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

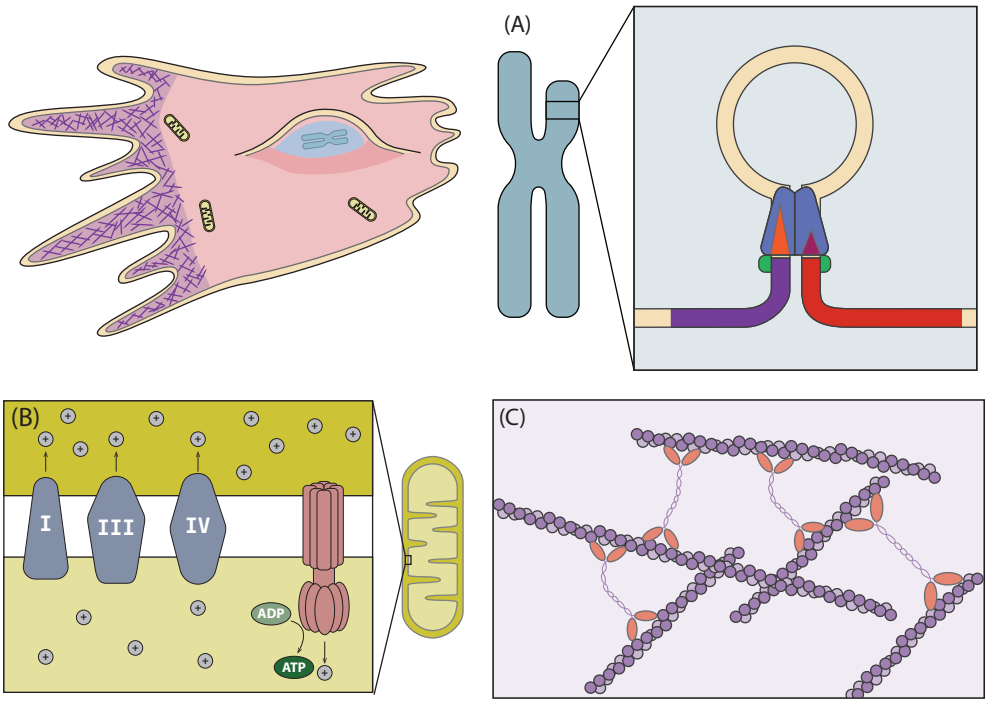

Figure 1.1: **Examples of proteins with coupling functions.** Cells utilize a range of proteins with coupling functions. (A) Some proteins bind to DNA at two sites in order to regulate or even aid in creating genes. (B) Proteins such as those found in the electron transport chain couple two environments 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

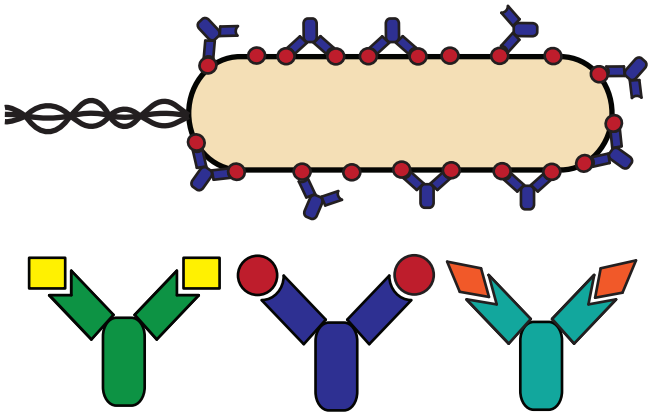

Figure 1.2: **Antibodies allow for the immune cells to identify invasive agents with greater specificity.** Antibodies differ from one another in their ability to bind different small protein markers called antigens on the surfaces of bacteria or infected cells. This identification allows other cells 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

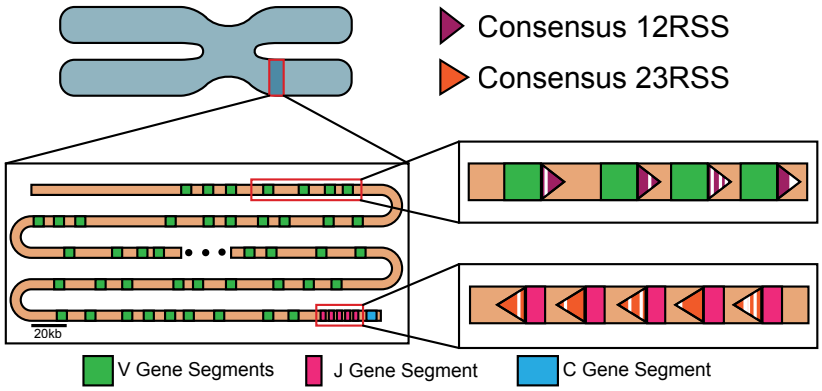

Figure 1.3: **Arrangement of antibody-encoding gene segments in the Igκ locus on chromosome 2 in humans.** The gene segments that make up an antibody were found to be fragmented in embryos and clustered according to the part of the antibody that they encode. Here, the variable (V) gene segments (green) 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

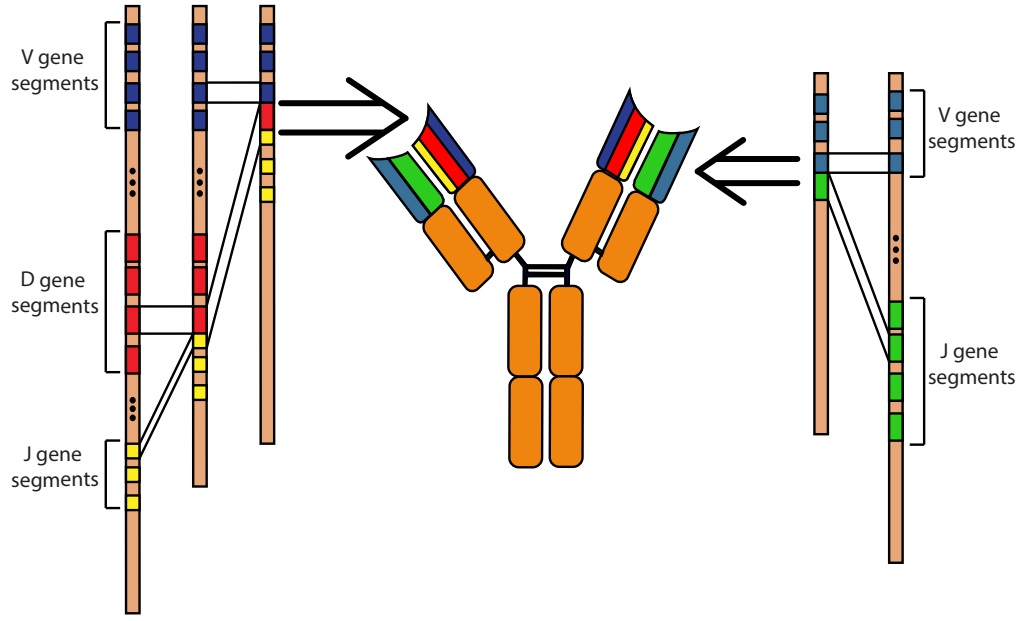

Figure 1.4: **Schematic of antibody gene encoding by V(D)J recombination.** Recombination occurs twice on one of the chromosomes to produce the heavy chain gene by combining a V (blue), D (red), and J (yellow) gene segment and once on another chromosome to produce the light chain gene by combining a 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

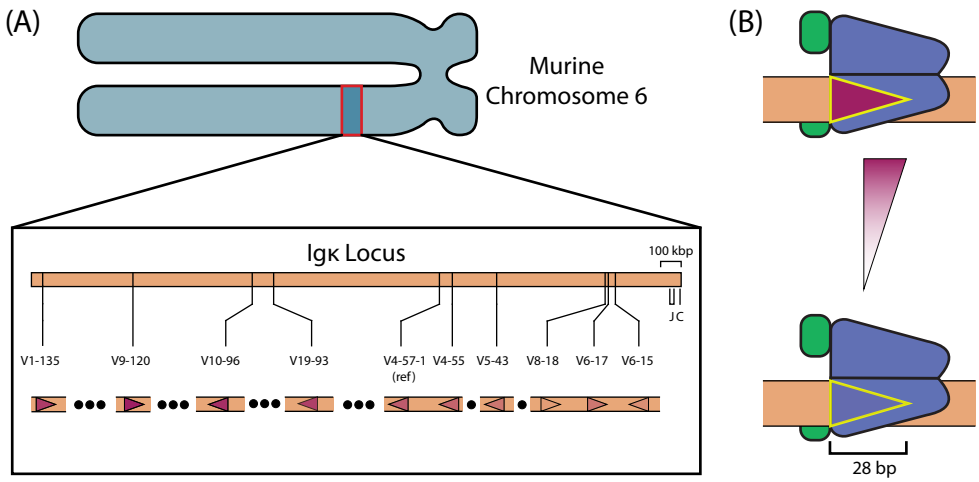

Figure 1.5: **Outline of different gene segments in the Igκ locus in chromosome 6 in mice.** (A) Gene segments can span about a few million base pairs, as shown for the Igκ locus in chromsome 6 in mice. (B) Each gene segment has a unique RSS sequence that can affect RAG binding and cutting. The triangles 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

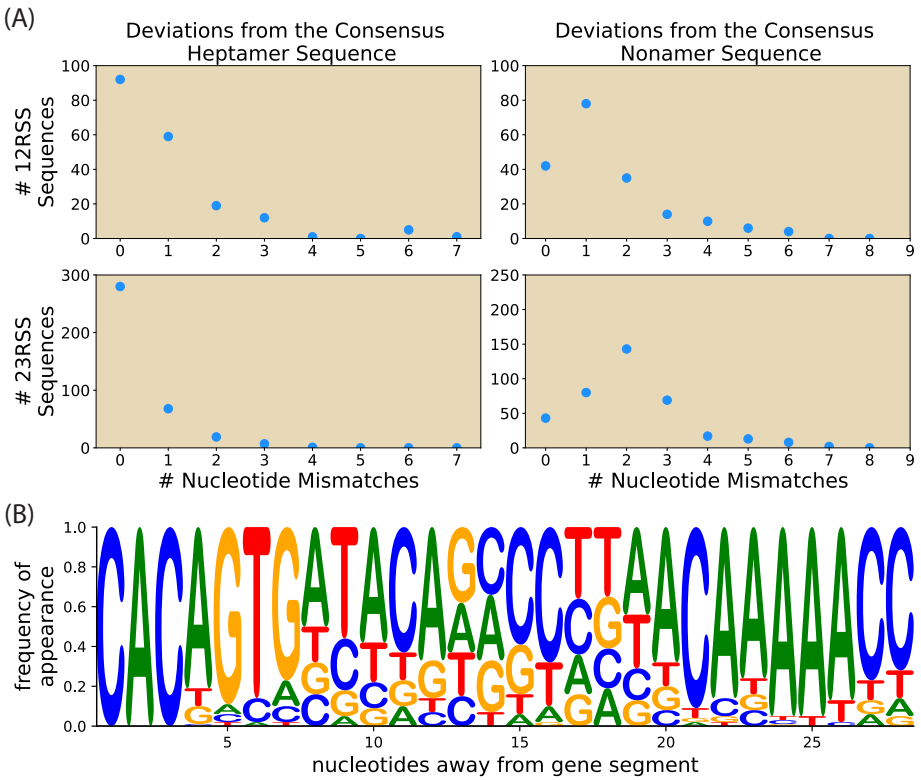

Figure 1.6: **Deviations in endogenous RSS heptamer and nonamer sequences from the consensus sequences in the mouse genome.** (A) Histograms of number of base pairs that each heptamer (left column) and nonamer (right column) deviates from their respective consensus sequences, with distinctions between 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].

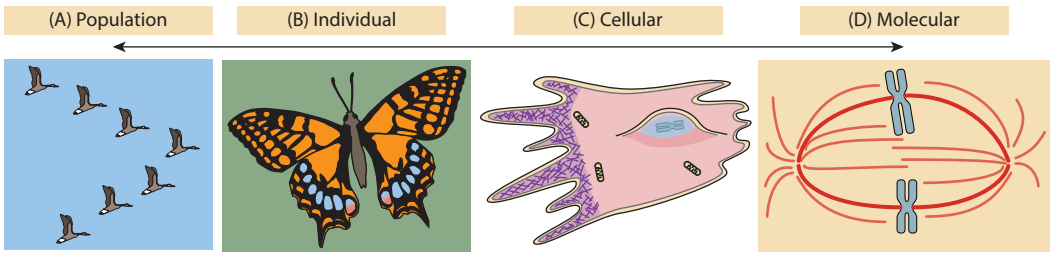

Figure 1.7: **Scales of patterning in biology.** The patterns that come from self-organization in biology can be found from the ecological level down to the molecular level. Examples range from (A) the migration of geese, (B) the patterning on butterfly wings, (C) the organization within a cell, and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

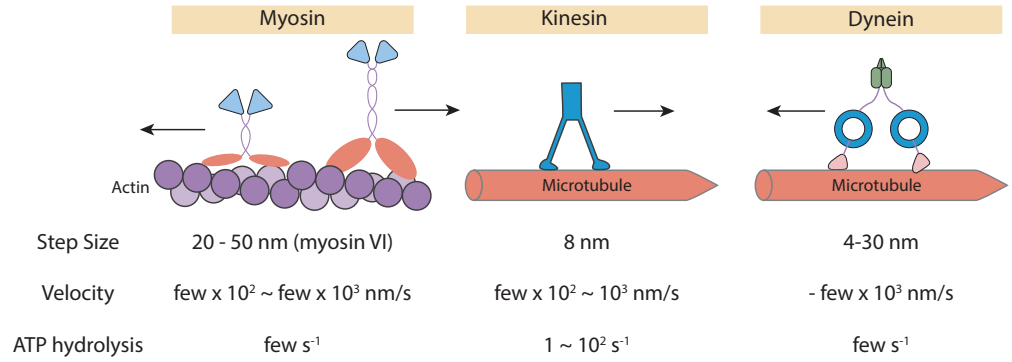

Figure 1.8: **Comparison of myosin, kinesin, and dynein motors.** While myosin motors walk along actin filaments, kinesin and dynein walk along microtubules. The direction that the myosin motor walks along the filament depends upon the specific motor while dynein motors process toward the minus end of 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

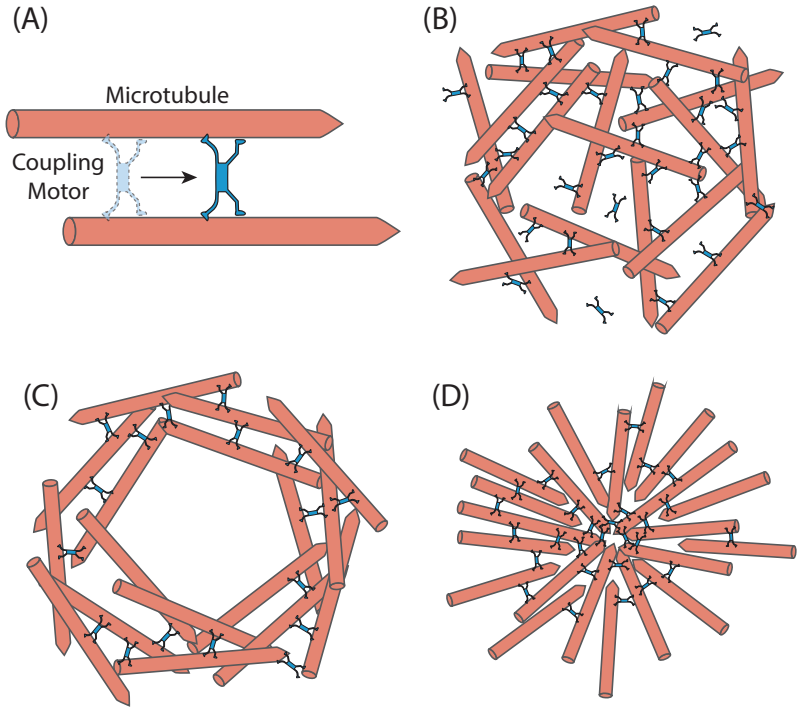

Figure 1.9: **Vortex and aster formed in *in vitro* motor-microtubule mixture.** (A) Motors were constructed such that they would multimerize, the result of which allowed them to simultaneously walk along two or more filaments. (B) Early experimental assays mixed these multimerized motors with microtubules 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

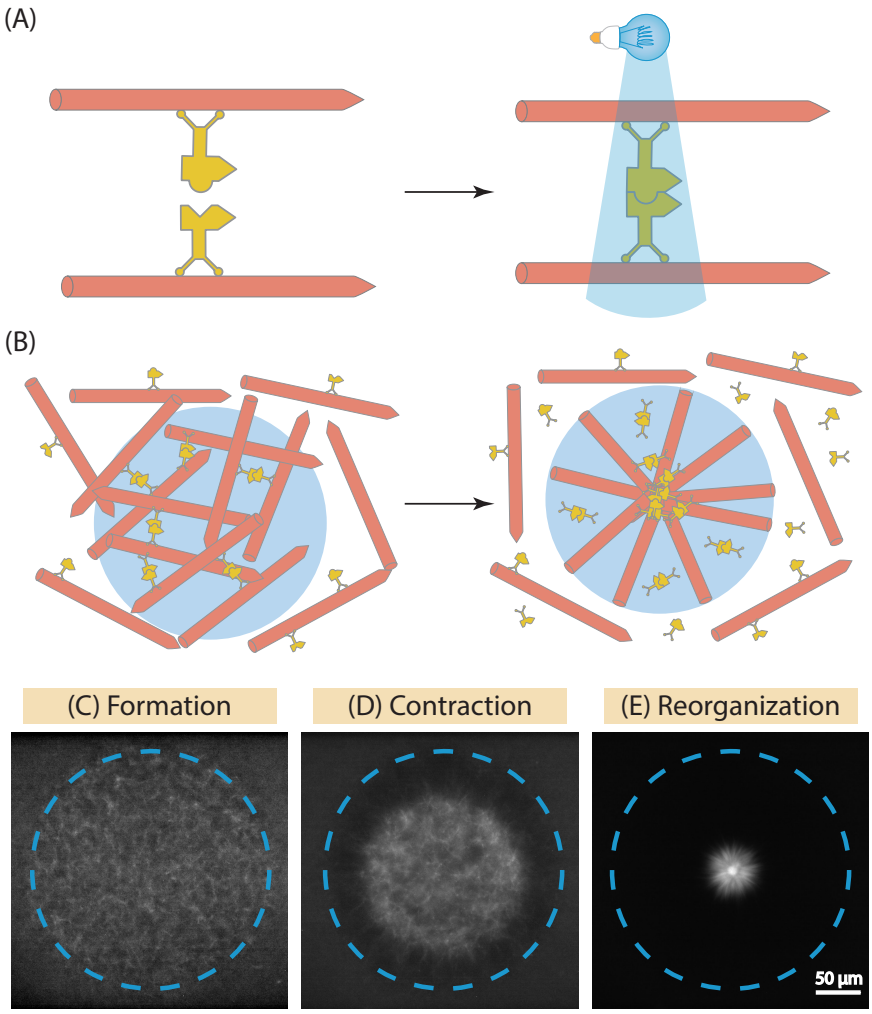

Figure 1.10: **Controlling filament coupling by light-induced motor dimerization.** (A) Motors are designed to include either an iLid or micro domain. When motors are exposed to blue light, the iLid domain undergoes a conformational change that allows for micro binding. The dimerized motors can then 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).