistically, we find that RAG in the absence of nicking can form loops at least as frequently as when it when it can nick the DNA.

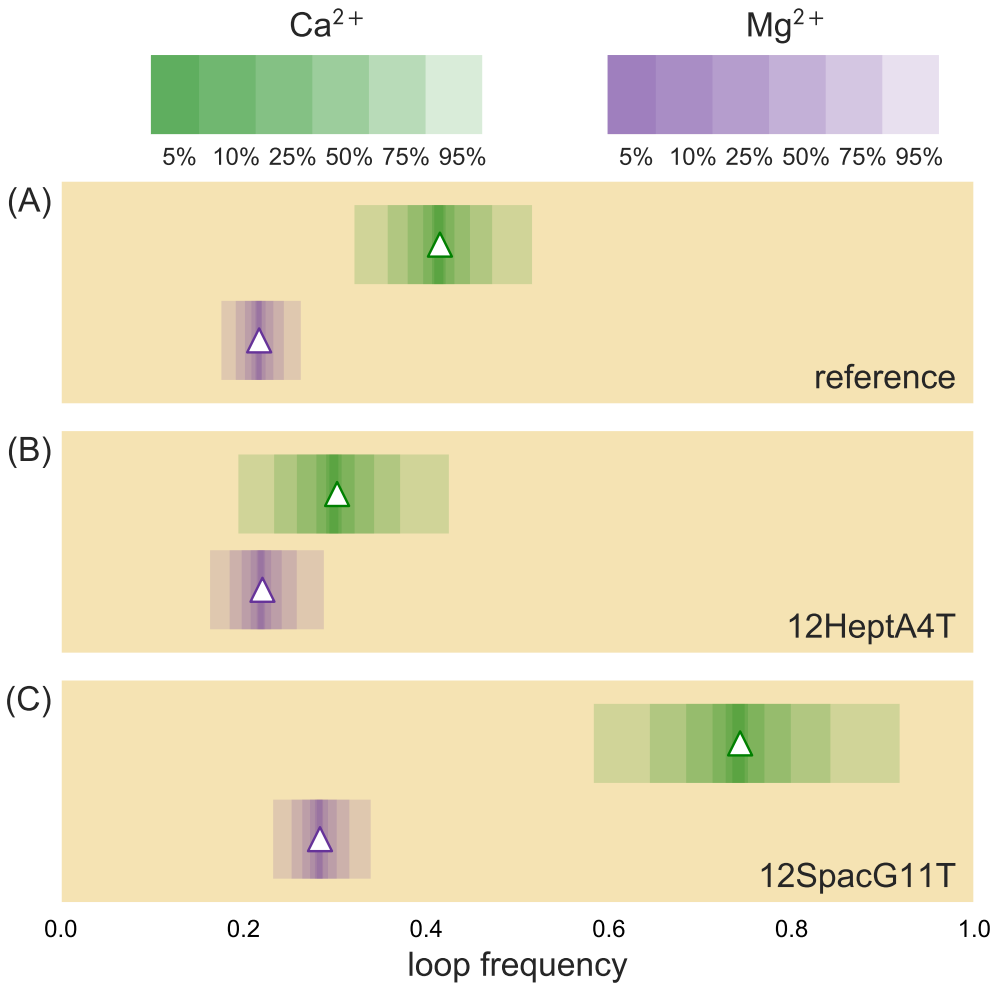


Figure A.7: **Ca^{2+} (green) and Mg^{2+} (purple) looping frequencies** for (A) reference 12RSS, (B) A-to-T change at the fourth position of the heptamer and (C) G-to-T change at the eleventh position of the spacer. Measured looping frequency shown as the triangles. Going from darker shading to lighter shading in rectangle bar indicates increasing of confidence interval percentage of the looping frequency from the bootstrapping method discussed in section A.2.2.

A.5 Coding flank contributions

For our study of the endogenous RSSs, we also modified the coding flanks adjacent to the RSSs during the cloning process to construct the DNA tethers. As shown in table A.1, most of these coding flanks have A and C nucleotides in the two or three base pairs upstream of the heptamer region. However, recent structural work have shown direct contacts between RAG1 residues and the coding flank [12–14]. Furthermore, various bulk assays have demonstrated that coding flank sequence can affect recombination efficiency [9–11]. These bulk assays suggest that coding flanks with A and C nucleotides near the heptamer tend to recombine more efficiently than those that have Ts instead. In attempting to determine whether these A- and C-rich coding flanks have much of an influence on the RAG-RSS dynamics, we looked to two pairs of TPM constructs where within each pair the RSS is identical, but the coding flank sequence is different.

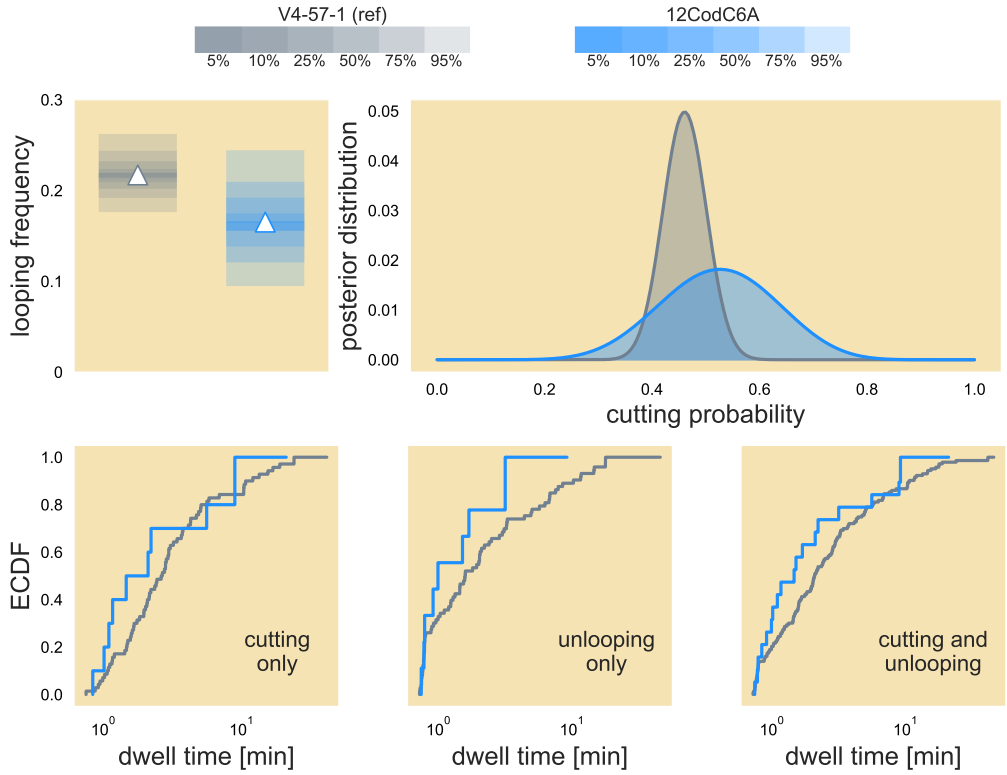


Figure A.8: **V4-57-1 (reference) RSS (grey) and coding flank change (blue) comparison of looping frequency, posterior distribution of the cutting probability and ECDFs of PC lifetimes for PCs that cut, those that unloop, and both.**

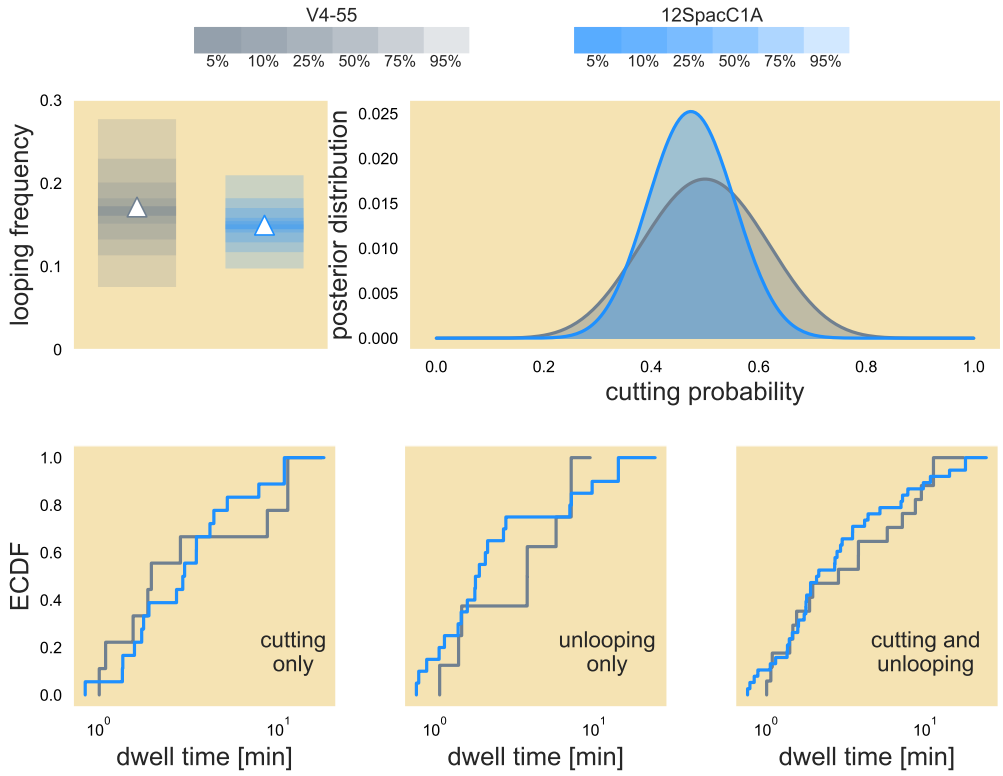


Figure A.9: **V4-55 12RSS (grey) and C-to-A change at spacer position 1 (blue) comparison of looping frequency, posterior distribution of the cutting probability and ECDFs of PC lifetimes for PCs that cut, those that unloop, and both.**

Fig. A.8 shows TPM results on the V4-57-1, or reference, RSS and a single bp change at the nucleotide immediately adjacent to the heptamer, where there is a C-to-A alteration. We find here no distinguishable difference in looping frequency or cleavage probability. Furthermore, we find that the dwell time distributions for PCs that cut, PCs that unloop, and both are identical between the reference and altered coding flank. This finding suggests that at least a single change from C to A near the heptamer does not have a dramatic effect on the RAG-RSS interaction.

We also examined two coding flanks that differ by multiple base pairs. The V4-55 RSS differs from the reference sequence at the first position of the spacer, where the C for the reference is changed to an A for the V4-55 RSS. However, the coding flank sequence differs at five nucleotides. Furthermore, the 6-bp coding flank of V4-55 is composed entirely of Cs and As and removes the Gs and Ts on the reference sequence at the first, third, and fourth positions

of the coding flank (where we index one as six base pairs from the start of the heptamer and six as immediately adjacent). We thus compared the C-to-A change at the spacer position 1 on the reference sequence with the V4-55 coding flank. As Fig. A.9 illustrates that despite the significant difference in sequence between these two constructs at the coding flank, our TPM assay reports little difference that separates these two sequences in looping frequency, dwell time distributions or cutting probability. We thus find that for most of the endogenous RSSs studied, the coding flank has little effect on the overall RAG-RSS interaction. This does not rule out the possibility that Gs or Ts in the first three positions immediately adjacent to the RSS can alter the RAG-RSS dynamics.

A.6 Cloning a different 12RSS in plasmids

To generate the synthetic RSSs used in this work, we used overhang PCR (polymerase chain reaction) and subsequently Gibson assembly (NEB Biolabs) to generate plasmids with the desired change. We selected the endogenous sequence V4-57-1 to serve as the "reference" sequence from which all synthetic RSSs were made. This sequence has been used previously [2] and exhibits a reasonable dwell time distribution, has moderately high looping frequency (compared to the other endogenous sequences), and has close to a 50% cleavage probability, as is shown in this study. This 12RSS sequence is located within the a pZE12 plasmid backbone [15]. The new RSS were inserted into this plasmid via overhang PCR with forward and reverse oligonucleotide primers (IDT) that contain a 15 base-pair overlap with the desired alteration in the middle of the sequence. The primers used in this work are listed in tables A.2 and A.3.

After purification of the PCR fragment and DpnI digestion (NEB Biolabs) of the PCR template, the fragment was circularized using Gibson assembly [16] and transformed into DH5α *Escherichia coli*. Transformants were then cultured and stored for plasmid purification and sequence verification.

A.7 Endogenous RSS sequences

Endogenous 12RSS	Sequence	n_{beads}	n_{loops}	n_{cuts}
DFL16.1-3'	AGCTAC CACAGTGCTATATCCATCAGCAAAAACC	83	37	18
DFL16.1-5'	AATAAA CACAGTAGTAGATCCCTTCACAAAAAGC	263	10	1
V1-135	TCCTCA CACAGTGATTCAGACCCGAACAAAAACT	268	46	14
V9-120	TCCTCC CACAGTGATACAAATCATAACATAAACC	248	41	20
V10-96	TCCTCC CACAATGATATAAGTCATAACATAAACC	286	43	17
V19-93	TCTACC CACAGTGATACAAATCATAACAAAAACC	284	58	26
V4-57-1 (reference)	GTCGAC CACAGTGCTACAGACTGGAACAAAAACC	700	152	70
V4-55	CACCCA CACAGTGATACAGACTGGAACAAAAACC	105	18	9
V5-43	GCCTCA CACAGTGATGCAGACCATAGCAAAAATC	186	27	3
V8-18	TCCCCC CACAGAGCTTCAGCTGCCTACACAAACC	146	5	0
V6-17	TCCTCC CACAGTGCTTCAGCCTCCTACACAAACC	126	34	10
V6-15	TCCTCT CACAGTACTTCAGCCTCCTACATAAACC	201	29	1
Jκ1 23RSS	GGATCC CACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAACC			

Table A.1: **Table of endogenous 12RSS sequences.** The 6 base pairs before the heptamer, known as the coding flank, is changed in the endogenous RSS studies and is included here. The spacer sequence for each RSS is underlined. Bold blue letters in the heptamer and nonamer regions denote deviations from the consensus sequences, CACAGTG and ACAAAAACC, respectively. The number of beads studied n_{beads}, the number of loops formed among the beads n_{loops}, and the number of cut loops n_{cuts} are given for each RSS. The bottom sequence is of the Jκ1 23RSS applied in all of the DNA constructs used in TPM.

A.8 Synthetic and endogenous 12RSS primer sequences

Tables A.2 and A.3 gives the list of primers used to construct the synthetic and endogenous RSSs. For synthetic RSSs, we apply the nomenclature '12' to denote that the 12RSS is altered, the region of the RSS where the change is made ('Hept' = heptamer, 'Non' = nonamer, 'Spac' = spacer, 'Cod' = coding flank), the original nucleotide, the position number in the region, where indexing starts at 1 and finally the new nucleotide. Therefore, if a change is made to the eighth position of the spacer, where a C is altered to T, the RSS is denoted '12SpacC8T'.

Synthetic 12RSS	**Primer**	$n_{\mathbf{beads}}$	$n_{\mathbf{loops}}$	$n_{\mathbf{cuts}}$
12CodC6A (Fwd)	AACACAGTGCTACAGACTGGAACAAAAACCCTGCAGTC	115	19	10
12CodC6A (Rev)	CTGTAGCACTGTGTTCGACCTGCAGCCCAAGCG			
12HeptC3G (Fwd)	ACCAGAGTGCTACAGACTGGAACAAAAACCCTGCAGTC	357	53	1
12HeptC3G (Rev)	CTGTAGCACTCTGGTCGACCTGCAGCCCAAGCG			
12HeptC3T (Fwd)	ACCATAGTGCTACAGACTGGAACAAAAACCCTGCAGTC	212	53	2
12HeptC3T (Rev)	CTGTAGCACTATGGTCGACCTGCAGCCCAAGCG			
12HeptA4T (Fwd)	ACCACTGTGCTACAGACTGGAACAAAAACCCTGCAGTC	607	134	52
12HeptA4T (Rev)	CTGTAGCACAGTGGTCGACCTGCAGCCCAAGCG			
12HeptG5A (Fwd)	ACCACAATGCTACAGACTGGAACAAAAACCCTGCAGTC	96	15	3
12HeptG5A (Rev)	CTGTAGCATTGTGGTCGACCTGCAGCCCAAGCG			
12HeptG5C (Fwd)	ACCACACTGCTACAGACTGGAACAAAAACCCTGCAGTC	821	96	18
12HeptG5C (Rev)	CTGTAGCAGTGTGGTCGACCTGCAGCCCAAGCG			
12HeptT6A (Fwd)	ACCACAGAGCTACAGACTGGAACAAAAACCCTGCAGTC	246	2	1
12HeptT6A (Rev)	CTGTAGCTCTGTGGTCGACCTGCAGCCCAAGCG			

12HeptT6C (Fwd)	ACCACAGCGCTACAGACTGGAACAAAAACCCTGCAGTC	461	24	2
12HeptT6C (Rev)	CTGTAGCGCTGTGGTCGACCTGCAGCCCAAGCG			
12HeptG7A (Fwd)	ACCACAGTACTACAGACTGGAACAAAAACCCTGCAGTC	343	109	28
12HeptG7A (Rev)	CTGTAGTACTGTGGTCGACCTGCAGCCCAAGCG			
12HeptG7C (Fwd)	ACCACAGTCCTACAGACTGGAACAAAAACCCTGCAGTC	146	30	10
12HeptG7C (Rev)	CTGTAGGACTGTGGTCGACCTGCAGCCCAAGCG			
12HeptG7T (Fwd)	ACCACAGTTCTACAGACTGGAACAAAAACCCTGCAGTC	219	47	7
12HeptG7T (Rev)	CTGTAGAACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacC1A (Fwd)	ACCACAGTGATACAGACTGGAACAAAAACCCTGCAGTC	254	38	18
12SpacC1A (Rev)	CTGTATCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacC1G (Fwd)	ACCACAGTGGTACAGACTGGAACAAAAACCCTGCAGTC	117	19	12
12SpacC1G (Rev)	CTGTACCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacA3G (Fwd)	ACCACAGTGCTGCAGACTGGAACAAAAACCCTGCAGTC	134	35	12
12SpacA3G (Rev)	CTGCAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacA3T (Fwd)	ACCACAGTGCTTCAGACTGGAACAAAAACCCTGCAGTC	120	28	18
12SpacA3T (Rev)	CTGAAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacC4G (Fwd)	ACCACAGTGCTAGAGACTGGAACAAAAACCCTGCAGTC	210	38	6
12SpacC4G (Rev)	CTCTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacC4T (Fwd)	ACCACAGTGCTATAGACTGGAACAAAAACCCTGCAGTC	306	128	43
12SpacC4T (Rev)	CTATAGCACTGTGGTCGACCTGCAGCCCAAGCG			

12SpacG6A (Fwd)	ACCACAGTGCTACAAACTGGAACAAAAACCCTGCAGTC	250	74	24
12SpacG6A (Rev)	TTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacG6T (Fwd)	ACCACAGTGCTACATACTGGAACAAAAACCCTGCAGTC	184	78	34
12SpacG6T (Rev)	ATGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacA7C (Fwd)	ACCACAGTGCTACAGCCTGGAACAAAAACCCTGCAGTC	139	37	15
12SpacA7C (Rev)	TTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacA7G (Fwd)	ACCACAGTGCTACAGGCTGGAACAAAAACCCTGCAGTC	168	21	10
12SpacA7G (Rev)	TTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacC8T (Fwd)	ACCACAGTGCTACAGATTGGAACAAAAACCCTGCAGTC	98	17	5
12SpacC8T (Rev)	TTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacT9A (Fwd)	ACCACAGTGCTACAGACAGGAACAAAAACCCTGCAGTC	112	22	12
12SPacT9A (Rev)	TTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacT9C (Fwd)	ACCACAGTGCTACAGACCGGAACAAAAACCCTGCAGTC	117	50	12
12SpacT9C (Rev)	TTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacT9G (Fwd)	ACCACAGTGCTACAGACGGGAACAAAAACCCTGCAGTC	96	8	6
12SpacT9G (Rev)	TTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacG10A (Fwd)	ACCACAGTGCTACAGACTAGAACAAAAACCCTGCAGTC	292	60	29
12SpacG10A (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacG10C (Fwd)	ACCACAGTGCTACAGACTCGAACAAAAACCCTGCAGTC	117	34	18
12SpacG10C (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			

12SpacG10T (Fwd)	ACCACAGTGCTACAGACTTGAACAAAAACCCTGCAGTC	65	20	15
12SpacG10T (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacG11A (Fwd)	ACCACAGTGCTACAGACTGAAACAAAAACCCTGCAGTC	184	29	12
12SpacG11A (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacG11C (Fwd)	ACCACAGTGCTACAGACTGCAACAAAAACCCTGCAGTC	172	26	8
12SpacG11C (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacG11T (Fwd)	ACCACAGTGCTACAGACTGTAACAAAAACCCTGCAGTC	941	267	83
12SpacG11T (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacA12C (Fwd)	ACCACAGTGCTACAGACTGGCACAAAAACCCTGCAGTC	132	15	7
12SpacA12C (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12SpacA12T (Fwd)	ACCACAGTGCTACAGACTGGTACAAAAACCCTGCAGTC	138	24	10
12SpacA12T (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12NonA1G (Fwd)	ACCACAGTGCTACAGACTGGAGCAAAAACCCTGCAGTC	392	108	38
12NonA1G (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12NonA3C (Fwd)	ACCACAGTGCTACAGACTGGAACCAAAACCCTGCAGTC	554	15	7
12NonA3C (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12NonA4C (Fwd)	ACCACAGTGCTACAGACTGGAACACAAACCCTGCAGTC	384	37	10
12NonA4C (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12NonA4T (Fwd)	ACCACAGTGCTACAGACTGGAACATAAACCCTGCAGTC	151	10	5
12NonA4T (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			

12NonA5T (Fwd)	ACCACAGTGCTACAGACTGGAACAATAACCCTGCAGTC	354	58	16
12NonA5T (Rev)	CTGTAGCACTGTGGTCGACCTGCAGCCCAAGCG			
12NonC8G (Fwd)	GCTACAGACTGGAACAAAAAGCCTGCAGTCAACCTCGA	131	24	9
12NonC8G (Rev)	TTTGTTCCAGTCTGTAGCACTGTGGTCGACCTGCAG			
12NonC8T (Fwd)	GCTACAGACTGGAACAAAAATCCTGCAGTCAACCTCGA	280	18	6
12NonC8T (Rev)	TTTGTTCCAGTCTGTAGCACTGTGGTCGACCTGCAG			
12NonC9T (Fwd)	GCTACAGACTGGAACAAAAACTCTGCAGTCAACCTCGA	109	20	11
12NonC9T (Rev)	TTTGTTCCAGTCTGTAGCACTGTGGTCGACCTGCAG			

Table A.2: **Forward (Fwd) and reverse (Rev) primers of synthetic RSSs.** Underlined sequence denotes the region where change is made. Bold-faced letter denotes the new nucleotide. The number of beads studied n_{beads}, the number of loops formed among those beads n_{loops}, and the number of cut loops n_{cuts} are given with the forward primer sequences.

Endogenous 12RSS	Primer
DFL16.1-3' (Fwd)	AGCTACCACAGTGCTATATCCATCAGCAAAAACCCTGCAGTCGAGTAATGCA
DFL16.1-3' (Rev)	GGTTTTTGCTGATGGATATAGCACTGTGGTATTCGAAGCTTGAGCTCG
DFL16.1-5' (Fwd)	AATAAACACAGTAGTAGATCCCTTCACAAAAAGCCTGCAGTCGAGTAATGCA
DFL16.1-5' (Rev)	GCTTTTTGTGAAGGGATCTACTACTGTGGTATTCGAAGCTTGAGCTCG
V1-135 (Fwd)	TCCTCACACAGTGATTCAGACCCGAACAAAAACTCTGCAGTCAACCTCGAGAAACG
V1-135 (Rev)	AGTTTTTGTTCGGGTCTGAATCACTGTGTGAGGACTGCAGCCCAAGCGTGTAG
V9-120 (Fwd)	TCCTCCCACAGTGATACAAATCATAACATAAACCCTGCAGTCAACCTCGAGAAACG
V9-120 (Rev)	GGTTTATGTTATGATTTGTATCACTGTGGGAGGACTGCAGCCCAAGCGTGTAG
V10-96 (Fwd)	TCCTCCCACAATGATATAAGTCATAACATAAACCCTGCAGTCAACCTCGAGAAACG
V10-96 (Rev)	GGTTTATGTTATGACTTATATCATTGTGGGAGGACTGCAGCCCAAGCGTGTAG
V19-93 (Fwd)	TCTACCCACAGTGATACAAATCATAACAAAAACCCTGCAGTCAACCTCGAGAAACG
V10-93 (Rev)	GGTTTTTGTTATGATTTGTATCACTGTGGGTAGACTGCAGCCCAAGCGTGTAG
V4-55 (Fwd)	CACCCACACAGTGATACAGACTGGAACAAAAACCCTGCAGTCAACCTCGAGAAACG
V4-55 (Rev)	GGTTTTTGTTCCAGTCTGTATCACTGTGTGGGTGCTGCAGCCCAAGCGTGTAG
V5-43 (Fwd)	GCCTCACACAGTGATGCAGACCATAGCAAAAATCCTGCAGTCAACCTCGAGAAACG
V5-43 (Rev)	GATTTTTGCTATGGTCTGCATCACTGTGTGAGGCCTGCAGCCCAAGCGTGTAG
V8-18 (Fwd)	TCCCCCCACAGAGCTTCAGCTGCCTACACAAACCCTGCAGTCAACCTCGAGAAACG
V8-18 (Rev)	GGTTTGTGTAGGCAGCTGAAGCTCTGTGGGGGGACTGCAGCCCAAGCGTGTAG
V6-17 (Fwd)	TCCTCCCACAGTGCTTCAGCCTCCTACACAAACCCTGCAGTCAACCTCGAGAAACG
V6-17 (Rev)	GGTTTGTGTAGGAGGCTGAAGCACTGTGGGAGGACTGCAGCCCAAGCGTGTAG
V6-15 (Fwd)	TCCTCTCACAGTACTTCAGCCTCCTACATAAACCCTGCAGTCAACCTCGAGAAACG
V6-15 (Rev)	GGTTTATGTAGGAGGCTGAAGTACTGTGAGAGGACTGCAGCCCAAGCGTGTAG

Table A.3: **Forward (Fwd) and reverse (Rev) primers for designing TPM constructs with endogenous 12RSSs.** Underlined regions denote the heptamer and nonamer regions.

A.9 Protein purification

A.9.1 Murine core RAG1 and core RAG2 co-purification

Maltose-binding protein(MBP)-tagged murine core RAG1 and core RAG2 are co-transfected into HEK293-6E suspension cells using BioT transfection agent and are expressed in the cells for 48 hours. Cells are centrifuged and collected before resuspending with a lysis buffer consisting of cOmplete Ultra protease inhibitor and Tween-20 detergent before lysis through a cell disruptor. Lysate is centrifuged to remove the cell membrane and the supernatant containing expressed RAG is mixed with amylose resin to bind the MBP region to the resin before loading onto a chromatography gravity column. Amylose-attached RAG is then washed using lysis buffer, wash buffer containing salts before eluting with buffer containing high concentrations of maltose to out-compete the MBP on the resin. RAG-contained eluate is then concentrated and dialyzed in buffer containing 25 mM Tris-HCl (pH 8.0), 150 mM KCl, 2 mM DTT and 10% glycerol before snap-freezing 5-10 µL aliquots and storing at -80°C.

A.9.2 HMGB1 purification

Though not discussed extensively in this paper, the high mobility group box 1 (HMGB1) protein binds nonspecifically to DNA and helps facilitate RAG binding onto the RSS. A plasmid containing a His-tagged HMGB1 gene is transformed into BL21(DE3) cells and grown in liquid cultures until they reach an OD600 of 0.7. Cultures are then induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG) to express HMGB1 for 4 hours at 30°C before cells are collected from the media. Cells are resuspended in binding buffer media containing cOmplete Ultra protease inhibitor, benzonase, Tween-20 and a low imidazole concentration before lysis through the cell disruptor. Lysate is cleared of membrane with an ultracentrifuge and loaded onto a nickel-NTA column to bind HMGB1. Nickel-bound HMGB1 is then washed with more binding buffer before washing with buffer containing low imidazole concentration. Washed HMGB1 are then eluted through the column with elution buffer containing higher concentration imidazole. Degraded HMGB1 is then removed by loading HMGB1 eluate onto SP column and collecting flow-through in 200 µL aliquots with an incrementally increasing salt gradient on the AKTA. Fractions that show highest change in voltage reading on the AKTA are run on a Western blot to confirm that protein of the correct size is collected before collecting. HMGB1 are transferred to a dialysis buffer containing 25 mM

Tris-HCl (pH 8.0), 150 mM KCl, 2 mM DTT and 10% glycerol through a buffer-exchange centrifuge column before snap-freezing 5-10 µL aliquots and freezing at -80°C.

BIBLIOGRAPHY

[1] L. Han et al., *Calibration of tethered particle motion experiments*, 1st ed., vol. 150, New York: Springer-Verlag, 2008, 123.

[2] G. A. Lovely et al., "Single-molecule analysis of RAG-mediated V(D)J DNA cleavage", Proc Natl Acad Sci 112 (2015), E1715.

[3] L. Han et al., "Concentration and length dependence of DNA looping in transcriptional regulation", PLoS ONE 4 (2009), e5621.

[4] S. Johnson, M. Linden, and R. Phillips, "Sequence dependence of transcription factor-mediated DNA looping", Nucleic Acids Res 40 (2012), 7728.

[5] G. Chure et al., "Connecting the dots between mechanosensitive channel abundance, osmotic shock, and survival at single-cell resolution", J Bacteriol 200 (2018), e00460.

[6] G. D. Ruxton, "The Unequal Variance T-Test Is an Underused Alternative to Student's t-Test and the Mann–Whitney U Test", en, Behav Ecol 17 (2006), 688.

[7] A. Agrawal, Q. M. Eastman, and D. G. Schatz, "Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system", Nature 394 (1998), 744.

[8] B. Carpenter et al., "Stan: A probabilistic programming language", J Stat Softw 76 (2017), 1.

[9] R. M. Gerstein and M. R. Lieber, "Coding end sequence can markedly affect the initiation of V(D)J", Genes Dev 7 (1993), 1459.

[10] U. R. Ezekiel et al., "The composition of coding joints formed in V(D)J recombination is strongly affected by the nucleotide sequence of the coding ends and their relationship to the recombination signal sequences", Mol Cell Biol 17 (1997), 4191.

[11] K. Yu and M. R. Lieber, "Mechanistic basis for coding end sequence effects in the initiation of V(D)J recombination", Mol Cell Biol 19 (1999), 8094.

[12] H. Ru et al., "Molecular mechanism of V(D)J recombination from synaptic RAG1-RAG2 complex structures", Cell 163 (2015), 1138.

[13] M.-S. Kim et al., "Cracking the DNA code for V(D)J recombination", Mol Cell 70 (2018), 1.

[14] H. Ru et al., "DNA melting initiates the RAG catalytic pathway", Nat Struct Mol Biol 25 (2018), 732.

[15] R. Lutz and H. Bujard, "Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements", Nucleic Acids Res 25 (1997), 1203.

[16] D. G. Gibson et al., "Enzymatic assembly of DNA molecules up to several hundred kilobases", Nat Meth 6 (2009), 343.

Appendix B

SUPPLEMENTAL INFORMATION FOR CHAPTER 3: GRID FRAP PATTERNING REVEALS A DISPERSIVE EFFECT IN THE BULK OF A LINEARLY CONTRACTING MICROTUBULE NETWORK

B.1 Materials and methods

B.1.1 Motor purification

Plasmids containing the gene encoding the motor-fluorescent protein-light-activated dimerization-FLAG tag construct with the pBiex-1 vector are transfected in Sf9 suspension cells for 60-72 hours at 27°C on shakers rotating at 120 rpm. Cells are then lightly centrifuged at 500 rpm for 12 minutes to remove the supernatant before resuspending in lysis buffer (100 mM NaCl, 2 mM $MgCl_2$, 0.25 mM EDTA, 0.5 mM EGTA, 0.25 % Igepal, 3.5% sucrose by weight, 10 mM imidazole pH 7.5, 10 µg/mL aprotinin, 10 µg/mL leupeptin, 1 mM ATP, 2.5 mM DTT, and 0.5 mM PMSF) and leaving on ice for 20 minutes. Cells are then spun down for 30 minutes at 154k $\times g$ after which the lysate is transferred to tubes containing mouse monoclonal anti-FLAG resin (Sigma A2220) and slowly rotated at 4°C for 1.5$\backsim$3 hrs to allow protein binding to the resin via the FLAG tag. Resin-bound protein are washed three times by spinning down at 2000$\times$ g, clearing the supernatant, then resuspending by tube inversion in wash buffer containing 15 mM KCl, 0.5 mM, 0.1 mM EGTA, 0.1 mM EDTA, 2 mM imidazole pH 7.5, 10 µg/mL aprotinin, 10 µg/mL leupeptin, 0.3 mM DTT, and ATP in 3 mM, 0.3 mM, and 0.03 mM concentrations for the first, second, and third washes, respectively. After the third wash, the protein are spun down again at 2000$\times$ g and most of the supernatant is removed, leaving the resin bed and roughly an equivalent amount of supernatant by volume in the tube. The resin bed is resuspended and FLAG peptide (Sigma F4799 or Thermo Scientific A36805) is added at a final concentration of 0.5 mg/mL before rotating for 3 hrs at 4°C. After incubation to allow the peptide to outcompete the protein for resin binding, the protein are spun down again at 2000$\times$ g with the supernatant extracted and further spun down using centrifuge columns with $\backsim$30 µm pore sizes to further separate proteins from any collected resin beads. Flow-through of clarified protein are spin concentrated

using a 50 kDa filter tube to a final concentration of 2-2.5 mg/mL before diluting in 100% glycerol of the same volume for storage.

B.1.2 Stabilized microtubule polymerization

Fluorescently labeled stabilized microtubules are prepared as in [1, 2]. After flash thawing at 37°C and kept on ice, a combination of ≈ 1.5 mg unlabeled and 100 µg labeled tubulin are diluted to 7.5 mg/mL and 0.5 mg/mL, respectively, in M2B 6.8 containing DTT and GMP-CPP at final concentrations of 1 mM and 6mM, respectively. The tubulin mixture is then incubated on ice for 5 minutes in an ultracentrifuge tube before ultracentrifugation at 90,000 rpm at 4°C for 8 minutes. Avoiding the pellet at the the bottom, the supernatant containing tubulin monomers are then placed in a new Eppendorf tube and incubated at 37°C for 1 hour, typically in a water bath, during which the tubulin is polymerizing and stabilizing with GMPCPP. The microtubule mixture is then aliquoted into individual PCR tubes while constantly being suspended in the mixture by stirring with a pipette tip. PCR tubes are then briefly spun down with a tabletop minicentrifuge before flash-freezing with liquid nitrogen and placed in a -80°C freezer for long-term storage. Microtubules are then prepared for experiments by immersing the PCR tube in 37°C water immediately when taken out of the freezer to quickly thaw.

B.1.3 Glass slide treatment

Corning glass slides and No. 1.5 Deckgläser coverslips are coated with an acrylamide solution to prevent the adhesion of proteins from the light-dimerized activation assay to the surface. The acrylamide coating is done similarly to that demonstrated in [3]. Prior to application of the solution, slides and coverslips are separated by placement in appropriately sized containers and rigorously cleaned through a series of solutions and sonicating. First, slides are immersed in 1% Hellmanex to remove dirt particulates, sonicated, repeatedly rinsed with deionized water (DI H_2O), then repeatedly rinsed with ethanol. Slides are then sonicated in 200 proof ethanol before rinsing again with DI H_2O. After rinsing, slides are sonicated in 0.1 M KOH and subsequently rinsed in double-distilled water (ddH_2O). Finally, trace metals are removed by immersing in 5% HCl for 4 hours. After repeatedly rinsing in ddH_2O, slides are stored overnight with MilliQ ultrapure water.

Upon cleaning and before the acrylamide coating, a silane solution is made first by mixing 98.5% 200 proof ethanol and 1% acetic acid before adding 0.5% trimethoxysilyl methacrylate and immediately pouring into the containers holding the slides and coverslips. After roughly 30 minutes, slides are rinsed twice in 200 proof ethanol before drying with N_2 air and baking at 110°C for 10-15 minutes to cure silane onto surface with oxygen bonding.

The polyacrylamide solution is made by mixing 950 mL ddH_2O with 50 mL 40% acrylamide and degassing under vacuum for 30 minutes. The solution is then under constant mixing on a stir plate with a stir bar during which time 350 µL TEMED and 700 mg ammonium persulfate (APS) are added to the solution. The acrylamide solution is immediately added to the slides and coverslips and incubated overnight. Slides are placed in 4°C for long-term storage.

B.1.4 Flow cell chamber preparation

Flow cells for all light-dimerized activation assays are prepared by thoroughly rinsing an acrylamide-coated glass slide and coverslip in ddH_2O and air drying with N_2 gas. A piece of parafilm with three channels each cut 3 mm wide is placed on the glass slide with the long axis of the channels running along the length of the slide. The coverslip is placed on top of the parafilm with pressure applied to flatten out the film. The flow cell is then briefly placed on a hot plate set at 65°C to warm the parafilm, allowing extra pressure on the contact points between the film and the glass to better seal the chambers.

B.1.5 Light-dimerized activation assay preparation

Photobleaching experiments require an energy mix to maintain stability and function of microtubules and motors while constantly supplying kinesin motors with ATP to contract the microtubule network. This energy mix is slightly altered from that used by Ross *et al.* [1] with the major changes being a change in acidity for K-PIPES from pH 6.8 to pH 6.1 and the absence of catalase to allow for photobleaching. iLid- and micro-tagged motors with the same fluorescent protein are each added to the reaction mixture at final concentrations of 40-100 nM with stabilized microtubules added at a final concentration of 1.5-2.5 µM tubulin. Concentrations of motors and tubulin are tuned to ensure that microtubule network contracts into an aster without an influx of microtubules from outside of the light-activation region.

B.1.6 Optical set-up

The sample is imaged and photobleached using a super planar fluorescence 20x objective from Nikon (numerical aperture 0.45). Image acquisition is performed using a FLiR Blackfly monochrome camera (BFLY-U3-23S6M-C) with two filters in front of it: a Semrock Brightline dual-band pass filter centered at 577 nm (28.3 nm FWHM bandwidth) and 690 nm (55.1 nm FWHM bandwidth) and a Semrock StopLine single-notch filter at 532 nm (17 nm notch bandwidth) to suppress transmission of the YFP YFP excitation to the camera.

Fig. B.1 gives a general idea of the layout of the microscopy components. Activation of motor dimerization and imaging of the microtubules is per-

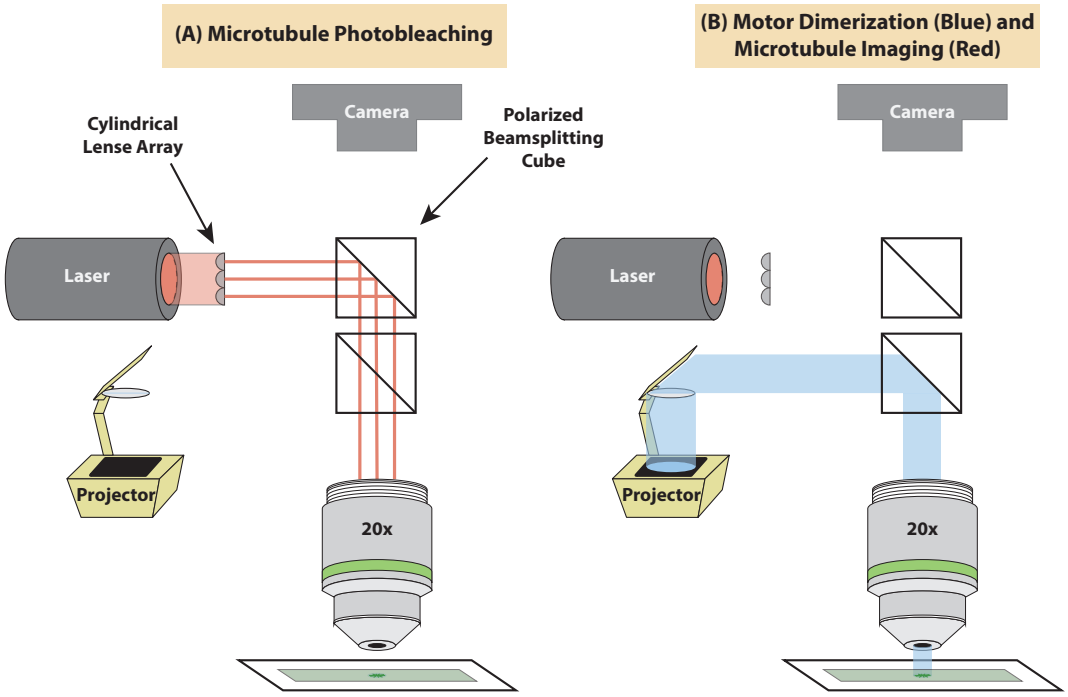


Figure B.1: **Arrangement of the laser and projector.** The laser and projector are set on different optical paths before reaching the sample. (A) The projector shines white light that passes through a filter in order to clip to the desired wavelength. These filters will either transmit blue light to perform the iLid-micro motor dimerization or red light to image the microtubule fluorescence channel. (B) The laser performs the photobleaching of the microtubules in a grid pattern by passing through a cylindrical lens array. The cylindrical lens array is mounted on a rotation mount (not shown) to bleach vertical and horizontal lines. A 20x Nikon objective is used for the imaging.

formed using a digital light projector DLP Lightcrafter Display 4710 EVM Gen2 from Texas Instruments. The DLP projects white light while a motorized filter wheel sets the transmissible range of wavelengths onto the sample (beam blocker for no light, 460/50 nm filter for blue light for iLid-micro dimerization and 630/38 for microtubule imaging). Photobleaching of microtubules is performed using a 645 nm laser. The laser path is set to pass through a cylindrical lens array that transforms the collimated light pattern into a series of lines along one axis. The cylindrical lens array is mounted onto a rotation mount to allow for photobleaching of vertical and horizontal lines to generate the grid pattern. To ensure that the photobleached lines persist for multiple frames of the image, the laser passes through a gimbal-mounted mirror that deflects the beam over a small range of angles. By deflecting the laser light off of the mirror through two lenses with the same focal length f and a second, stationary mirror placed $4 \times f$ away from the gimbal-mounted mirror before passing the laser through the cylindrical lense array, the transformed laser lines can be swept out. We use this beam steering approach to photobleach thicker lines.

To perform the activation and imaging patterns, we supply µManager with a TIFF stacks of matching pixel dimensions as the projector and use a Beanshell script modified from Ross *et al.* to use the correct TIFF image in the stack. The TIFF stack contains a blank image (all pixel values 0) for when the laser is turned on (which is also used in conjunction with the beam blocker to prevent light from passing onto the sample outside of the activation and imaging cycles); a maximum pixel intensity image for the microtubule imaging, and a circular pattern in a blank background for the circular iLid-micro dimerization activation pattern. The primary modification to the Beanshell script is the incorporation of a timer for when the photobleaching will be performed. Once the experiment reaches the desired time, the imaging pauses while the Beanshell script turns on the laser and executes a series of custom written executables that sweep out the laser lines to create thicker parallel photobleached lines, turn off the laser, rotate the cylindrical lense array, then reactivate the laser and sweep out the laser lines in the orthogonal direction to generate the grid pattern. Upon finishing this command, the laser is shut off and imaging resumes. The entire photobleaching is performed within a roughly 10-15 second window.

B.2 Unit cell segmentation and fluorescence preservation in subsequent frames

Fluorescent unit cells of a photobleached microtubule network are segmented in the cropped image sets where the microtubules outside of the activation region are neglected. We first reduce the background signal in each image by performing a heavy Gaussian blur ($\sigma = 20$ pixels) and subtracting off the Gaussian blur from the original image. Images are then normalized to fall between 0 and 1. In order to identify each fluorescent square, we use the triangle thresholding algorithm [4] as it accurately segmented the unit cells in the first image taken after the photobleaching was performed. Other thresholding methods either segmented unit cells to be much smaller and therefore misses a large amount of fluorescent regions of the unit cell or segmented unit cells to be much larger, which affects the amount of time that unit cells are identified as distinct. After the thresholding is applied, the segmented image is cleaned up by removing segmented objects that are too small (less than a third of the area of a unit cell immediately after photobleaching), objects that are too large (more than 3 times larger than an expected unit cell immediately after photobleaching) or images that are too close to the border, which typically removes microtubules outside of the iLid-micro light-dimerized region. To close off any patches within a fluorescent unit cell due to the thresholding, we perform a morphological closing is performed. With the segmented images, the centroid position, area, and total fluorescence of each unit cell are obtained as well as the pixel-weighted centroid of the entire segmented image to obtain the microtubule network center.

Subsequent images of the same dataset undergo the same background subtraction to segmented image clean-up. However, as the some fraction of the fluorescent microtubules begin to disperse, the image segmentation may not pick up fewer of the fluorescent microtubules at the boundary of the unit cell with the photobleached region as they may be considered too low in signal to be distinguished from the background. As a result, for later images than the first image after photobleaching, we correct the segmentation by adding on pixels around the boundary of the segmented unit cells until we return to the correct total fluorescence. To do this, each unit cell is then paired with itself from the previous time step by determining nearest centroids. Due to the minimal reduction in fluorescence intensity from the DLP during imaging as shown in Section B.3, we compare the total fluorescence intensity of the unit

cell in the frame of interest post-segmentation to its total intensity from the first frame. If the total intensity is less than 99% of the initial intensity, we continually add a single-pixel thick layer around the unit cell until the unit cell finally falls within 99% of the initial intensity. If after an iteration the total intensity becomes greater than the initial intensity, we remove the dimmest pixel around the outer layer until the intensities roughly match. Unit cell centroids, areas, and fluorescence intensities are then computed in addition to the pixel-weighted center of the entire contracting network after this intensity-adjusted processing for all of the unit cells. Image processing of a unit cell terminates when it is found to overlap with another unit cell during the fluorescence intensity correction scheme as this indicates that the unit cells have begun to merge and by the next time point will no longer be distinguished.

B.3 Projector effects on microtubule fluorescence intensity

In analyzing the photobleached microtubule field as the network contracts, we used the total fluorescence intensity of the unit cells as a conserved quantity during the unit cell segmentation. One concern might be whether the microtubule fluorescence decreases in time due to the effects of the projector, which illuminates the field of view for imaging purposes. To investigate this, we imaged the microtubule field without activating the iLid-micro dimerization using the same exposure times ($\backsim$100 ms) and imaging frequency (10 s per frame). We then examined the mean image intensity and standard deviation of the pixel intensity as a function of time.

SI Fig. B.2(A) illustrates the effects of the projector on the microtubule field. The mean intensity of the field of view, as normalized against the mean intensity at $t = 0$ seconds, indicates that the fluorescence field fluctuates only a few tenths of a percent but does not appear to decrease over an hour. These fluctuations are likely due to the diffusion of the microtubules in the flow cell, as SI Fig. B.2(B) shows the normalized mean intensity of the microtubule fluorescence channel but in the absence of microtubules. Here, we see that that there are fewer fluctuations in the fluorescence intensity, further supporting that the small fluctuations in fluorescence intensity in successive imaging stages comes from diffusion of the microtubules. Nevertheless, we show here that the fluorescence intensity is well preserved and use this as our justification for using total fluorescence intensity as the conserved metric for unit cell segmentation.

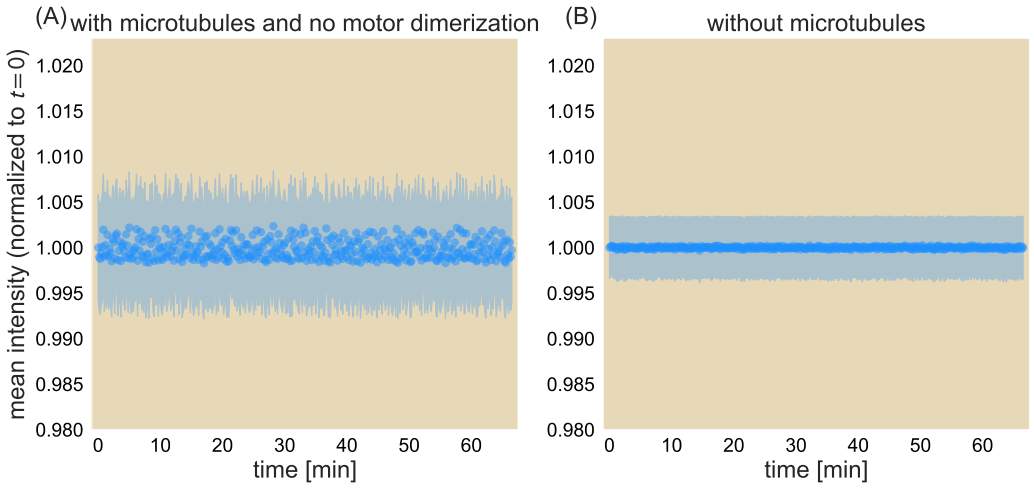


Figure B.2: **Image intensity of the microtubule field as a function of time.** (A) Mean intensity of the microtubule field normalized against that of the first image. Blue shaded region represents one standard deviation in the mean intensity (normalized by the same initial mean value). (B) Mean intensity of the same fluorescence channel in the absence of microtubules. Blue shaded region once again represents the standard deviation of the image region.

B.4 Data analysis

B.4.1 Contraction rate computation

In the main text, we use the centroids of fluorescent unit cells obtained as outlined in Section B.2 of Appendix B to demonstrate that contraction speed of the microtubule network scales linearly with distance from the network center. We first obtain the speed that each unit cell centroid is moving toward the center as a function of time. For each unit cell, we observe a linear relation between the centroid distance from the network center and time after photobleaching of the form

$$r = v_c\, t + r_0, \tag{B.1}$$

where r is the unit cell centroid distance from the network center, v_c is the speed of the unit cell (which will take to be positive here but directed toward the origin), t is the time since photobleaching, and r_0 is the initial centroid distance from the network center immediately after photobleaching.

Using the extracted contraction speed and distances for all of the unit cells for a given motor type, we next computed the rate of contraction of the microtubule network. We note that we expect a linear relation between radius r and centroid speed v_c of the form

$$v_c = \alpha\, r + v_0, \tag{B.2}$$

where α is the contraction rate and v_0 is the contraction speed at the network center. Although we expect the speed at the network center to be 0, we relax this assumption for our analysis. To more carefully compute the rate of contraction of the network and determine the range of credibility of the computed rate, we use a Bayesian approach. Specifically, we compute the probability of α and v_0 given our data on the contraction speeds for each unit cell and their distance from the network center, $P\left[\alpha, v_0 | \left\{(r_0, v_c)_i\right\}\right]$, where i denotes each unit cell. Here, we use the centroid distance immediately after photobleaching but found that another criterion such as the median of the centroid distance over the course of the time window analyzed does not dramatically affect the results due to the relatively small travel $\left(\frac{\Delta r}{r_0} < 10\% \text{ for } \Delta r \text{ the distance traveled over the entire time course}\right)$ the unit cells undergo.

We note from Bayes' Theorem that

$$\begin{aligned} P\left[\alpha, v_0 | \left\{(r_0, v_c)_i\right\}\right] &= \frac{P\left[\left\{(r_0, v_c)_i\right\} | \alpha, v_0\right] P\left(\alpha, v_0\right)}{P\left[\left\{(r_0, v_c)_i\right\}\right]}, \\ &= \frac{\prod_i P\left[(r_0, v_c)_i | \alpha, v_0\right]}{\prod_i P\left[(r_0, v_c)_i\right]} P\left(\alpha, v_0\right), \\ &\propto \prod_i P\left[(r_0, v_c)_i | \alpha, v_0\right] P\left(\alpha, v_0\right), \end{aligned} \tag{B.3}$$

where we drop the denominator on the right-hand side as it does not involve the parameters we want to find, thus making the two sides proportional to each other. Here, $P\left[(r_0, v_c)_i | \alpha, v_0\right]$ is the likelihood distribution of getting the $(r_0, v_c)_i$ that we did given α and v_0 while $P\left(\alpha, v_0\right)$ is the prior distribution of our two parameters.

We expect that our priors on α and v_0 are independent of each other, so we can break up the probability function into a product of two:

$$P\left(\alpha, v_0\right) = P\left(\alpha\right) P\left(v_0\right). \tag{B.4}$$

Meanwhile, we can rearrange each likelihood function as a product of two probabilities. The probability of getting $(r_0, v_c)_i$ given our parameters is also the probability of getting $v_{c,i}$ given our parameters and $r_{0,i}$ times the probability of getting $r_{0,i}$, or

$$\begin{aligned} P\left[(r_0, v_c)_i | \alpha, v_0\right] &= P\left(v_{c,i} | \alpha, v_0, r_{0,i}\right) P\left(r_{0,i}\right), \\ &\propto P\left(v_{c,i} | \alpha, v_0, r_{0,i}\right), \end{aligned} \tag{B.5}$$

where we change to a proportionality again as $P(r_{0,i})$ is independent of our parameters. Here, we expect that our contraction speed for a given unit cell $v_{c,i}$ comes from a Normal distribution where the mean value is $\alpha\, r_{0,i} + v_0$ and standard deviation σ. This means that we will also need a prior on σ. This means that our distribution really takes the form of

$$P\left[\alpha, v_0, \sigma | \left\{(r_0, v_c)_i\right\}\right] \propto P(\alpha)\, P(v_0)\, P(\sigma) \prod_i P\left(v_{c,i} | \alpha, v_0, \sigma, r_{0,i}\right). \tag{B.6}$$

As a result, we say that our likelihood takes the form

$$v_{c,i} \backsim \text{Normal}\left(\alpha r_{0,i} + v_0, \sigma^2\right). \tag{B.7}$$

We then defined our priors to be that α is drawn from the half-normal distribution where $\alpha > 0$ as we are working with speeds of contraction, σ is also drawn from a half-normal distribution and enforced to be positive, and v_0 is drawn from a normal distribution about $v = 0$. We make the offset a normal rather than a half-normal distribution as there may be a value of $r > 0$ for which the contraction stops, which for a positive slope would mean a negative speed at $r = 0$. Put together, we have the following priors:

$$\alpha \backsim \text{Half-Normal}(0, 1), \tag{B.8}$$

$$\sigma \backsim \text{Half-Normal}(0, 1), \tag{B.9}$$

$$v_0 \backsim \text{Normal}(0, 1). \tag{B.10}$$

We sampled the joint distribution of (α, v_0, σ) by Hamiltonian Markov chain Monte Carlo using the Stan probabilistic program [5]. From each (α, v_0) that is sampled we compute the mean value $\mu = \alpha\, r + v_0$ for $0 \leq r \leq R$ where R is the distance of the farthest centroid from the network center and report the median and 95% credible region for at each distance r as presented in Fig. 2 of the main text and Fig. B.6.

B.5 Deformation of a square due solely to contraction

In the main text, we observed that each fluorescent unit cell on average conserves its area while its center of mass moves toward the network center with speed that is linearly dependent on the distance from the center. We compute the expected area of each unit cell had the network elastically contracted due solely to the observed global contraction. We define the contraction velocity field $\mathbf{v}(x, y)$ as

$$\mathbf{v}(x, y) \equiv -\alpha\left(x\hat{x} + y\hat{y}\right), \tag{B.11}$$

where α is the contraction rate as computed in SI Sec. B.4.1 and reported in the main manuscript. This means that after a time interval Δt a point (x, y) subject to this advective flow will be displaced in the x- and y- directions according to

$$\begin{aligned} dX &= -\alpha x \Delta t, \\ dY &= -\alpha y \Delta t, \end{aligned} \tag{B.12}$$

so the point at the later time (x', y') relates to its earlier time point by

$$\begin{aligned} x' &= x + dX = x\left(1 - \alpha \Delta t\right) \\ y' &= y + dY = y\left(1 - \alpha \Delta t\right). \end{aligned} \tag{B.13}$$

Suppose we looked at the four corners of a unit cell, labeled as A, B, C, D as depicted in Fig. B.3. If we assign their coordinates as

$$\begin{aligned} \mathrm{A} &\to \left(x_{\mathrm{A}}, y_{\mathrm{A}}\right), \\ \mathrm{B} &\to \left(x_{\mathrm{B}}, y_{\mathrm{B}}\right), \\ \mathrm{C} &\to \left(x_{\mathrm{C}}, y_{\mathrm{C}}\right), \\ \mathrm{D} &\to \left(x_{\mathrm{D}}, y_{\mathrm{D}}\right), \end{aligned} \tag{B.14}$$

we see that by picking a square, we can simplify any two diagonal points to be dependent on coordinate values from the other two diagonal points, so with a choice of using coordinates from A and D, the coordinates become

$$\begin{aligned} \mathrm{A} &\to \left(x_{\mathrm{A}}, y_{\mathrm{A}}\right), \\ \mathrm{B} &\to \left(x_{\mathrm{D}}, y_{\mathrm{A}}\right), \\ \mathrm{C} &\to \left(x_{\mathrm{A}}, y_{\mathrm{D}}\right), \\ \mathrm{D} &\to \left(x_{\mathrm{D}}, y_{\mathrm{D}}\right). \end{aligned} \tag{B.15}$$

Under the deformation mapping, their new coordinates, labeled as A′, B′, C′, and D′ get mapped on as

$$\begin{aligned} \mathrm{A}' &\to \left[x_{\mathrm{A}}\left(1 - \alpha \Delta t\right), y_{\mathrm{A}}\left(1 - \alpha \Delta t\right)\right], \\ \mathrm{B}' &\to \left[x_{\mathrm{D}}\left(1 - \alpha \Delta t\right), y_{\mathrm{A}}\left(1 - \alpha \Delta t\right)\right], \\ \mathrm{C}' &\to \left[x_{\mathrm{A}}\left(1 - \alpha \Delta t\right), y_{\mathrm{D}}\left(1 - \alpha \Delta t\right)\right], \\ \mathrm{D}' &\to \left[x_{\mathrm{D}}\left(1 - \alpha \Delta t\right), y_{\mathrm{D}}\left(1 - \alpha \Delta t\right)\right]. \end{aligned} \tag{B.16}$$

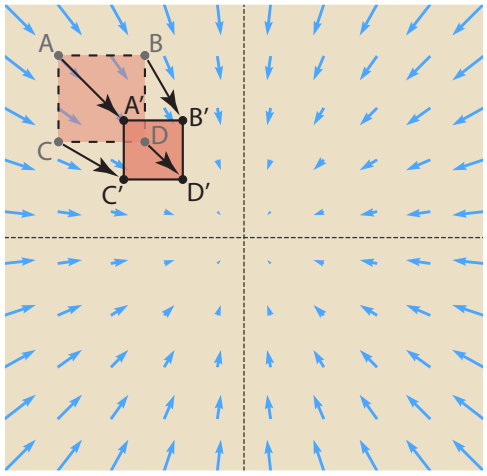


Figure B.3: **Schematic of unit cell contraction due purely to the advective velocity field.** An advective velocity field scales linearly with distance from the origin while pointing radially inward and are shown in blue. The points at the corners of the square (A, B, C, D) are mapped after some time Δt to (A′, B′, C′, D′).

Eqs. B.16 tells us that under this particular velocity field, any two points that are horizontally or vertically aligned will maintain the same horizontal or vertical alignment, respectively, even at later times. Thus, a square will preserve its shape in time.

We next examine what happens to the area of a unit cell had the only effect been the global contraction. In this case, we can compare the area of the square before and after the deformation. To compute the area swept out by (A,B,C,D), we multiply the line segment between B and D, L_{BD} with the line segment between C and D, L_{CD}:

$$\begin{aligned}\sigma_{(\mathrm{A,B,C,D})} &= L_{\mathrm{BD}} \times L_{\mathrm{CD}}, \\ &= \left[\sqrt{(x_{\mathrm{B}} - x_{\mathrm{D}})^2 + (y_{\mathrm{B}} - y_{\mathrm{D}})^2}\right] \times \left[\sqrt{(x_{\mathrm{D}} - x_{\mathrm{C}})^2 + (y_{\mathrm{D}} - y_{\mathrm{C}})^2}\right], \\ &= (y_{\mathrm{A}} - y_{\mathrm{D}}) \times (x_{\mathrm{D}} - x_{\mathrm{A}}), \end{aligned} \tag{B.17}$$

where we use Eq. B.15 to write in terms of the coordinates of A and D. In comparison, the area of the deformed unit cell swept out by (A′, B′, C′, D′)

takes the form

$$\begin{aligned}
\sigma_{(\mathrm{A}',\mathrm{B}',\mathrm{C}',\mathrm{D}')} &= L_{\mathrm{B}'\mathrm{D}} \times L_{\mathrm{C}'\mathrm{D}'}, \\
&= \left[\sqrt{\left(x_{\mathrm{B}'} - x_{\mathrm{D}'}\right)^2 + \left(y_{\mathrm{B}'} - y_{\mathrm{D}'}\right)^2}\right] \times \left[\sqrt{\left(x_{\mathrm{D}'} - x_{\mathrm{C}'}\right)^2 + \left(y_{\mathrm{D}'} - y_{\mathrm{C}'}\right)^2}\right], \\
&= \left(y_{\mathrm{A}'} - y_{\mathrm{D}'}\right) \times \left(x_{\mathrm{D}'} - x_{\mathrm{A}'}\right), \\
&= \left[y_{\mathrm{A}}\left(1 - \alpha\Delta t\right) - y_{\mathrm{D}}\left(1 - \alpha\Delta t\right)\right] \times \left[x_{\mathrm{D}}\left(1 - \alpha\Delta t\right) - x_{\mathrm{A}}\left(1 - \alpha\Delta t\right)\right], \\
&= \left(y_{\mathrm{A}} - y_{\mathrm{D}}\right)\left(1 - \alpha\Delta t\right) \times \left(x_{\mathrm{D}} - x_{\mathrm{A}}\right)\left(1 - \alpha\Delta t\right), \\
&= \left(y_{\mathrm{A}} - y_{\mathrm{D}}\right) \times \left(x_{\mathrm{D}} - x_{\mathrm{A}}\right)\left(1 - \alpha\Delta t\right)^2, \\
&= \sigma_{(\mathrm{A},\mathrm{B},\mathrm{C},\mathrm{D})}\left(1 - \alpha\Delta t\right)^2. \qquad \text{(B.18)}
\end{aligned}$$

Thus we find that the area of the unit cell subject solely to the contraction would decrease by $(1 - \alpha\Delta t)^2$ after a time period Δt. This comes in contrast to the results that we present here where the area of the fluorescent unit squares remains constant during the contraction process suggesting a mechanism that disperses microtubules against the global contraction.

B.6 Microtubule length extraction

Stabilized microtubules imaged under total internal reflection fluorescence (TIRF) microscopy such as the ones shown in Fig. B.4A were analyzed similar to that discussed in [1] in order to extract their lengths. Briefly, due to the even illumination that can occur in the image, images were first background corrected using a local thresholding method known as Niblack thresholding [6] with window size of 3 pixels and k value of 0.001, which determines how many standard deviations below the mean pixel value that one sets the cut-off within the window. Although the array is a series of pixel values to be weighed against the original image, we found that this array already improved the image contrast. Due to better flattening of the image but a nonbinary image, we used Otsu thresholding on the Niblack theshold array to extract the microtubules from the background. The result is shown in Fig. B.4B.

Using the binary image which contains extracted microtubules, we imposed a morphological closing algorithm to reconnect any microtubules that were broken during the Niblack thresholding from being picked up as signal. This closing was performed using a 3 pixel x 3 pixel square array, suggesting that disconnected microtubules needed to be within $3\sqrt{2}$ pixels of each other at their ends to be connected again. From here, we removed any microtubules

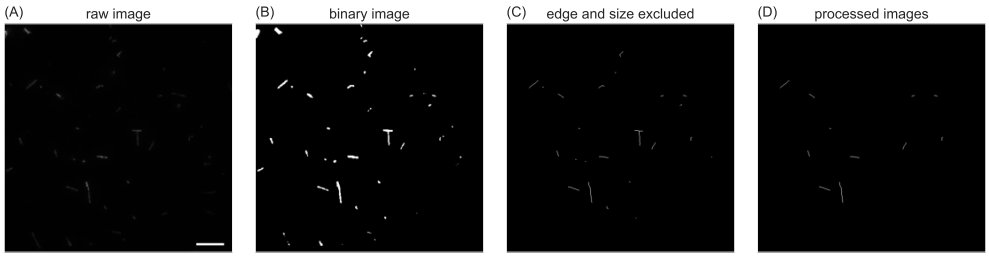


Figure B.4: **Processing steps of microtubule images.** (A) Raw image. Scale bar denotes 10 µm. (B) Images processed after computing a Niblack threshold and using Otsu thresholding on the Niblack threshold array. (C) Putative MTs skeletonized after removing objects too close to the image border or too small. (D) Removal of any MTs that cross over each other to get the final MTs used for analysis.

that were too close to the edge of the image as they may extend outside of the camera field of view, any objects that were fewer than 10 pixels in area as we considered them too small to know with enough certainty whether they were microtubules or small blemishes in the image. Putative microtubules underwent a morphological thinning so that they were converted to one-pixel wide lines along which we could compute their lengths. The result of the edge and size exclusion and skeletonizing are shown in Fig. B.4C.

As a final step before measuring the lengths, we removed any microtubules that seemed to cross over. This was performed by removing objects where two line segments along the same microtubule strand formed angles of at least 75°, leaving behind a processed image such as Fig. B.4D. From here, we used any remaining microtubules and measured their lengths and compiled them. Fig. B.5 shows empirical cumulative distribution functions of these microtubules from the four MT polymerization assays performed over the course of the work presented here. n denotes the number of microtubules that were extracted from the image processing and used in the ECDF for each replicate. Here, we see that for most of the work performed the MTs had lengths between $1 - 3$ µm with median lengths between $1.5 - 2$ µm.

B.7 Motor constructs

While several of the motors used here in the analysis are obtained from previous work, including K401 expressed in bacteria [1], K401 expressed in insects and Ncd236 expressed in insects [7], we also designed constructs for the study of Ncd281 [8]. Specifically, the sequences are inserted into pBiex-1 vectors and

Motor Construct	Sequence Layout
micro variant	pBiex-1:FLAG-GG-mVenus-$(GSG)_2$-micro-$(GSG)_4$-Ncd281
iLid variant	pBiex-1:FLAG-GG-mVenus-$(GSG)_2$-iLid-$(GSG)_4$-Ncd281

Table B.1: **Ncd281 construct design.** All constructs are designed in the pBiex-1 vector and produced by Twist Biosciences.

includes a FLAG tag for protein purification, mVenus for motor fluorescence visualization, either a micro or iLid domain as described in [9] and Ncd281 as described in [8]. Between these different domains are multiple repeats of a 'GSG' amino acid sequence which offers flexible links between the regions. Table B.1 illustrates these sequences. Constructs were produced by Twist Biosciences.

B.8 Measuring motor speeds and their effects on contraction rate and unit cell area

In the work shown in the main manuscript, we showed that by changing the motor used in the system from Ncd236 to the slower Ncd281 the contraction

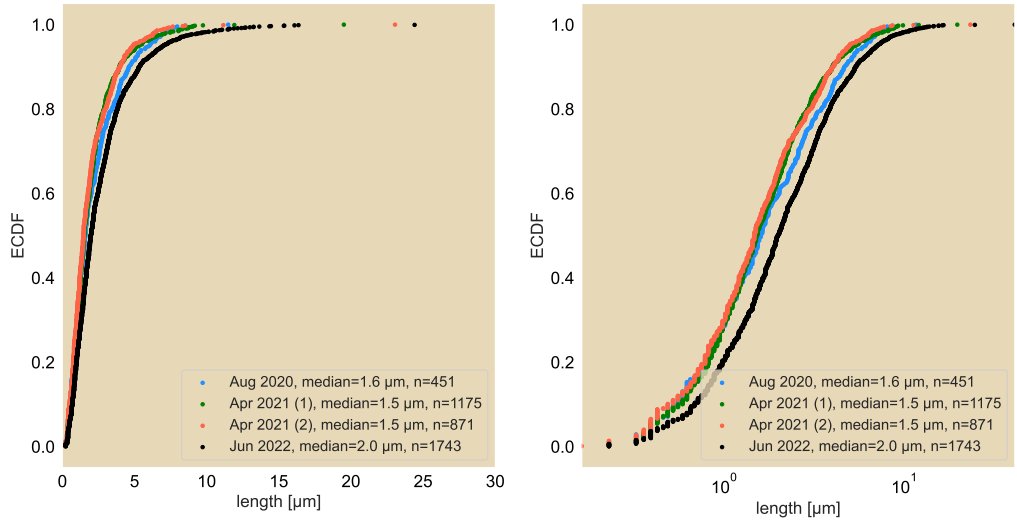


Figure B.5: **Empirical cumulative distributions of microtubule length from microtubules stabilized from polymerization preparations for experiments used in this manuscript.** Microtubules were prepared four times over the course of the work presented here, thus shown as four different datasets. Left and right plots show the same data but on different x-scales (linear for the left plot and logarithmic scale on the right). The two polymerization preparations performed in April 2021 were performed separately by two of the authors of this manuscript on the same day. n denotes the number of microtubules whose lengths were used in the ECDFs.

rate of the microtubule network decreased while the unit cell area remained uniform in time, suggesting that motor speed largely drives a local reorganization of the microtubules in the bulk of the network despite the global contraction. We similarly performed our photobleaching approach to the active contraction assay on two motors that are faster than Ncd236. Both are K401 constructs with one expressed in bacteria as in [1] while the other is expressed in insects from the constructs designed in [7]. These motors have different speeds, which we measure through gliding assays.

Fig. B.6 shows the effects of the different motor speeds on contraction speed as a function of distance from the center of the contracting network and normalized area as a function of time. Unlike Ncd281 (column B) where the contraction rate decreases relative to Ncd236 (column A), the insect-expressed (column C) and bacterial-expressed (column D) K401, both of which are faster than Ncd236, the contraction rate increases. Interestingly, despite the bacterial K401 being slower than the insect K401, they have similar contraction rates, with contraction rates of $0.0065^{+0.0009}_{-0.0008}$ s^{-1} and 0.0072 ± 0.0012 s^{-1}, respectively. Even so, we find that despite dramatic increases the contraction rate, the unit cell areas on average remain constant.

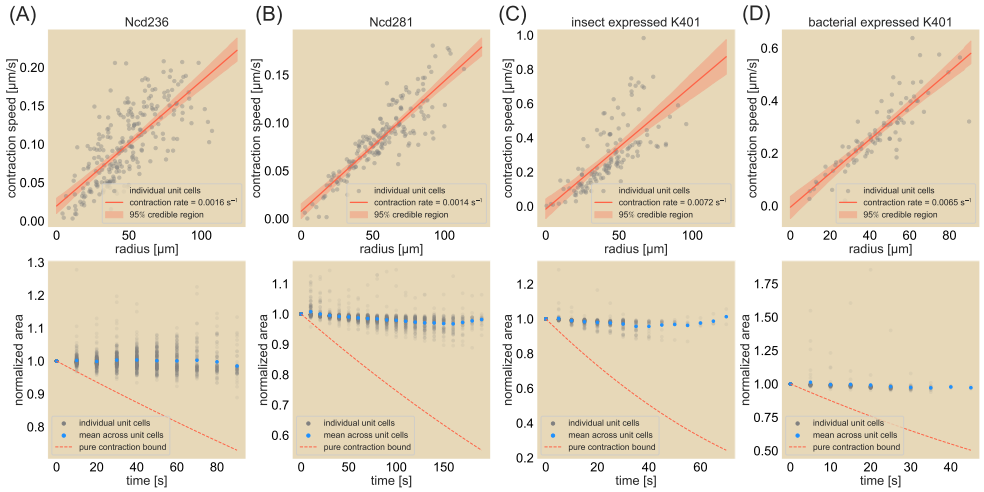


Figure B.6: **Contraction rates and unit cell area in time for four different motors.** (Top row) Contraction speed against radius for unit cell centroids with most likely contraction rate fit (red line) and 95% credible region (shaded region) and (bottom row) unit cell area as a function of time for (A) Ncd236, (B) Ncd281, (C) K401 expressed in insect cells, and (D) K401 expressed in bacteria.

B.9 The recovery of a typical FRAP-like disc is time-sensitive in the advection-diffusion model

As we derive in the Section B.12, the general solution to the PDE

$$\frac{\partial c}{\partial t} = D\nabla^2 c + \nabla \cdot \left[\frac{v_{\mathrm{m}}}{R}\mathbf{r}c\right], \tag{B.19}$$

assuming no angular dependence takes the form

$$c(r,t) = c_{\mathrm{ss}}\, e^{-\frac{r^2}{2\lambda^2}} + e^{-\frac{r^2}{2\lambda^2}} \sum_{i=1}^{\infty} c_i e^{-Dk_i^2 t}\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big), \tag{B.20}$$

where c_{ss} is the coefficient for the steady-state concentration term, $\lambda \equiv \sqrt{\frac{DR}{v_{\mathrm{m}}}}$, k_i are the eigenvalues specific to the boundary condition, c_i are the coefficients based on initial conditions, and ${}_1F_1(a;b;z)$ is the Kummer confluent hypergeometric function

$${}_1F_1(a;b;z) = \sum_{l=0}^{\infty} \frac{(a)_l}{(b)_l}\frac{z^l}{l!}, \tag{B.21}$$

where the Pochhammer symbol $(a)_l = \frac{(a+l-1)!}{(a-1)!}$. The most well-known example of Eq. B.21 is the case where $a = b$, which yields ${}_1F_1(a;a;z) = e^z$. The eigenvalues $\{k_i\}$ are found by satisfying the boundary conditions and are those terms that satisfy the equation

$$\Big(\frac{\lambda^2 k_i^2}{2}\Big){}_1F_1\Big(1 - \frac{\lambda^2 k_i^2}{2}; 2; \frac{R^2}{2\lambda^2}\Big) = 0. \tag{B.22}$$

Eq. B.80 shows that the steady-state profile of the concentration is a Gaussian distribution with standard deviation λ.

We now seek to identify the coefficients of the terms, which are specific to the initial conditions. Here, we will analytically examine three cases for initial conditions: 1) uniform concentration, 2) a uniform concentration except with molecules removed in the region $r \le R_0$ as found in many FRAP assays, and 3) a FRAP-like removal of molecules in the region $r \le R_0$ after the system initially reaches a steady-state Gaussian concentration profile.

B.9.1 Uniform concentration

We start with the case where the concentration is uniform everywhere:

$$c(r,0) = c_0. \tag{B.23}$$

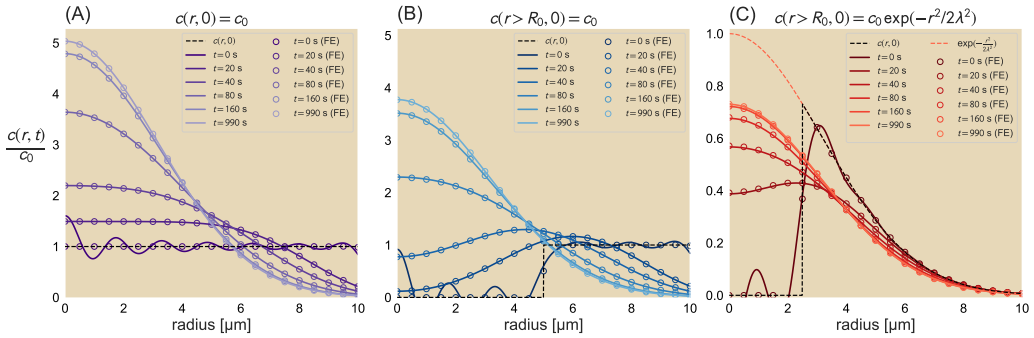


Figure B.7: **Radial advection-diffusion for various initial conditions.** (A) Uniform concentration throughout the system. (B) Uniform concentration for $r > R_0$ and no molecules for $r \leq R_0$. (C) A Gaussian distribution for $r > R_0$ and no molecules for $r \leq R_0$. Analytical solutions are presented as solid lines while solutions obtained by finite elements are shown as hollow points. The initial condition for each situation is shown as a dashed red line. For all studies, $D = 0.1\,\frac{\mu m^2}{s}$, $R = 10$ μm, and $v_m = 0.1\,\frac{\mu m}{s}$. For (B), we set $R_0 = \frac{R}{2}$ while for (C) we set $R_0 = \frac{R}{4}$. For (C), the steady-state profile prior to removing molecules for $r \leq R_0$ is shown as a dashed red line. All analytical solutions use the first 12 eigenvalues that satisfy Eq. B.79.

The solution to the PDE with this initial condition takes the form of

$$c(r,t) = \frac{c_0}{2} e^{-\frac{r^2}{2\lambda^2}} \Bigg\{ \frac{\frac{R^2}{\lambda^2}}{1 - e^{-\frac{R^2}{2\lambda^2}}} + \sum_{i=1}^{\infty} \frac{R^2\, e^{-D\,k_i^2 t}\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 2; \frac{R^2}{2\lambda^2}\Big)}{\int_0^R r'\, e^{-\frac{r'^2}{2\lambda^2}} \Big[{}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r'^2}{2\lambda^2}\Big)\Big]^2 \mathrm{d}r'} \times {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big) \Bigg\}. \tag{B.24}$$

Fig. B.7A shows the concentration profile as a function of radius and for various time points given this initial condition. Here, we used $D = 0.1\,\frac{\mu m^2}{s}$, $R = 10$ μm, and $v_m = 0.1\,\frac{\mu m}{s}$. Solid lines indicate different time points for the specific analytical solution given the uniform initial condition. These analytical solutions also show strong agreement with simulations performed by FEM which are denoted by hollow points. Here, we use the first 12 eigenvalues k_i for the analytical solution. Similar to the decomposition of a square wave into a sum of sinusoidal functions yielding imperfect agreement with the original function, we see here that the use of a limited number of eigenvalues that satisfy Eq. B.79 leads to fluctuations about the original function for $t = 0$ (see Appendix B.14 on Gibbs phenomenon). Nevertheless, we see that these

fluctuations in the analytical condition quickly smooth out for $t > 0$. For the given parameters, the concentration at larger radii decreases quickly due to the higher advection overcoming diffusion. As shown at $t = 20$ seconds and $t = 40$ seconds, the concentration appears roughly uniform at lower concentrations but the length scale of this uniformity appears to decrease. At $t = 990$ seconds, the concentration profile reaches the Gaussian steady-state solution where the concentration gradient allows diffusion to counter the advective flow.

B.9.2 Uniform concentration for $r > R_0$

We apply a similar initial condition as that used in Sec. B.9.1, but remove any molecules within a distance R_0 from the origin as typically performed in FRAP experiments. This initial condition is mathematically described by

$$c(r,0) = \begin{cases} 0 & \text{if } r \leq R_0, \\ c_0 & \text{if } r > R_0. \end{cases} \tag{B.25}$$

The solution for this initial condition is similar to Eq. B.24 but with different limits of integration (see Appendix B.11 on Sturm-Liouville Theory and S2 for application of the theory in 2D):

$$c(r,t) = \frac{c_0}{2} e^{-\frac{r^2}{2\lambda^2}} \left\{ \frac{\frac{R^2}{\lambda^2} - \frac{R_0^2}{\lambda^2}}{1 - e^{-\frac{R^2}{2\lambda^2}}} + \sum_{i=1}^{\infty} \alpha_i e^{-D\,k_i^2 t}\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r^2}{2\lambda^2} \Big) \right\}, \tag{B.26}$$

where

$$\alpha_i = \frac{R^2\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 2; \frac{R^2}{2\lambda^2} \Big) - R_0^2\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 2; \frac{R_0^2}{2\lambda^2} \Big)}{\int_0^R r'\, e^{-\frac{r'^2}{2\lambda^2}} \Big[{}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r'^2}{2\lambda^2} \Big) \Big]^2 \mathrm{d}r'}. \tag{B.27}$$

As $R_0 \to 0$ in Eq. B.26 we recover Eq. B.24. Fig. B.7B shows traces of the concentration profile at the same times as in Fig. B.7A. Here, $R_0 = \frac{R}{2}$. Once again, we see that the analytical solution for $t = 0$ fluctuates about the defined initial condition but quickly smooth out and agree well with FEM results (hollow points) for $t > 0$. By removing molecules at $r \leq R_0$, a wave of molecules move toward the origin from a combination of advection toward the origin and diffusion moving molecules against the concentration gradient while the concentration at $r \to R$ recedes. Once again, we recover a Gaussian profile, but at a lower maximum than that observed in Fig. B.7A due to the lower initial number of molecules.

B.9.3 Gaussian profile for $r > R_0$

Finally, consider a situation where molecules in this advective-diffusive system are allowed to reach steady-state before photobleaching all molecules within a certain radius of the center $r \leq R_0$. The initial conditions would appear as

$$c(r,0) = \begin{cases} 0 & \text{if } r \leq R_0, \\ c_0\, e^{-\frac{r^2}{2\lambda^2}} & \text{if } r > R_0. \end{cases} \tag{B.28}$$

We show analytically that the concentration profile is

$$c(r,t) = c_0 e^{-\frac{r^2}{2\lambda^2}} \Bigg\{ \frac{e^{-\frac{R_0^2}{2\lambda^2}} - e^{-\frac{R^2}{2\lambda^2}}}{1 - e^{-\frac{R^2}{2\lambda^2}}} - \frac{1}{2} \sum_{i=1}^{\infty} \beta_i e^{-Dk_i^2 t}\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r^2}{2\lambda^2} \Big) \Bigg\}, \tag{B.29}$$

where

$$\beta_i = \frac{R_0^2\, {}_1F_1\Big(1 + \frac{\lambda^2 k_i^2}{2}; 2; -\frac{R_0^2}{2\lambda^2}\Big)}{\int_0^R r'\, e^{-\frac{r'^2}{2\lambda^2}} \Big[{}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r'^2}{2\lambda^2} \Big) \Big]^2 \mathrm{d}r'}. \tag{B.30}$$

Once again the analytical solution agrees with simulations of the same initial condition shown in Fig. B.7C for $R_0 = \frac{R}{4}$. We note here that as $R_0 \to 0$ we recover the steady-state solution again as the time-dependent terms vanish and the ratio of exponentials in the time-independent term goes to unity. Fig. B.7C shows again the imperfection of the analytical solution for $t = 0$ and the initial condition but a strong agreement with FEM results. In this situation, the concentration toward the outer edge of the system remains largely unchanged as diffusion and advection are balanced toward the boundary. However, at smaller radii of the system, there is a shift in concentration as molecules enter the $r \leq R_0$ region and for the chosen parameter values, the overall concentration profile returns to a Gaussian distribution within 3 minutes.

Across all three initial conditions, the trend toward a Gaussian distribution as the steady-state profile shows that in experimental systems exhibiting such an advective-diffusive behavior the use of FRAP becomes sensitive to the time when photobleaching is applied. If the concentration profile in the system has already begun to move away from a uniform distribution, such as the initial contraction of a highly connected filament network, then the molecule

redistribution until steady state is achieved will show different recovery profiles from that of an experiment where photobleaching is applied at a time when the system is already close to reaching the steady-state profile. Such results provide the two extremes of "fluorescence recovery" in potential *in vitro* assays that evolve from a uniform concentration to a Gaussian-shaped distribution subject to this advection-diffusion system.

B.10 1D telescoping model

In this work, we present a theory for the redistribution of particles influenced by diffusion and advection with a linear velocity profile directed toward the origin. This theoretical analysis is meant to explore the filament concentration when subject to a linear contraction velocity profile. We start by illustrating this in a 1D system of length L. The velocity as a function of position is described by

$$v(x) = -v_{\mathrm{m}}\frac{x}{L},\ 0 \le x \le L, \tag{B.31}$$

where v_{m} is the maximum particle velocity in this system, located at $x = L$. We also note that the velocity is negative to indicate that the particles are moving toward $x = 0$. The general one-dimensional advection-diffusion equation says that the concentration changes in space and time $c(x,t)$ in the form

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - \frac{\partial}{\partial x}\Big[v(x)\,c\Big], \tag{B.32}$$

for D the diffusion constant. With a linear velocity profile, Eq. B.32 takes the form

$$\begin{aligned}\frac{\partial c}{\partial t} &= D\frac{\partial^2 c}{\partial x^2} + \frac{\partial}{\partial x}\Big(v_{\mathrm{m}}\frac{x}{L}c\Big),\\ &= D\frac{\partial^2 c}{\partial x^2} + \frac{v_{\mathrm{m}}\,x}{L}\frac{\partial c}{\partial x} + \frac{v_{\mathrm{m}}}{L}c.\end{aligned} \tag{B.33}$$

To solve Eq. B.33, we apply a separation of variables where our ansatz for the concentration of particles is

$$c(x,t) = \Phi(x)T(t). \tag{B.34}$$

We apply Eq. B.34 to Eq. B.33 and divide by $D\Phi(x)T(t)$ to get

$$\frac{1}{DT}\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}x^2} + \frac{v_\mathrm{m}x}{DL}\frac{1}{\Phi}\frac{\mathrm{d}\Phi}{\mathrm{d}x} + \frac{v_\mathrm{m}}{DL}. \tag{B.35}$$

Due to the left-hand and right-hand sides of the equation depending only on t and x, respectively, we can say that both sides of Eq. B.35 are the same constant $-k^2$. We then solve the left-hand side of Eq. B.35:

$$\begin{aligned}\frac{1}{DT}\frac{\mathrm{d}T}{\mathrm{d}t} &= -k^2,\\ T(t) &= e^{-Dk^2t}.\end{aligned} \tag{B.36}$$

We are left to solve the right-hand side of Eq. B.35. Here, we get

$$\begin{aligned}-k^2 &= \frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}x^2} + \frac{v_\mathrm{m}x}{DL}\frac{1}{\Phi}\frac{\mathrm{d}\Phi}{\mathrm{d}x} + \frac{v_m}{DL},\\ 0 &= \frac{\mathrm{d}^2\Phi}{\mathrm{d}x^2} + \frac{v_\mathrm{m}\,x}{DL}\frac{\mathrm{d}\Phi}{\mathrm{d}x} + \Phi\Big[\frac{v_\mathrm{m}}{DL} + k^2\Big].\end{aligned} \tag{B.37}$$

We define a parameter $\alpha^2 = \frac{v_\mathrm{m}}{DL}$. When implemented into Eq. B.37, we get

$$\begin{aligned}0 &= \frac{\mathrm{d}^2\Phi}{\mathrm{d}x^2} + \alpha^2 x\frac{\mathrm{d}\Phi}{\mathrm{d}x} + \Phi\Big[\alpha^2 + k^2\Big],\\ 0 &= \frac{1}{\alpha^2}\frac{\mathrm{d}^2\Phi}{\mathrm{d}x^2} + x\frac{\mathrm{d}\Phi}{\mathrm{d}x} + \Phi\Big[1 + \Big(\frac{k}{\alpha}\Big)^2\Big],\\ &= \frac{\mathrm{d}^2\Phi}{\mathrm{d}\tilde{x}^2} + \tilde{x}\frac{\mathrm{d}\Phi}{\mathrm{d}\tilde{x}} + \Phi\Big(1 + \tilde{k}^2\Big),\end{aligned} \tag{B.38}$$

where we redefined $\tilde{x} = \alpha x$ and $\tilde{k} = \frac{k}{\alpha}$.

In order to solve for Φ, we applied the ODE into Wolfram Alpha. The general solution takes the form

$$\Phi(\tilde{x}) = c_\mathrm{ss}\, e^{-\frac{\tilde{x}^2}{2}} + c_1\, e^{-\frac{\tilde{x}^2}{2}} H_{\tilde{k}^2}\Big(\frac{\tilde{x}}{\sqrt{2}}\Big) + c_2\, e^{-\frac{\tilde{x}^2}{2}}\, {}_1F_1\Big(-\frac{\tilde{k}^2}{2}; \frac{1}{2}; \frac{\tilde{x}^2}{2}\Big), \tag{B.39}$$

where $H_n(x)$ is the nth Hermite polynomial and ${}_1F_1(a;b;z)$ is the Kummer confluent hypergeometric function. When we apply no-flux boundary conditions to the problem, we are looking to satisfy the conditions $J_x|_{x=0} = 0$ and

$J_x|_{x=L} = 0$. We note that $J_x = D\frac{\mathrm{d}\Phi}{\mathrm{d}x} - v(x)\Phi(x)$ rather than simply $\frac{\mathrm{d}\Phi}{\mathrm{d}x} = 0$ at the boundaries where the advection of material coming in must be countered by diffusion going outward to ensure that the number of particles is constant in the system.

Fortunately, both boundary conditions are satisfied for the steady-state solution. However, when we apply these conditions to the Hermite polynomials, the condition at $x = 0$ requires that $\tilde{k}^2$ be an even integer, but the boundary condition at $x = L$ requires that

$$H'_{2n}\Big(\frac{\alpha\, L}{\sqrt{2}}\Big) = 0. \tag{B.40}$$

To be able to satisfy this boundary condition, we would have to ensure that the derivative of each even function of the Hermite polynomial is 0 at $\frac{\alpha\, L}{\sqrt{2}}$. However, as L, v_m, and D are defined properties of the system, we are left to argue that the coefficients of the Hermite polynomials are 0. Finally, we check that the Kummer confluent hypergeometric function can satisfy our boundary conditions. We start with $x = 0$:

$$\begin{aligned}
\frac{\mathrm{d}\Phi}{\mathrm{d}x}\Big|_{x=0} &= 0,\\
c_2\frac{\mathrm{d}\tilde{x}}{\mathrm{d}x}\frac{\mathrm{d}}{\mathrm{d}\tilde{x}}\Big[e^{-\frac{\tilde{x}^2}{2}}{}_1F_1\Big(-\frac{\tilde{k}^2}{2};\frac{1}{2};\frac{\tilde{x}^2}{2}\Big)\Big] &= 0,\\
c_2\,\alpha\, e^{-\frac{\tilde{x}^2}{2}}\Big[-\tilde{x}\,{}_1F_1\Big(-\frac{\tilde{k}^2}{2};\frac{1}{2};\frac{\tilde{x}^2}{2}\Big) + \sum_{l=1}^{\infty}\tilde{x}\frac{\Big(-\frac{\tilde{k}^2}{2}\Big)_l}{\Big(\frac{1}{2}\Big)_l}\frac{\Big(\frac{\tilde{x}^2}{2}\Big)^{l-1}}{(l-1)!}\Big] &= 0,\\
0 &= 0.
\end{aligned} \tag{B.41}$$

Eq. B.41 shows that all terms of the function for Φ will satisfy the boundary conditions without a need to specify $\tilde{k}$. Applying the boundary condition at $x = L$ gives:

$${}_1F_1\Big(1-\frac{\tilde{k}^2}{2};\frac{3}{2};\frac{(\alpha L)^2}{2}\Big) = 0. \tag{B.42}$$

In order to get to this solution, we used the case that $(a)_l = (a)(a+1)(a+2)...(a+l-1) = a\,(a+1)_{l-1}$ so that we return to a Hypergeometric function. In essence, we then need to solve for $\tilde{k}$ through Eq. B.42 in order to obtain each value of k in our original problem.

Fig. B.8 plots the left-hand side of Eq. B.42 as a function of $\tilde{k}$ when $\alpha L = 1$. That is, for simplicity, we set all of the parameters of the system to unity. In

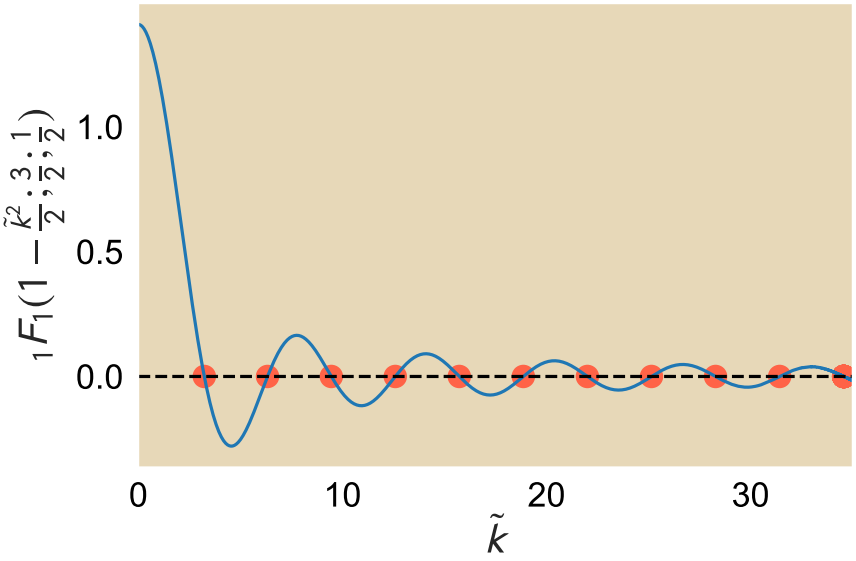


Figure B.8: **Zeros of $\tilde{k}$ for $_1F_1\Big(1-\frac{\tilde{k}^2}{2};\frac{3}{2};\frac{(\alpha L)^2}{2}\Big)=0$ where $\alpha L=1$.** Red dots are overlayed with the points where the Kummer confluent hypergeometric function crosses the x-axis.

this case, we can see a roughly periodic nature to the hypergeometric function. The first five solutions for $\tilde{k}$ are $\tilde{k}=3.231$, 6.329, 9.456, 12.589, and 15.727, which we will refer to later.

So far, the solution to Eq. B.33 with no-flux boundary conditions is

$$c(x,t)=c_{\mathrm{ss}}\exp\Big(-\frac{v_{\mathrm{m}}x^2}{2\,D\,L}\Big) + \sum_{j=1}^{\infty}c_j\exp\Big(-Dk_j^2t-\frac{v_{\mathrm{m}}x^2}{2DL}\Big)\,{}_1F_1\Big(-\frac{DLk_j^2}{2v_{\mathrm{m}}};\frac{1}{2};\frac{v_{\mathrm{m}}x^2}{2DL}\Big), \quad \text{(B.43)}$$

where k_j is determined from finding the values of $\tilde{k}_j\equiv\frac{k_j}{\alpha}$ for which $_1F_1\Big(1-\frac{\tilde{k}^2}{2};\frac{3}{2};\frac{(\alpha L)^2}{2}\Big)=0$. In the case where we set $L=1$, $D=1$, and $v_{\mathrm{m}}=1$, we are solving

$$c(x,t)=c_{\mathrm{ss}}\exp\Big(-\frac{x^2}{2}\Big)+\sum_{j=1}^{\infty}c_j\exp\Big(-k_j^2t-\frac{x^2}{2}\Big)\,{}_1F_1\Big(-\frac{k_j^2}{2};\frac{1}{2};\frac{x^2}{2}\Big), \quad \text{(B.44)}$$

$$_1F_1\Big(1-\frac{k_j^2}{2};\frac{3}{2};\frac{1}{2}\Big)=0.$$

Here, we have determined the first few values of k_j that satisfy the no-flux boundary condition. We now find the coefficients $\{c_j\}$ from solving the initial

condition. There are many possible initial conditions we could consider, but suppose we let a one-dimensional aster assay carry out to form a steady-state aster. At $t < 0$, the concentration of fluorescent molecules in the system is the steady-state concentration profile $c_{\text{ss}}\, e^{-\frac{x^2}{2}}$, but then we photobleach the molecules at positions $x < x_0 < 1$. In this case, our initial conditions appear as

$$c(x,0) = \begin{cases} 0 & \text{if } x \leq x_0, \\ c_0\, e^{-\frac{x^2}{2}} & \text{if } x > x_0. \end{cases} \tag{B.45}$$

In order to solve the initial conditions, we must multiply both sides of Eq. B.43 by an eigenfunction with some value of k_h that satisfies the boundary conditions, $e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)$. We also use the weighting function $w(x) = e^{\frac{x^2}{2}}$ as derived in Appendix B.11:

$$\begin{aligned}
\int_0^1 c(x,0)\, w(x)\, e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\, \mathrm{d}x &= c_{\text{ss}} \int_0^1 e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\, \mathrm{d}x \\
&\quad + \Bigg[\sum_{j=1}^{\infty} c_j \int_0^1 e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_j^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big) \\
&\quad \times {}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\mathrm{d}x\Bigg], \\
c_0 \int_{x_0}^1 e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\, \mathrm{d}x &= c_{\text{ss}} \int_0^1 e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\, \mathrm{d}x \\
&\quad + \Bigg[\sum_{j=1}^{\infty} c_j \int_0^1 e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_j^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big) \\
&\quad \times {}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\mathrm{d}x\Bigg].
\end{aligned} \tag{B.46}$$

We first tackle the left-hand side of Eq. B.46. By invoking a relation for Kummer confluent hypergeometric functions of the first kind:

$${}_1F_1\Big(a; b; x\Big) = e^x\, {}_1F_1\Big(b-a; b; -x\Big), \tag{B.47}$$

the integral can be altered to take the form

$$c_0 \int_{x_0}^{1} e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\,\mathrm{d}x = c_0 \int_{x_0}^{1} {}_1F_1\Big(\frac{1+k_h^2}{2}; \frac{1}{2}; -\frac{x^2}{2}\Big)\,\mathrm{d}x,$$
$$= c_0 \frac{x}{2} \sum_{i=0}^{\infty} \frac{\Big(\frac{1}{2}+\frac{k_h^2}{2}\Big)_i}{\Big(\frac{1}{2}\Big)_i \Big(i+\frac{1}{2}\Big)} \frac{\Big(-\frac{x^2}{2}\Big)^i}{i!}\Bigg|_{x_0}^{1}, \quad \text{(B.48)}$$

and by using $(a)_i(a+i) = a(a+1)...(a+i-1)(a+i) = a(a+1)_i$, and using Eq. B.47 we get

$$c_0 \int_{x_0}^{1} e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\,\mathrm{d}x = c_0 \frac{x}{2\frac{1}{2}} \sum_{i=0}^{\infty} \frac{\Big(\frac{1}{2}+\frac{k_h^2}{2}\Big)_i}{\Big(\frac{3}{2}\Big)_i} \frac{\Big(-\frac{x^2}{2}\Big)^i}{i!}\Bigg|_{x_0}^{1},$$
$$= c_0\, x\, {}_1F_1\Big(\frac{1}{2}+\frac{k_h^2}{2}; \frac{3}{2}; -\frac{x^2}{2}\Big)\Bigg|_{x_0}^{1},$$
$$= c_0\, x\, e^{-\frac{x^2}{2}}\, {}_1F_1\Big(1-\frac{k_h^2}{2}; \frac{3}{2}; \frac{x^2}{2}\Big)\Bigg|_{x_0}^{1},$$
$$= c_0 \Big[e^{-\frac{1}{2}}\, {}_1F_1\Big(1-\frac{k_h^2}{2}; \frac{3}{2}; \frac{1}{2}\Big)$$
$$- x_0\, e^{-\frac{x_0^2}{2}}\, {}_1F_1\Big(1-\frac{k_h^2}{2}; \frac{3}{2}; \frac{x_0^2}{2}\Big)\Big],$$
$$= -c_0\, x_0 e^{-\frac{x_0^2}{2}}\, {}_1F_1\Big(1-\frac{k_h^2}{2}; \frac{3}{2}; \frac{x_0^2}{2}\Big). \quad \text{(B.49)}$$

Where the first term in the penultimate line is 0 due to Eq. B.42.

Integrating the term with the steady-state solution simply leads to an integral of the hypergeometric function:

$$\int_{0}^{1} e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\mathrm{d}x = e^{-\frac{1}{2}}\, {}_1F_1\Big(1-\frac{k_h^2}{2}; \frac{3}{2}; \frac{1}{2}\Big),$$
$$= 0. \quad \text{(B.50)}$$

So the first integral on the right-hand side vanishes. This makes sense as the steady-state function, being an eigenfunction of the PDE is orthogonal to the eigenfunction chosen.

Finally, we solve for the second integral on the right-hand side. We showed in Eq. B.66 of Sec. B.11 that for $j \neq h$, the integral is 0. This leaves only

one integral to tackle, where $j = h$. For this problem, this integral must be performed numerically. The coefficients are then solved as

$$c_h = -c_0 \frac{x_0 e^{-\frac{x_0^2}{2}}\, {}_1F_1\Big(1 - \frac{k_h^2}{2}; \frac{3}{2}; \frac{x_0^2}{2}\Big)}{\int_0^1 e^{-\frac{x^2}{2}} \Big[{}_1F_1\Big(-\frac{k_h^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\Big]^2 \mathrm{d}x}. \tag{B.51}$$

Finally, we determine the coefficient for the steady-state solution. To achieve this, we multiply both sides by the weighting function $w(x)$ and the steady-state eigenfunction as prescribed in Eq. B.62 of Sec. B.11. In this case, the product of the two functions cancel, so we integrate each side over the system size:

$$\begin{aligned} c_0 \int_{x_0}^1 e^{-\frac{x^2}{2}}\, \mathrm{d}x &= c_{\mathrm{ss}} \int_0^1 e^{-\frac{x^2}{2}} \mathrm{d}x + \sum_{j=1}^{\infty} c_j \int_0^1 e^{-\frac{x^2}{2}}\, {}_1F_1\Big(-\frac{k_j^2}{2}; \frac{1}{2}; \frac{x^2}{2}\Big)\mathrm{d}x, \\ c_0 \sqrt{\frac{\pi}{2}} \mathrm{erf}\Big(\frac{x}{\sqrt{2}}\Big)\Big|_{x_0}^1 &= c_{\mathrm{ss}} \sqrt{\frac{\pi}{2}} \mathrm{erf}\Big(\frac{x}{\sqrt{2}}\Big)\Big|_0^1, \\ c_{\mathrm{ss}} &= c_0 \left[1 - \frac{\mathrm{erf}\Big(\frac{x_0}{\sqrt{2}}\Big)}{\mathrm{erf}\Big(\frac{1}{\sqrt{2}}\Big)}\right], \end{aligned} \tag{B.52}$$

where $\mathrm{erf}(x)$ is the Gauss error function and the integrals with the hypergeometric functions vanish as demonstrated from Eq. B.50. When we assemble all of the terms for this particular initial condition and reintroduce the parameters, the solution takes the form

$$\begin{aligned} c(x,t) = c_0\, e^{-\frac{v_{\mathrm{m}} x^2}{2DL}} \Bigg\{ 1 - \frac{\mathrm{erf}\Big(x_0 \sqrt{\frac{v_{\mathrm{m}}}{2DL}}\Big)}{\mathrm{erf}\Big(\sqrt{\frac{v_{\mathrm{m}} L}{2D}}\Big)} \\ - \sum_{j=1}^{\infty} \frac{x_0 e^{-\frac{v_{\mathrm{m}} x_0^2}{2DL}}\, {}_1F_1\Big(1 - \frac{DLk_j^2}{2v_{\mathrm{m}}}; \frac{3}{2}; \frac{v_{\mathrm{m}} x_0^2}{2DL}\Big)}{\int_0^L e^{-\frac{v_{\mathrm{m}} x^2}{2DL}} \Big[{}_1F_1\Big(-\frac{DLk_j^2}{2v_{\mathrm{m}}}; \frac{1}{2}; \frac{x^2}{2}\Big)\Big]^2 \mathrm{d}x} e^{-Dk_j^2 t}\, {}_1F_1\Big(-\frac{DLk_j^2}{2v_{\mathrm{m}}}; \frac{1}{2}; \frac{v_{\mathrm{m}} x^2}{2DL}\Big) \Bigg\}. \end{aligned} \tag{B.53}$$

Fig. B.9A illustrates the initial Gaussian profile (red dashed line) prior to photobleaching from $x < x_0$ (solid black line). Fig. B.9B shows the FRAP recovery process at various time units as solved in Eq. B.53. We observe that the increase in concentration toward $x = 0$ and the decrease in concentration toward $x = L = 1$ appear to generally match one another over the course of the recover. We also see that by $t = 0.500$, we have returned to a Gaussian profile as the steady-state profile, but with a reduced peak concentration.

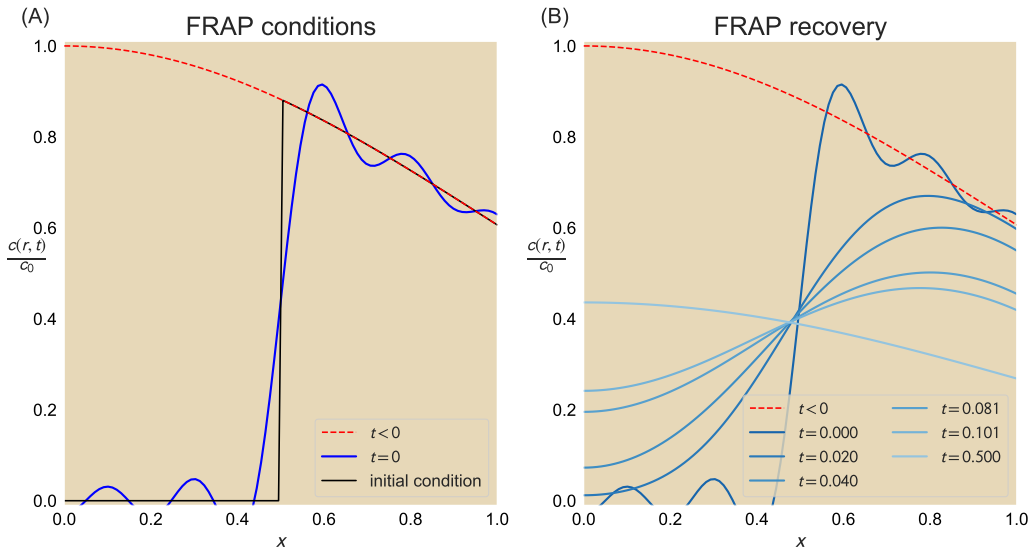


Figure B.9: **FRAP for 1D advection diffusion with linear velocity profile.** (A) Initial steady-state profile of the concentration (red dashed line) before photobleaching the system for $x < x_0$ (solid black line). The blue line is obtained from Eq. B.53 for the first nonzero values of k_j for the given problem. (B) Time evolution of the concentration after photobleaching. Decreasing shades of blue designate later time points of the concentration profile.

B.11 Sturm-Liouville Theory

The Sturm-Liouville theory says that all second-order linear ordinary differential equations can be written in the form

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big[p(x)\frac{\mathrm{d}y}{\mathrm{d}x}\Big] + q(x)\,y(x) = -\lambda\,w(x)\,y(x). \tag{B.54}$$

Importantly, $w(x)$ is the weighting function, which provides the means for satisfying the orthogonality relations for finding coefficients of each term in the series solution to the partial differential equation. Specifically, if we were to write the ODE in the form

$$P(x)\,y''(x) + Q(x)\,y'(x) + R(x)\,y(x) = f(x), \tag{B.55}$$

for functions $P(x)$, $Q(x)$, $R(x)$, and $f(x)$, then there is a multiplicative function that can be determined by

$$m(x) = \exp\Big(\int \frac{Q(x) - P'(x)}{P(x)}\mathrm{d}x\Big). \tag{B.56}$$

This multiplicative function is then multiplied to Eq. B.55 and recast into the form shown in Eq.B.54. Thus, with $P(\tilde{x}) = 1$ and $Q(\tilde{x}) = \tilde{x}$,

$$\begin{aligned} m(\tilde{x}) &= \exp\Big(\int \tilde{x}\,\mathrm{d}\tilde{x}\Big), \\ &= \exp\Big(\frac{\tilde{x}^2}{2}\Big), \end{aligned} \tag{B.57}$$

and the ODE takes the form

$$0 = \frac{\mathrm{d}}{\mathrm{d}\tilde{x}}\Big[e^{\frac{\tilde{x}^2}{2}}\frac{\mathrm{d}\Phi}{\mathrm{d}\tilde{x}}\Big] + \Phi\Big(1+\tilde{k}^2\Big)e^{\frac{\tilde{x}^2}{2}}, \tag{B.58}$$

or in the form of Eq. B.54:

$$\frac{\mathrm{d}}{\mathrm{d}\tilde{x}}\Big[e^{\frac{\tilde{x}^2}{2}}\frac{\mathrm{d}\Phi}{\mathrm{d}\tilde{x}}\Big] + e^{\frac{\tilde{x}^2}{2}}\Phi = -\tilde{k}^2 e^{\frac{\tilde{x}^2}{2}}\Phi, \tag{B.59}$$

so that $p(x) = q(x) = w(x) = e^{\frac{\tilde{x}^2}{2}}$ and $\lambda = \tilde{k}^2$. We note the weighting function here is the same as the multiplicative function for the 1D advection-diffusion equation reported here.

Next, we show the orthogonality conditions of the eigenfunctions. Suppose that solving Eq. B.54 creates a series of eigenfunctions $\{y_j(x)\}$. Suppose that a given eigenfunction $y_i(x)$ has the eigenvalue λ_i so that

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big[p(x)\frac{\mathrm{d}y_i}{\mathrm{d}x}\Big] + q(x)\,y_i(x) = -\lambda_i\,w(x)\,y_i(x). \tag{B.60}$$

Suppose that each eigenfunction of the system, bounded by $a \le x \le b$, obeys the boundary conditions

$$\begin{aligned} \alpha_1 y_i(a) + \alpha_2 y_i'(a) &= 0, \\ \beta_1 y_i(b) + \beta_2 y_i'(b) &= 0. \end{aligned} \tag{B.61}$$

To test the orthogonality conditions, we multiply both sides by $y_j(x)$, a particular eigenfunction of the differential equation, and integrate over the entire system:

$$\int_a^b \frac{\mathrm{d}}{\mathrm{d}x}\Big[p(x)\frac{\mathrm{d}y_i}{\mathrm{d}x}\Big]y_j(x) + q(x)\,y_i(x)\,y_j(x)\mathrm{d}x = -\lambda_i \int_a^b w(x)\,y_i(x)\,y_j(x)\mathrm{d}x,$$
$$p(x)\frac{\mathrm{d}y_i}{\mathrm{d}x}y_j(x)\Big|_a^b - \int_a^b p(x)\frac{\mathrm{d}y_i}{\mathrm{d}x}\frac{\mathrm{d}y_j}{\mathrm{d}x}\mathrm{d}x + \int_a^b q(x)\,y_i(x)\,y_j(x)\mathrm{d}x = -\lambda_i \int_a^b w(x)\,y_i(x)\,y_j(x)\mathrm{d}x. \tag{B.62}$$

Had Eq. B.60 involved $y_j(x)$ and we multiplied both sides of the equation by $y_i(x)$, then Eq. B.62 would have the subscripts reversed:

$$p(x)\frac{\mathrm{d}y_j}{\mathrm{d}x}y_i(x)\Big|_a^b - \int_a^b p(x)\frac{\mathrm{d}y_i}{\mathrm{d}x}\frac{\mathrm{d}y_j}{\mathrm{d}x}\mathrm{d}x + \int_a^b q(x)\,y_i(x)\,y_j(x)\mathrm{d}x = -\lambda_j \int_a^b w(x)\,y_i(x)\,y_j(x)\mathrm{d}x. \tag{B.63}$$

Suppose we subtracted Eq. B.63 from Eq. B.62 and applied our boundary conditions:

$$\begin{aligned}
-(\lambda_i-\lambda_j)\int_a^b w(x)\,y_i(x)\,y_j(x)\mathrm{d}x &= p(x)\frac{\mathrm{d}y_i}{\mathrm{d}x}y_j(x)\Big|_a^b - p(x)\frac{\mathrm{d}y_i}{\mathrm{d}x}y_j(x)\Big|_a^b,\\
-(\lambda_i-\lambda_j)\int_a^b w(x)\,y_i(x)\,y_j(x)\mathrm{d}x &= p(b)\Big[\frac{\mathrm{d}y_i}{\mathrm{d}x}\Big|_b y_j(b) - \frac{\mathrm{d}y_j}{\mathrm{d}x}\Big|_b y_i(b)\Big]\\
&\quad - p(a)\Big[\frac{\mathrm{d}y_i}{\mathrm{d}x}\Big|_a y_j(a) - \frac{\mathrm{d}y_j}{\mathrm{d}x}\Big|_a y_i(a)\Big],\\
-(\lambda_i-\lambda_j)\int_a^b w(x)\,y_i(x)\,y_j(x)\mathrm{d}x &= p(b)\Big[\frac{\beta_1}{\beta_2}y_i(b)\,y_j(b) - \frac{\beta_1}{\beta_2}y_i(b)\,y_j(b)\Big]\\
&\quad - p(a)\Big[\frac{\alpha_1}{\alpha_2}y_i(a)\,y_j(a) - \frac{\alpha_1}{\alpha_2}y_i(a)\,y_j(a)\Big],\\
-(\lambda_i-\lambda_j)\int_a^b w(x)\,y_i(x)\,y_j(x)\mathrm{d}x &= 0.
\end{aligned} \tag{B.64}$$

If $i = j$, then the left-hand side is already zero.

$$-\lambda_i \int_a^b w(x)\Big[y_i(x)\Big]^2\mathrm{d}x = p(x)\frac{\mathrm{d}y_i}{\mathrm{d}x}y_i(x)\Big|_a^b - \int_a^b p(x)\Big[\frac{\mathrm{d}y_i}{\mathrm{d}x}\Big]^2\mathrm{d}x + \int_a^b q(x)\Big[y_i(x)\Big]^2\mathrm{d}x. \tag{B.65}$$

We will return to the case where $i = j$ to find the coefficients of eigenfunction. If $i \neq j$, then the eigenvalues are different here and the integral is zero:

$$\int_a^b w(x)\, y_i(x)\, y_j(x)\mathrm{d}x = 0, \text{for } i \neq j. \tag{B.66}$$

Though not true for the 1D case, Eq. B.65 may serve as a convenient equation for analytically solving the coefficients for each eigenfunction.

B.12 2D telescoping model

In the 2D telescoping case, we assume that we are carrying out an aster assay experiment where we dimerize motors (and thus couple microtubules) in a circular region of radius R. We assume that the distributions of motors and microtubules are strictly radially dependent and thus have no angular dependence. Finally, we model the velocity profile of the microtubule movement by assuming radially inward advection of particles where those that lie further away from the origin move faster than those toward the center:

$$\mathbf{v} = -v_\mathrm{m}\frac{r}{R}\hat{r}. \tag{B.67}$$

The advection-diffusion equation then takes the form

$$\begin{aligned}
\frac{\partial c}{\partial t} &= D\nabla^2 c - \nabla\cdot(\mathbf{v}c), \\
&= \frac{D}{r}\frac{\partial}{\partial r}\Big(r\frac{\partial c}{\partial r}\Big) + \frac{v_\mathrm{m}}{R}\frac{1}{r}\frac{\partial}{\partial r}(r^2 c), \\
&= D\frac{\partial^2 c}{\partial r^2} + \frac{D}{r}\frac{\partial c}{\partial r} + \frac{v_\mathrm{m} r}{R}\frac{\partial c}{\partial r} + \frac{2v_\mathrm{m}c}{R}, \\
&= D\frac{\partial^2 c}{\partial r^2} + \Big(\frac{D}{r} + \frac{v_\mathrm{m} r}{R}\Big)\frac{\partial c}{\partial r} + \frac{2v_\mathrm{m}c}{R}, \\
\frac{1}{D}\frac{\partial c}{\partial t} &= \frac{\partial^2 c}{\partial r^2} + \Big(\frac{1}{r} + \frac{v_\mathrm{m} r}{DR}\Big)\frac{\partial c}{\partial r} + \frac{2v_\mathrm{m}c}{DR}.
\end{aligned} \tag{B.68}$$

We first follow the procedure of separation of variables $c(r,t) = \Phi(r)T(t)$ and determine that the time-dependent component takes on the familiar form of $e^{-Dk^2 t}$. This ansatz is then applied to Eq. B.68 and rewrite the spatial component of the concentration as

$$\begin{aligned}
-k^2\Phi &= \frac{\mathrm{d}^2\Phi}{\mathrm{d}r^2} + \Big(\frac{1}{r} + \frac{v_\mathrm{m} r}{DR}\Big)\frac{\mathrm{d}\Phi}{\mathrm{d}r} + \frac{2v_\mathrm{m}\Phi}{DR}, \\
0 &= r\frac{\mathrm{d}^2\Phi}{\mathrm{d}r^2} + \Big(1 + \frac{v_\mathrm{m} r^2}{DR}\Big)\frac{\mathrm{d}\Phi}{\mathrm{d}r} + \Big(\frac{2v_\mathrm{m}}{DR} + k^2\Big)r\Phi.
\end{aligned} \tag{B.69}$$

We will define a new length scale $\lambda^2 \equiv \frac{DR}{v_\mathrm{m}}$ as well as a change of variables $\rho \equiv \frac{r}{\lambda}$ and $\tilde{k} \equiv \lambda k$. In this case, Eq. B.69 takes the altered form

$$0 = \rho\frac{\mathrm{d}^2\Phi}{\mathrm{d}\rho^2} + (1+\rho^2)\frac{\mathrm{d}\Phi}{\mathrm{d}\rho} + \Big(2 + \tilde{k}^2\Big)\rho\Phi. \tag{B.70}$$

We obtain the multiplicative function by following the prescription from Eq. B.56 in Sec. B.11:

$$m(\rho) = e^{\frac{\rho^2}{2}}. \tag{B.71}$$

When we multiply Eq. B.70 by the multiplicative function, we get

$$\begin{aligned} 0 &= \rho\, e^{\frac{\rho^2}{2}} \frac{\mathrm{d}^2\Phi}{\mathrm{d}\rho^2} + (1+\rho^2) e^{\frac{\rho^2}{2}} \frac{\mathrm{d}\Phi}{\mathrm{d}\rho} + \left(2 + \tilde{k}^2\right) \rho\, e^{\frac{\rho^2}{2}} \Phi, \\ \frac{\mathrm{d}}{\mathrm{d}\rho}\left[\rho\, e^{\frac{\rho^2}{2}} \frac{\mathrm{d}\Phi}{\mathrm{d}\rho}\right] &+ 2\rho\, e^{\frac{\rho^2}{2}} \Phi = -\tilde{k}^2 \rho\, e^{\frac{\rho^2}{2}} \Phi. \end{aligned} \tag{B.72}$$

Eq. B.72 shows that unlike the 1D advection-diffusion telescoping model, the weighting function differs from the multiplicative function due to the inclusion of the prefactor ρ. In this case, the weighting function $w(\rho)$ as well as $p(\rho)$ and $q(\rho)$ are given as

$$w(\rho) = p(\rho) = q(\rho) = \rho\, e^{\frac{\rho^2}{2}}. \tag{B.73}$$

Furthermore, we observe that, as in the 1D case, the eigenvalues take the form $\tilde{k}^2$. Solutions of Φ from Eq. B.72 are obtained from Wolfram Alpha and take the form

$$\begin{aligned} \Phi_{\mathrm{ss}}(\rho) &= c_{\mathrm{ss}}\, e^{-\frac{\rho^2}{2}}, \\ \Phi_{\mathrm{dyn}}(\rho) &= c_1\, e^{-\frac{\rho^2}{2}}\, {}_1F_1\left(-\frac{\tilde{k}^2}{2}; 1; \frac{\rho^2}{2}\right) + c_2\, G^{2,0}_{1,2}\left(\frac{\rho^2}{2}\middle| \begin{matrix} -\frac{\tilde{k}^2}{2} \\ 0, 0 \end{matrix}\right), \end{aligned} \tag{B.74}$$

where $G^{m,n}_{p,q}\left(z\middle| \begin{matrix} a_1,\ldots,a_p \\ b1,\ldots,b_q \end{matrix}\right)$ is the Meijer G-function (we split up the eigenfunctions as dynamic and steady-state terms for now). We note here that the arguments of the Meijer G-function are such that the function diverges at the origin. As our system is defined as $0 \leq r \leq R$, we can say that $c_2 = 0$. Thus, our eigenfunctions are

$$\begin{aligned} \Phi_{\mathrm{ss}}(\rho) &= c_{\mathrm{ss}}\, e^{-\frac{\rho^2}{2}}, \\ \Phi_{\mathrm{dyn}}(\rho) &= c_1\, e^{-\frac{\rho^2}{2}}\, {}_1F_1\left(-\frac{\tilde{k}^2}{2}; 1; \frac{\rho^2}{2}\right), \end{aligned} \tag{B.75}$$

where we note that in the case of $\tilde{k} = 0$, we go from the dynamic eigenfunction to the static eigenfunction.

B.12.1 No-flux boundary condition

In the work presented here, there is no inflow or outflow of material at the boundary. Thus, we impose the boundary condition $\mathbf{J}\Big|_{r=R} = 0$. This means that

$$J_r\Big|_{r=R} = D\frac{\mathrm{d}\Phi}{\mathrm{d}r} - v(R)\Phi(R) = D\frac{\mathrm{d}\Phi}{\mathrm{d}r}\Big|_{r=R} + v_\mathrm{m}\Phi(R) = 0. \tag{B.76}$$

We know that Eq. B.76 is satisfied for the steady-state eigenfunction in the same way that the 1D steady-state solution satisfied the boundary condition. We then need to ensure that the boundary condition is satisfied for the dynamic eigenfunction. We start by taking the derivative of the eigenfunction:

$$\begin{aligned}\frac{\mathrm{d}\Phi}{\mathrm{d}r} &= -\frac{c_1\,\rho}{\lambda}\,e^{-\frac{\rho^2}{2}}\Big[\frac{\tilde{k}^2}{2}{}_1F_1\Big(1-\frac{\tilde{k}^2}{2};2;\frac{\rho^2}{2}\Big) + {}_1F_1\Big(-\frac{\tilde{k}^2}{2};1;\frac{\rho^2}{2}\Big)\Big],\\ \frac{\mathrm{d}\Phi}{\mathrm{d}r}\Big|_{r=R} &= -\frac{c_1\,v_\mathrm{m}}{D}\,e^{-\frac{v_\mathrm{m}R}{2D}}\Big[\Big(\frac{DRk^2}{2v_\mathrm{m}}\Big){}_1F_1\Big(1-\frac{DRk^2}{2v_\mathrm{m}};2;\frac{v_\mathrm{m}R}{2D}\Big) + {}_1F_1\Big(-\frac{DRk^2}{2v_\mathrm{m}};1;\frac{v_\mathrm{m}R}{2D}\Big)\Big].\end{aligned} \tag{B.77}$$

so when applied to the boundary condition, we get

$$\begin{aligned}D\frac{\mathrm{d}\Phi}{\mathrm{d}r}\Big|_{r=R} + v_\mathrm{m}\Phi(R) = &-c_1\,v_\mathrm{m}\,e^{-\frac{v_\mathrm{m}R}{2D}}\Big(\frac{DRk^2}{2v_\mathrm{m}}\Big){}_1F_1\Big(1-\frac{DRk^2}{2v_\mathrm{m}};2;\frac{v_\mathrm{m}R}{2D}\Big)\\ &- c_1\,v_\mathrm{m}\,e^{-\frac{v_\mathrm{m}R}{2D}}{}_1F_1\Big(-\frac{DRk^2}{2v_\mathrm{m}};1;\frac{v_\mathrm{m}R}{2D}\Big)\\ &+ c_1\,v_\mathrm{m}\,e^{-\frac{v_\mathrm{m}R}{2D}}{}_1F_1\Big(-\frac{DRk^2}{2v_\mathrm{m}};1;\frac{v_\mathrm{m}R}{2D}\Big).\end{aligned} \tag{B.78}$$

We are then left with the simplified equation:

$$\Big(\frac{DRk^2}{2v_\mathrm{m}}\Big){}_1F_1\Big(1-\frac{DRk^2}{2v_\mathrm{m}};2;\frac{v_\mathrm{m}R}{2D}\Big) = 0. \tag{B.79}$$

Here, $k = 0$ is satisfied, which yields the steady-state solution. Fig. B.10 shows the zeros when we set $\frac{R}{\lambda} = 3.16$. The first few non-zero eigenvalues are then $\tilde{k} = 0.474$, 0.759, 1.058, 1.354, and 1.672. Here, we observe a similar oscillator pattern to the zeros of the system. Once again, we see that there are multiple values of k that satisfy the boundary conditions. This means that the solution to the advection-diffusion problem once both boundary and initial conditions are satisfied, is a superposition of the different eigenfunctions:

$$c(r,t) = c_\mathrm{ss}\,e^{-\frac{v_\mathrm{m}r^2}{2DR}} + e^{-\frac{v_\mathrm{m}r^2}{2DR}}\sum_{i=1}^{\infty} c_i e^{-Dk_i^2 t}\,{}_1F_1\Big(-\frac{DRk_i^2}{2v_\mathrm{m}};1;\frac{v_\mathrm{m}r^2}{2DR}\Big). \tag{B.80}$$

We emphasize here the parallel between Eq. B.43 in the 1D case and Eq. B.80 in the 2D case. The primary difference between the two equations is the second argument in the Kummer confluent hypergeometric function. For simplicity, we will reintroduce the length scale $\lambda \equiv \sqrt{\frac{DR}{v_m}}$ so that the equation is simplified as

$$c(r,t) = c_{ss}\, e^{-\frac{r^2}{2\lambda^2}} + e^{-\frac{r^2}{2\lambda^2}} \sum_{i=1}^{\infty} c_i e^{-Dk_i^2 t}\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big). \tag{B.81}$$

B.12.2 Initial condition: uniform concentration

In the manuscript, we show three analytical solutions to the PDE with zero flux at the boundaries and each satisfying different initial conditions. We derive the three specific solutions in the following subsections. Here, we will tackle the uniform concentration initial conditions by examining the case where the concentration is 0 for $r \leq R_0$ and at uniform concentration c_0 for $r > R_0$. Once we have solved this general case, we will show the case where $R_0 = 0$.

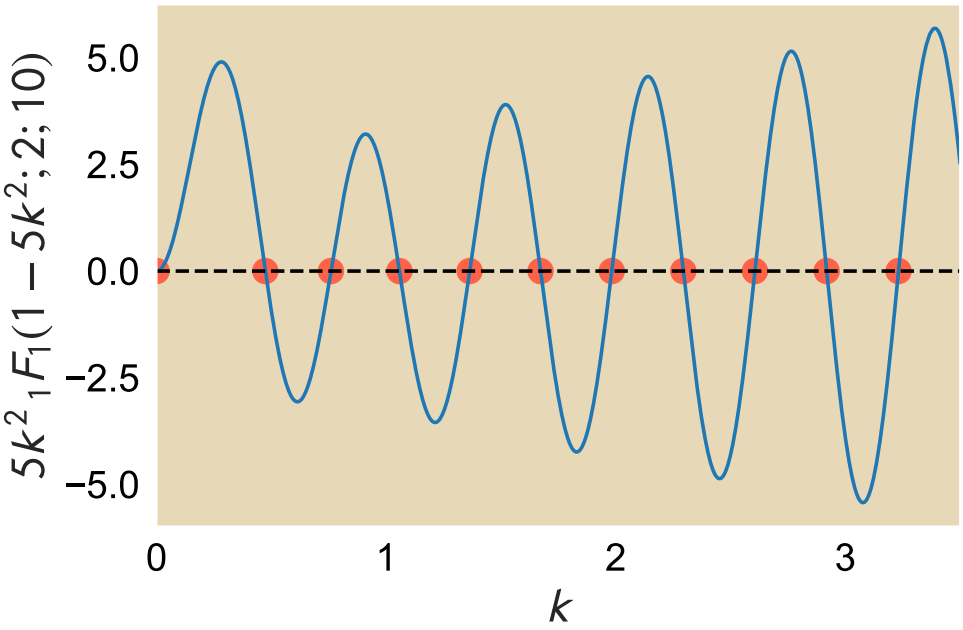


Figure B.10: **Zeros of k for $\frac{\lambda^2 k^2}{2}\,{}_1F_1\Big(1-\frac{\lambda^2 k^2}{2}; 2; \frac{R^2}{2\lambda^2}\Big) = 0$ where $\frac{R}{\lambda} = 3.16$.** Red dots are overlayed with the points where the Kummer confluent hypergeometric function crosses the x-axis.

The piecewise defined function then appears as

$$c(r,0) = \begin{cases} 0 & \text{if } r \le R_0, \\ c_0 & \text{if } r > R_0. \end{cases} \tag{B.82}$$

At $t = 0$, our equation looks like

$$c(r,0) = c_{\text{ss}}\, e^{-\frac{r^2}{2\lambda^2}} + e^{-\frac{r^2}{2\lambda^2}} \sum_{i=1}^{\infty} c_i \, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big), \tag{B.83}$$

We multiply both sides by the weighting function $w(r) = re^{\frac{r^2}{2\lambda^2}}$ and an eigenfunction of the differential equation $\Phi_h(r) = e^{-\frac{r^2}{2\lambda^2}} \, {}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)$ for identifying the coefficients of the non-steady state terms or $\Phi_{\text{ss}}(r) = e^{-\frac{r^2}{2\lambda^2}}$ for determining the steady-state term. For the steady-state term, we have

$$\begin{aligned}
c_0 \int_{R_0}^{R} r \, \mathrm{d}r &= c_{\text{ss}} \int_0^R r\, e^{-\frac{r^2}{2\lambda^2}} \mathrm{d}r + \sum_{i=1}^{\infty} c_i \int_0^R r\, e^{-\frac{r^2}{2\lambda^2}} \, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)\mathrm{d}r, \\
c_0 \frac{r^2}{2}\Big|_{R_0}^{R} &= -c_{\text{ss}}\lambda^2 e^{-\frac{r^2}{2\lambda^2}}\Big|_0^R + \sum_{i=1}^{\infty} c_i \int_0^R r \, {}_1F_1\Big(1 + \frac{\lambda^2 k_i^2}{2}; 1; -\frac{r^2}{2\lambda^2}\Big)\mathrm{d}r, \\
c_0\Big(\frac{R^2}{2\lambda^2} - \frac{R_0^2}{2\lambda^2}\Big) &= c_{\text{ss}}\Big(1 - e^{-\frac{R^2}{2\lambda^2}}\Big) + \sum_{i=1}^{\infty} \frac{c_i}{\lambda^2} \int_0^R \sum_{j=0}^{\infty} \mathrm{d}r\, r \, \frac{\Big(1 + \frac{\lambda^2 k_i^2}{2}\Big)_j}{(1)_j} \frac{\Big(-\frac{r^2}{2\lambda^2}\Big)^j}{j!}, \\
c_0\Big(\frac{R^2}{2\lambda^2} - \frac{R_0^2}{2\lambda^2}\Big) &= c_{\text{ss}}\Big(1 - e^{-\frac{R^2}{2\lambda^2}}\Big) - \sum_{i=1}^{\infty} c_i \sum_{j=0}^{\infty} \frac{\Big(1 + \frac{\lambda^2 k_i^2}{2}\Big)_j}{(1)_j} \frac{\Big(-\frac{r^2}{2\lambda^2}\Big)^{j+1}}{(j+1)!}\Big|_0^R.
\end{aligned} \tag{B.84}$$

We use the fact that $(1)_j = j!$ and $(j+1)! = (2)_j$ so

$$\begin{aligned}
c_0\Big(\frac{R^2}{2\lambda^2} - \frac{R_0^2}{2\lambda^2}\Big) &= c_{\text{ss}}\Big(1 - e^{-\frac{R^2}{2\lambda^2}}\Big) + \sum_{i=1}^{\infty} c_i \frac{r^2}{2\lambda^2} \sum_{j=0}^{\infty} \frac{\Big(1 + \frac{\lambda^2 k_i^2}{2}\Big)_j}{(2)_j} \frac{\Big(-\frac{r^2}{2\lambda^2}\Big)^j}{j!}\Big|_0^R, \\
c_0\Big(\frac{R^2}{2\lambda^2} - \frac{R_0^2}{2\lambda^2}\Big) &= c_{\text{ss}}\Big(1 - e^{-\frac{R^2}{2\lambda^2}}\Big) + \sum_{i=1}^{\infty} c_i \frac{r^2}{2\lambda^2} \, {}_1F_1\Big(1 + \frac{\lambda^2 k_i^2}{2}; 2; -\frac{r^2}{2\lambda^2}\Big)\Big|_0^R, \\
c_0\Big(\frac{R^2}{2\lambda^2} - \frac{R_0^2}{2\lambda^2}\Big) &= c_{\text{ss}}\Big(1 - e^{-\frac{R^2}{2\lambda^2}}\Big) + \sum_{i=1}^{\infty} c_i \frac{r^2}{2\lambda^2} e^{-\frac{r^2}{2\lambda^2}} \, {}_1F_1\Big(1 - \frac{\lambda^2 k_i^2}{2}; 2; \frac{r^2}{2\lambda^2}\Big)\Big|_0^R, \\
c_{\text{ss}} &= \frac{c_0}{2} \frac{\frac{R^2}{\lambda^2} - \frac{R_0^2}{\lambda^2}}{1 - e^{-\frac{R^2}{2\lambda^2}}},
\end{aligned} \tag{B.85}$$

where we use Eq. B.79 to remove the upper bound of the integral involving the hypergeometric function. We now find the coefficients for the non-steady state terms. We do so by multiplying both sides by $\Phi_h(r) = e^{-\frac{r^2}{2\lambda^2}}\,{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)$ instead,

$$\begin{aligned} c_0 \int_{R_0}^{R} r\,{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)\mathrm{d}r = c_{\mathrm{ss}} \int_0^R r e^{-\frac{r^2}{2\lambda^2}}\,{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)\mathrm{d}r \\ + \Bigg[\sum_{i=1}^{\infty} c_i \int_0^R r\, e^{-\frac{r^2}{2\lambda^2}}\,{}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big) \\ \times\,{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)\mathrm{d}r\Bigg]. \end{aligned} \tag{B.86}$$

Fortunately, we have already done the first integral on the right-hand side of the equation, so we only have to take care of the integral on the left-hand side. We further argue that by the Sturm-Liouville theory all of the integrals in the summation vanish except in the case where $i = h$. The equation then boils down to

$$\begin{aligned} c_0 \int_{R_0}^{R} r\,{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)\mathrm{d}r &= c_h \int_0^R r\, e^{-\frac{r^2}{2\lambda^2}} \Big[{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)\Big]^2 \mathrm{d}r, \\ c_0 \int_{R_0}^{R} \sum_{j=0}^{\infty} r \frac{\Big(-\frac{\lambda^2 k_h^2}{2}\Big)_j}{(1)_j} \frac{\Big(\frac{r^2}{2\lambda^2}\Big)^j}{j!}\mathrm{d}r &= c_h \int_0^R r\, e^{-\frac{r^2}{2\lambda^2}} \Big[{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)\Big]^2 \mathrm{d}r, \\ c_0 \sum_{j=0}^{\infty} \lambda^2 \frac{\Big(-\frac{\lambda^2 k_h^2}{2}\Big)_j}{(1)_j} \frac{\Big(\frac{r^2}{2\lambda^2}\Big)^{j+1}}{(j+1)!}\Bigg|_{R_0}^{R} &= c_h \int_0^R r\, e^{-\frac{r^2}{2\lambda^2}} \Big[{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)\Big]^2 \mathrm{d}r, \\ c_0 \frac{r^2}{2}\,{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 2; \frac{r^2}{2\lambda^2}\Big)\Big|_{R_0}^{R} &= c_h \int_0^R r\, e^{-\frac{r^2}{2\lambda^2}} \Big[{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)\Big]^2 \mathrm{d}r, \\ c_h &= \frac{c_0}{2} \frac{R^2\,{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 2; \frac{R^2}{2\lambda^2}\Big) - R_0^2\,{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 2; \frac{R_0^2}{2\lambda^2}\Big)}{\int_0^R r\, e^{-\frac{r^2}{2\lambda^2}} \Big[{}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)\Big]^2 \mathrm{d}r}, \end{aligned} \tag{B.87}$$

where we numerically integrate the denominator. When assembled together, the solution comes out to

$$c(r,t) = \frac{c_0}{2}\, e^{-\frac{r^2}{2\lambda^2}} \Bigg\{ \frac{\frac{R^2}{\lambda^2} - \frac{R_0^2}{\lambda^2}}{1 - e^{-\frac{R^2}{2\lambda^2}}} + \sum_{i=1}^{\infty} \frac{R^2\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 2; \frac{R^2}{2\lambda^2}\Big) - R_0^2\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 2; \frac{R_0^2}{2\lambda^2}\Big)}{\int_0^R r'\, e^{-\frac{r'^2}{2\lambda^2}} \Big[{}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r'^2}{2\lambda^2}\Big)\Big]^2 \mathrm{d}r'} \times e^{-Dk_i^2 t}\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big) \Bigg\}. \tag{B.88}$$

In the case where $R_0 = 0$, the solution for uniform concentration throughout the system is instead

$$c(r,t) = \frac{c_0}{2}\, e^{-\frac{r^2}{2\lambda^2}} \Bigg\{ \frac{\frac{R^2}{\lambda^2}}{1 - e^{-\frac{R^2}{2\lambda^2}}} + \sum_{i=1}^{\infty} \frac{R^2\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 2; \frac{R^2}{2\lambda^2}\Big)}{\int_0^R r'\, e^{-\frac{r'^2}{2\lambda^2}} \Big[{}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r'^2}{2\lambda^2}\Big)\Big]^2 \mathrm{d}r'} \times e^{-Dk_i^2 t}\, {}_1F_1\Big(-\frac{\lambda^2 k_i^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big) \Bigg\}, \tag{B.89}$$

as shown in the manuscript.

B.12.3 Initial condition: Gaussian concentration for $r > R_0$

We finish the 2D advection-diffusion model with the initial condition of a Gaussian concentration profile outside of a region $r > R_0$ and 0 within that region. Written explicitly, the initial condition is

$$c(r,0) = \begin{cases} 0 & \text{if } r \le R_0, \\ c_0\, e^{-\frac{r^2}{2\lambda^2}} & \text{if } r > R_0. \end{cases} \tag{B.90}$$

We apply the same situation where we multiply both sides by the weighting function $w(r) = re^{\frac{r^2}{2\lambda^2}}$ and an eigenfunction of the differential equation $\Phi_h(r) = e^{-\frac{r^2}{2\lambda^2}}\, {}_1F_1\Big(-\frac{\lambda^2 k_h^2}{2}; 1; \frac{r^2}{2\lambda^2}\Big)$ for identifying the coefficients of the non-steady state terms or $\Phi_{\text{ss}}(r) = e^{-\frac{r^2}{2\lambda^2}}$ for determining the steady-state term. Relying on integrals performed in Subsec. B.12.2, we start with identifying the coefficients

of the dynamic terms c_i,

$$\begin{aligned}
\int_0^R c(r,0)w(r)\Phi_h(r)\mathrm{d}r &= c_{\mathrm{ss}}\int_0^R e^{-\frac{r^2}{2\lambda^2}}w(r)\Phi_h(r)\mathrm{d}r \\
&\quad+\sum_{i=1}^{\infty}c_i\int_0^R e^{-\frac{r^2}{2\lambda^2}}\,{}_1F_1\Big(-\frac{\lambda^2k_i^2}{2};1;\frac{r^2}{2\lambda^2}\Big) \\
&\quad\times w(r)\Phi_h(r)\mathrm{d}r, \\
c_0\int_{R_0}^R r\,e^{-\frac{r^2}{2\lambda^2}}\,{}_1F_1\Big(-\frac{\lambda^2k_h^2}{2};1:\frac{r^2}{2\lambda^2}\Big)\mathrm{d}r &= c_{\mathrm{ss}}\int_0^R r\,e^{-\frac{r^2}{2\lambda^2}}\,{}_1F_1\Big(-\frac{\lambda^2k_h^2}{2};1;\frac{r^2}{2\lambda^2}\Big)\mathrm{d}r \\
&\quad+\sum_{i=1}^{\infty}c_i\int_0^R w(r)\Phi_i(r)\Phi_h(r)\mathrm{d}r, \\
c_0\frac{r^2}{2}e^{-\frac{r^2}{2\lambda^2}}\,{}_1F_1\Big(1-\frac{\lambda^2k_h^2}{2};2;\frac{r^2}{2\lambda^2}\Big)\Big|_{R_0}^R &= c_h\int_0^R r\,e^{-\frac{r^2}{2\lambda^2}}\Big[{}_1F_1\Big(-\frac{\lambda^2k_h^2}{2};1;\frac{r^2}{2\lambda^2}\Big)\Big]^2\mathrm{d}r, \\
-c_0\frac{R_0^2}{2}e^{-\frac{R_0^2}{2\lambda^2}}\,{}_1F_1\Big(1-\frac{\lambda^2k_h^2}{2};2;\frac{R_0^2}{2\lambda^2}\Big) &= c_h\int_0^R r\,e^{-\frac{r^2}{2\lambda^2}}\Big[{}_1F_1\Big(-\frac{\lambda^2k_h^2}{2};1;\frac{r^2}{2\lambda^2}\Big)\Big]^2\mathrm{d}r, \\
c_h &= -c_0\frac{\frac{R_0^2}{2}e^{-\frac{R_0^2}{2\lambda^2}}\,{}_1F_1\Big(1-\frac{\lambda^2k_h^2}{2};2;\frac{R_0^2}{2\lambda^2}\Big)}{\int_0^R r\,e^{-\frac{r^2}{2\lambda^2}}\Big[{}_1F_1\Big(-\frac{\lambda^2k_h^2}{2};1;\frac{r^2}{2\lambda^2}\Big)\Big]^2\mathrm{d}r},
\end{aligned} \tag{B.91}$$

where we removed the terms $i \neq h$ as shown from Eq. B.66. To determine the initial conditions of the steady-state coefficient term, we would instead multiply by the weighting function and the steady-state eigenfunction $e^{-\frac{r^2}{2}}$ to yield

$$\begin{aligned}
c_0\int_{R_0}^R r\,e^{-\frac{r^2}{2\lambda^2}}\mathrm{d}r &= c_{\mathrm{ss}}\int_0^R r\,e^{-\frac{r^2}{2\lambda^2}}\mathrm{d}r+\sum_{i=1}^{\infty}c_i\int_0^R r\,e^{-\frac{r^2}{2\lambda^2}}\,{}_1F_1\Big(-\frac{\lambda^2k_i^2}{2};1;\frac{r^2}{2\lambda^2}\Big)\mathrm{d}r, \\
-c_0\,e^{-\frac{r^2}{2\lambda^2}}\Big|_{r=R_0}^R &= -c_{\mathrm{ss}}\,e^{-\frac{r^2}{2\lambda^2}}\Big|_{r=0}^R, \\
c_{\mathrm{ss}} &= c_0\frac{e^{-\frac{R_0^2}{2\lambda^2}}-e^{-\frac{R^2}{2\lambda^2}}}{1-e^{-\frac{R^2}{2\lambda^2}}}.
\end{aligned} \tag{B.92}$$

When all is assembled, the solution with the no-flux boundary conditions and FRAPed initial condition yields

$$c(r,t) = c_0\, e^{-\frac{v_{\mathrm{m}} r^2}{2DR}} \Bigg\{ \frac{e^{-\frac{v_{\mathrm{m}} R_0^2}{2DR}} - e^{-\frac{v_{\mathrm{m}} R}{2D}}}{1 - e^{-\frac{v_{\mathrm{m}} R}{2D}}} - \sum_{i=1}^{\infty} \frac{\frac{R_0^2}{2} e^{-Dk_i^2 t}\, {}_1F_1\Big(1 + \frac{DRk_i^2}{2v_{\mathrm{m}}}; 2; -\frac{v_{\mathrm{m}} R_0^2}{2DR}\Big)}{\int_0^R r\, e^{-\frac{v_{\mathrm{m}} r^2}{2DR}} \Big[{}_1F_1\Big(-\frac{DRk_i^2}{2v_{\mathrm{m}}}; 1; \frac{v_{\mathrm{m}} r^2}{2DR}\Big)\Big]^2 \mathrm{d}r}\, {}_1F_1\Big(-\frac{DRk_i^2}{2v_{\mathrm{m}}}; 1; \frac{v_{\mathrm{m}} r^2}{2DR}\Big) \Bigg\}. \tag{B.93}$$

B.13 Numerically solving advection-diffusion equations with COMSOL

COMSOL Multiphysics® simulations are constructed with consideration of four particular details in mind: design of the geometry, set-up of the differential equations, incorporation of images as initial conditions, and sweeping through parameters. A discussion of the mesh is discussed in Sec. B.14.

B.13.1 Geometry

Because simulations would be performed using images as initial conditions, and because the microtubule network has a roughly circular geometry, we designed a circle geometry in COMSOL where the radius was a parameter based upon the photobleach dataset used. This could range from as small as 70 µm for the networks nearing the end of contraction and upwards of 250 µm which sets the initial activation size for the experiments.

B.13.2 Setting up the differential equations

Although there are multiple partial differential equation forms in COMSOL that can be used for the advection-diffusion equation studied here, we elect to use the coefficient form PDE and define our variable of interest as u with units of mol/m^3 and a source term units of mol/(m^3·s). Although our past derivations use the variable c, we use u in the differential equation due to the occurrence of the coefficient c in the coefficient form PDE in COMSOL. We note that the coefficient form PDE as shown in COMSOL is of the form

$$e_a \frac{\partial^2 u}{\partial t^2} + d_a \frac{\partial u}{\partial t} + \nabla \cdot (-c\nabla u - \alpha u + \gamma) + \beta \cdot \nabla u + au = f, \tag{B.94}$$

where e_a, d_a, c, a, and f are scalar coefficients while α, γ, and β are vectors. We note that since our advection-diffusion (using u for concentration here) is of the form

$$\frac{\partial u}{\partial t} = D\nabla^2 u + \frac{v_{\mathrm{m}}}{R} \nabla \cdot (\mathbf{r} u)\,, \tag{B.95}$$

if we rewrite the equation a little bit to match the form of Eq. B.94, we get

$$\frac{\partial u}{\partial t} + \nabla \cdot \left(-D\nabla u - \frac{v_{\mathrm{m}}}{R}\mathbf{r}u \right) = 0. \tag{B.96}$$

We can see here that to make Eq. B.96 match Eq. B.94, then e_a, a, all of the elements of γ, all of the elements of β, and f are all 0 while

$$d_a = 1 \text{ s}^{-1}, \tag{B.97}$$

$$c = D, \tag{B.98}$$

$$\alpha = \begin{bmatrix} \frac{v_{\mathrm{m}}}{R}x \\ \frac{v_{\mathrm{m}}}{R}y \end{bmatrix}, \tag{B.99}$$

where we note that we define D to take on dimensions of length^2/time and $\frac{v_{\mathrm{m}}}{R}$ to have units of time^{-1} in COMSOL.

In our experiments, we were careful to ensure that there was negligible to no detectable amount of microtubules flowing from outside of the light-activated region into network. We similarly impose a no-flux boundary condition by using the Zero Flux boundary condition option in COMSOL.

B.13.3 Incorporation of images as initial conditions

One of the conveniences of using COMSOL is the ability to use experimental data as part of the simulations. Here, we elected to use the first photobleached frame as our initial condition for our images. Before doing so, we took our image of interest and imposed a small Gaussian smoothing ($\sigma = 1$pixel) to gently smooth out the microtubule concentration field before renormalizing the image and returning it into an 8-bit image ('uint8'). We then exported the image as a TIFF file. Within COMSOL, under our Component $\rightarrow$ Definitions branch of the simulation, we defined an Image Function and gave it the notation u_im. Within the image function, we selected the image of interest under the Browse option. We then need to line up the image such that the center of the contracting microtubule network is at the origin or else the advection-diffusion model will drive the advective contraction toward a different part of the network. To do so, we specify the coordinates where we determine the minimum and maximum x and y values based on the pixel-weighted center of the network as discussed in Appendix B.2 and dimensions of the image. We then import the image and verify the image was the one we wanted by plotting.

When incorporating the image in COMSOL, under the Coefficient Form PDE node in the Initial Conditions tab, we set the initial time derivative of u to 0 while the initial value for u is set as $u\_im(x, y)$, where (x, y) specifies the spatial dimensions of the image. Fortunately, by the geometry we specify, we will not pick up any parts of the image outside of the region of interest. We further multiply this function by a coefficient such that we obtain roughly the correct units of concentration as required.

B.13.4 Parameter sweep

To perform the parameter sweep, we include the Parametric Sweep option in the Study section of the simulation and define the parameters of interest under Global Definitions → Parameters. Within the parameters, we specify the parameters D for our diffusion constant and *alpha* for our contraction rate, which replaces $\frac{v_m}{R}$ in the equations above, including our definition of α in Eq. B.99. Under the Parametric Sweep, we can then chose D and *alpha* as our parameters to be swept. By selected our range of *alpha* to be 0.0016 to 0.0024 s^{-1} in increments of 0.0002 s^{-1} while D ranged from 0.05 to 0.2 $\mu m^2/s$ in increments of 0.05 $\mu m^2/s$. All possible combinations of D and *alpha* were permitted for the simulations.

B.14 Gibbs phenomenon in analytical solutions and mesh granularity in FEM

A common observation found for many of the analytical solutions is the disagreement between the analytical solution at $t = 0$ and the defined initial condition that the solution is intended to recapitulate. As shown in Fig. 1 of the main manuscript, the analytical solution, which is composed of twelve nonzero eigenvalues and the steady-state function, creates oscillations about the intended initial condition. This disagreement is a demonstration of the Gibbs phenomenon, as famously revealed by the imperfect decomposition of a square wave into a sum of sinusoidal functions. Fig. B.12 demonstrates the evolution of each of the three analytical solutions examined in the main manuscript when more eigenvalues are included in the solution. Specifically, for $c(r, 0) = c_0$ (Fig. B.12A), $c(r > R_0, 0) = c_0$ (Fig. B.12B), and $c(r > R_0, 0) = c_0 \exp(-r^2/2\lambda^2)$ (Fig. B.12C), all of which are represented by dashed black lines, more eigenvalues reduce the level of error between the analytical solution and the initial condition. For the two initial conditions involving a uniform concentration, the

use of one eigenvalue in addition to the steady-state solution (purple line) leads to a large negative concentration at $r = 0$ but begins to better recapitulate the initial conditions by the addition of 12 non-zero eigenvalues. Deviations from the initial condition decrease dramatically by that point. This is further observed for the clipped Gaussian distribution: while the Gaussian tail is quantitatively captured by the the addition of only a few eigenvalues, the analytical solution begins to better recapitulate the concentration profile about $r = R_0$ with the addition of more terms in the solution. Nevertheless, even after using twelve eigenvalues, the solution shows small oscillations about the exact initial condition and is a continued feature with the addition of more eigenvalues.

The deviations in the constructed solutions from the true values are also apparent in finite element methods through the choice of granularity in the mesh. As FEM involves solving the governing equation over a particular domain, having a very fine grained mesh allows for the FEM solution to more accurately reflect the true solution to the problem at the cost of computational time. On the other hand, a very coarse-grained mesh involves less computing power to solve the original equations but may coarse grain away details smaller than the element size, requiring a balance between accurately solving the original PDE(s) and computational efficiency.

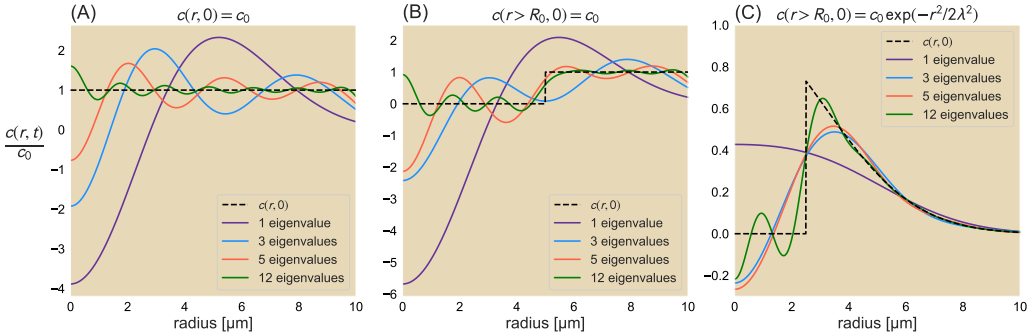


Figure B.11: **Gibbs phenomenon for analytical solutions.** Concentration profiles of the analytical solution for the initial conditions (A) $c(r,0) = c_0$, (B) $c(r > R_0, 0) = c_0$, and (C) $c(r > R_0, 0) = c_0 \exp(-r^2/2\lambda^2)$ with the steady-state solution and the first nonzero eigenvalue solution (purple line), the first three nonzero eigenvalue solutions (blue), the first five terms (red), and the first twelve terms (green). The intended initial conditions are represented as dashed black lines.

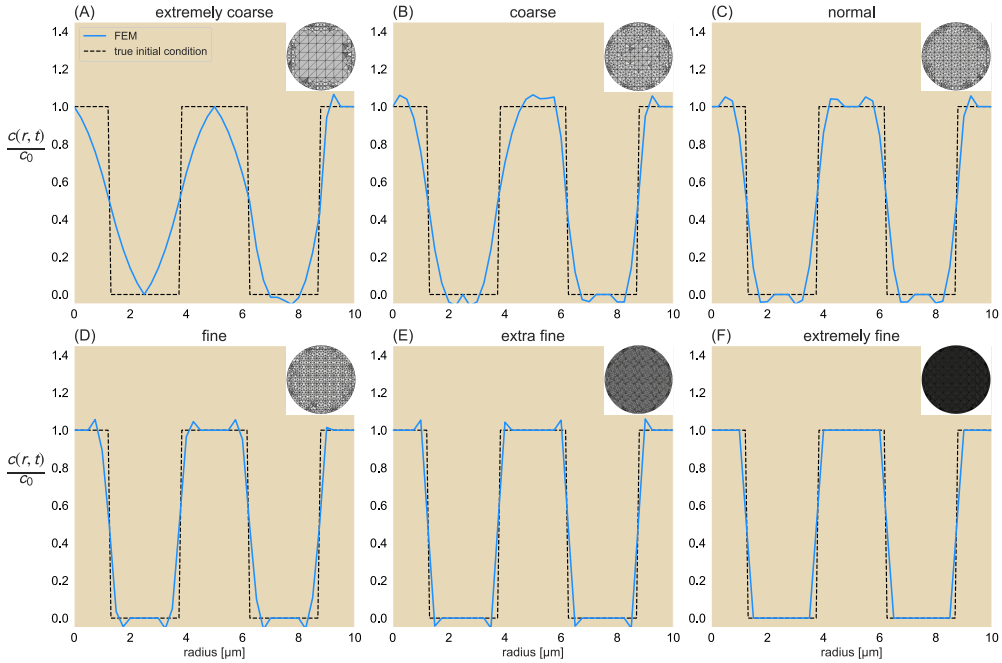


Figure B.12: **Effects of mesh granularity on FEM solution.** Concentration profiles at $t = 0$ for six different element sizes as defined by the COMSOL Multiphysics physics-controlled mesh: (A) extremely coarse, (B) coarse, (C) normal, (D) fine, (E) extra fine, and (F) extremely fine. Finite elements output is represented by the blues lines while the true initial conditions are given as the black dashed lines. For visualization purposes, the appearance of the meshes used for the defined geometry are shown as insets in the upper right-hand corner of the respective subfigures. Concentration profile is from a line trace along the horizontal axis from the origin of the geometry to the boundary.

Fig. B.12 shows how the granularity of the mesh affects the FEM solutions. We compare the concentration profiles produced by FEM (solid blue lines) against the true initial condition (dashed black lines) for six different element sizes as found in the physics-controlled mesh feature in COMSOL Multiphysics: (A) extremely coarse, (B) coarse, (C) normal, (D) fine, (E) extra fine, and (F) extremely fine. We see that using the most coarse-grained feature produces a more sinusoidal shape of matching frequency and amplitude to the square wave pattern of the initial condition. However, with successive decreases in element size (increase in mesh fineness) the FEM solution more closely reflect the initial condition. Fig. B.12B-E show that increase the mesh fineness leaves fewer deviations from the true values, largely located near the discontinuities in the profile. The insets in the upper right of each figure shows the mesh pattern for the geometry for the study. As Fig. B.12F shows, while the extremely fine

mesh does not overshoot above the c_0 values or undershoot the $c(r,0) = 0$ regions, the finite size of the elements in the mesh causes the discontinuous region to take on a value between the two regions instead.

B.15 Parameter sweeping and Péclet numbers

In this section, we ask how changes in the diffusion constant D and contraction velocity v_m are reflected in the grid patterned advection-diffusion model. This interplay reveals itself by transforming Eq. B.19 into dimensionless form. Suppose instead of v_max we wrote that out as a function of the speed of individual motors which move along and move microtubules. We noted that the maximum velocity occurred at the outer edge of the activation circle. Assuming a telescoping model where a filament network contracts due to a series of alternating filaments and motors connecting them, we start by treating the maximum velocity as the speed of the motors multiplied by the minimum number of filaments required to connect the origin to the outer edge of the activation zone. This is simply a case of filaments being serially aligned at their ends. This scheme then means that for a filament of average length L and activation circle of radius R

$$v_\mathrm{m} = v_L \frac{R}{L} \tag{B.100}$$

where v_L is a natural velocity scale. If we further redefine some variables to make them dimensionless, such as $x \to L\tilde{x}$ and $t \to \frac{L}{v_L}\tilde{t}$, we can alter Eq. B.68 to

$$\begin{aligned}
\frac{\partial c}{\partial t} &= D\nabla^2 c - \nabla\cdot(\mathbf{v}c)\,, \\
&= D\nabla^2 c + \frac{v_\mathrm{m}}{R}\nabla\cdot(\mathbf{r}c)\,, \\
\frac{v_L}{L}\frac{\partial c}{\partial \tilde{t}} &= \frac{D}{L^2}\tilde{\nabla}^2 c + \frac{v_L}{L}\tilde{\nabla}\cdot(\tilde{\mathbf{r}}c), \\
\frac{\partial c}{\partial \tilde{t}} &= \frac{D}{v_L L}\tilde{\nabla}^2 c + \tilde{\nabla}\cdot(\tilde{\mathbf{r}}c)\,, \\
\frac{\partial c}{\partial \tilde{t}} &= \frac{1}{\mathrm{Pe}}\tilde{\nabla}^2 c + \tilde{\nabla}\cdot(\tilde{\mathbf{r}}c)\,,
\end{aligned} \tag{B.101}$$

where Pe is the Péclet number:

$$\mathrm{Pe} \equiv \frac{v_L L}{D}. \tag{B.102}$$

This dimensionless parameter tells us how the contraction speed of a connected network and the diffusion constant dictate whether the contraction process

or diffusion process dominates. For fixed length L such as the length of a microtubule, increasing Péclet number tells us that the advection is dominant and thus Eq. B.101 is largely the advective term, while smaller values of Pe tell us that diffusion is the dominant term.

Amusingly, had we defined the natural time variable to be $\tilde{t} = \frac{D}{L^2}t$, then Eq. B.101 would be modified as

$$\frac{\partial c}{\partial \tilde{t}} = \tilde{\nabla}^2 c + \mathrm{Pe}\tilde{\nabla} \cdot (\tilde{\mathbf{r}}c), \tag{B.103}$$

Pe illustrates the relationship between the advection in the system and the diffusion. For the parameters used for Fig. 4, if we take the characteristic length scale to be on the order 1, roughly the length of the microtubule in our experiments (see Appendix B.6 then $v_{\mathrm{L}} = 0.01\ \frac{\mu\mathrm{m}}{\mathrm{s}}$ and we obtain a Péclet number of 0.1. As this value is much smaller than unity, we see that the diffusion term dominates over the short timescale.

To further demonstrate the tradeoffs between advection and diffusion, we examined the redistribution of the concentration with the same gridlike pattern for different Péclet numbers. To do so, we kept v_{m} fixed and varied D for a set of simulations while for another set of FEM studies we kept D fixed while changing v_{m}. Fig. B.13 shows the concentration along the x-axis that extends from the origin to the boundary at $r = 10$ µm and as depicted by the purple line in the $t = 0$ plot in Fig. 4A. Fig. B.13A looks at a time series of the concentration profile for different diffusion constants while v_{m} is fixed at 0.1 $\frac{\mu\mathrm{m}}{\mathrm{s}}$ while Fig. B.13B shows the concentration profile for different v_{m} with D kept constant at 0.1 $\frac{\mu\mathrm{m}^2}{\mathrm{s}}$. Using the purple line in Fig. B.13A as the original parameter combination used in Fig. 4, we see that increasing the diffusion constant (green and blue) causes the individual squares of initial concentration c_0 to quickly disperse to create a more uniform concentration before the advection creates the Gaussian steady-state profile (and ones with longer standard deviations than the original parameters). This observation makes sense as the Péclet number gets lower and lower with increases in D, causing the diffusion term to dominate more than the advection. This increase in diffusion further illustrates the wider Gaussian distribution obtained at steady state, as the length scale λ depends on the square root of D and inversely on the advection speed v_{m}.

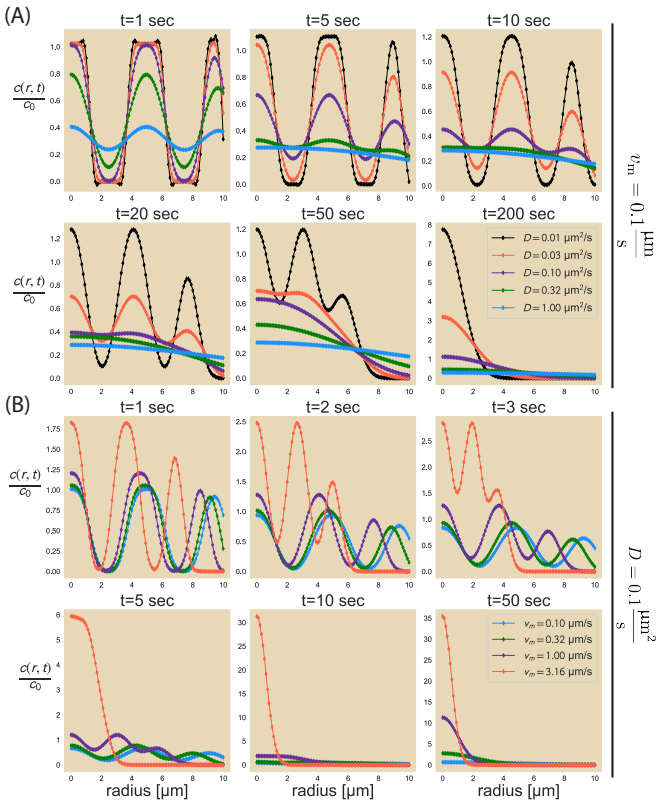


Figure B.13: **Traces of concentration as a function of radius for various combinations of diffusion constant D and velocity $v_{\mathbf{m}}$.** (A) Concentration profiles for D spanning three orders of magnitude with the velocity fixed at $v_{\mathrm{m}} = 0.1\,\frac{\mu\mathrm{m}}{\mathrm{s}}$. As shown in the yellow boxes for the heatmap for $t = 0$ sec, the concentration throughout the black gridlines is 0 while the concentration in the white cells is constant c_0. (B) Concentration profiles for varying v_{m} with the diffusion constant fixed at $D = 0.1\,\frac{\mu\mathrm{m}^2}{\mathrm{s}}$ Traces of all concentration profiles are obtained from a 1D slice along the x-axis from the origin of the circle to the boundary at $R = 10$ μm as shown by the purple line in Fig. 4A.

In contrast, decreasing the diffusion constant which increases the Péclet number preserves the oscillatory pattern of the concentration profile as the advection pushes the material toward the origin. We see that by $t = 50$ sec, the red and black curves that denote $D = 0.03\,\frac{\mu\mathrm{m}^2}{\mathrm{s}}$ (Pe $= 0.3$) and $D = 0.01\,\frac{\mu\mathrm{m}^2}{\mathrm{s}}$ (Pe $= 1.0$), respectively, still exhibit wave-like shapes at the $t = 50$ sec mark. As a result of the reduced diffusive effects, the concentration at the center is much higher and falls off much more quickly as the length scale λ is shorter (1.7 μm for the red curve and 1 μm for the black curve).

Tuning the advection for fixed diffusion constant as shown in Fig. B.13B similarly demonstrates the competition between diffusion attempting to level out

the concentration profile and advective flow trying to concentrate molecules toward the center. With the purple line corresponding with Pe = 1, we see that decreasing advection and lowering Pe, as demonstrated earlier, causes the differences between local minima and maxima in concentration to decrease faster than the minima and maxima move toward the origin. On the other hand, by increasing v_{m} and thus increasing Pe to make advection more dominant (red line of Fig. B.13B), we see that the advective flow causes the minima and maxima to be pushed toward the origin in less than 5 sec and create a sharper Gaussian peak (a discussion about the jagged profile for the $v_{\mathrm{m}} = 3.16\,\frac{\mu\mathrm{m}}{\mathrm{s}}$ plot can be founded in the Appendix on the Gibbs phenomenon and FEM mesh setting). Taken together, when diffusion dominates Pe < 1 the concentration of molecules tends toward a more uniform behavior before advection pushes them to the origin, causing peaks and troughs in the concentration to disperse and become indistinguishable. On the other hand, when advection dominates Pe > 1 peaks and troughs move toward the origin faster than they disperse, and lead to more tightly distributed Gaussian steady-state profiles.

BIBLIOGRAPHY

[1] T. D. Ross et al., "Controlling organization and forces in active matter through optically defined boundaries", Nature 572 (2019), 224.

[2] N. Georgoulia, *Tubulin polymerization with GTP/GMPCPP/Taxol*, 2012.

[3] S. J. DeCamp, "*Dogic lab - acrylamide coating protocol.*", (2016).

[4] G. W. Zack, W. E. Rogers, and S. A. Latt, "Automatic measurement of sister chromatid exchange frequency", Journal of Histochemistry and Cytochemistry 25 (1977).

[5] B. Carpenter et al., "Stan: A probabilistic programming language", J Stat Softw 76 (2017).

[6] W. Niblack, *An Introduction to Digital Image Processing*, Prentice-Hall, 1986.

[7] R. A. Banks, V. Galstyan, H. J. Lee, S. Hirokawa, A. Ierokomos, T. Ross, Z. Bryant, M. Thomson, and R. Phillips, "Motor processivity and speed determine structure and dynamics of motor-microtubule assemblies", bioRxiv (under review at eLife) (2021).

[8] N. F. Endres et al., "A lever-arm rotation drives motility of the minus-end-directed kinesin Ncd", Nature 439 (2006), 875.

[9] G. Guntas et al., "Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins", Proc Natl Acad Sci 112 (2015), 112.