ABSTRACT

Title of dissertation: THEORETICAL STUDIES OF
THE WORKINGS OF
PROCESSIVE MOLECULAR MOTORS

Huong T Vu, Doctor of Philosophy, 2017

Directed by: Professor D. Thirumalai
Biophysics Program

Processive molecular motors, such as kinesins, myosins and helicases, take multiple discrete steps on linear polar tracks such as microtubules, filamentous actin, and DNA/RNA substrates. Insights into the mechanisms and functions of this important class of biological motors have been obtained through observations from single-molecule experiments and structural studies. Such information includes the distribution of n, the number of steps motors take before dissociating, and v, the motor velocity, in the presence and absence of an external resistive force from single molecule experiments; and different structures of different states of motors at different conditions. Based on those available data, this thesis focuses on using both analytical and computational theoretical tools to investigate the workings of processive motors. Two examples of processive motors considered here are kinesins that walk on microtubules while transporting vesicles, and helicases which translocate on DNA/RNA substrate while unwinding the helix substrate. New physical principles and predictions related to their motility emerge from the proposed theories.

The most significant results reported in this thesis are:

Exact and approximate equations for velocity distribution, $P(v)$, and run-length distribution, $P(n)$, have been derived. Application of the theory to kinesins shows that $P(v)$ is non-Gaussian and bimodal at high resistive forces. This unexpected behavior is a consequence of the discrete spacing between the α/β tubulins, the building blocks of microtubule. In the case of helicases, we demonstrate that $P(v)$ of typical helicases T7 and T4 shows signatures of heterogeneity, inferred from large variations in the velocity from molecule to molecule. The theory is used to propose experiments in order to distinguish between different physical basis for heterogeneity.

We generated a one-μs atomic simulation trajectory capturing the docking process of the neck-linker, a crucial element deemed to be important in the motility of Kinesin-1. The conformational change in the neck linker is important in the force generation in this type of motor. The simulations revealed new conformations of the neck-linker that have not been noted in previous structural studies of Kinesin-1, but which are demonstrated to be relevant to another superfamily member, Kinesin-5. By comparing the simulation results with currently available data, we suggest that the two superfamilies might actually share more similarities in the neck-linker docking process than previously thought.

THEORETICAL STUDIES OF THE WORKINGS OF PROCESSIVE MOLECULAR MOTORS

by

Huong T Vu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Dr. D Thirumalai, Chair/Advisor
Dr. Garegin Papoian, Co-Chair
Dr. Dorothy Beckett
Dr. Robert Briber
Dr. John Weeks
Dr. Arpita Upadhyaya

Dedication

for Mom and Dad

...

kính tặng cha và mẹ yêu

Acknowledgments

First and foremost, I would like to thank my advisor, Professor D. Thirumalai for giving me an invaluable opportunity to join the Biophysics program in UMD and to work in his group. I have been given enormous chances to study independently, while at the same time, to recognize how important teamwork is. Once he engaged in my work, he gave me critical comments to the arguments, challenged the ideas, and augmented thoughtful points. There were always more things to do after every round of his comments, and the research problem always got deeper and richer. In his group, I have learned how high-quality research should be done, understood my own strength and weakness, and recognized how I want to be in the future. No doubt that what I have learned in his group will frame how I do research in the future.

I would also like to thank my advisor committee, Dr. Beckett, and Dr. Briber, who I can talk to about my concerns or problems. I owe them for all incredible advice, recommendations, and supports since my first year. Thank Dr. Beckett for letting me stay in her hotel room during a conference. Thanks are due also to Professor Weeks, Professor Upadhyaya and Professor Papoian for agreeing to serve on my thesis committee and for squeezing their invaluable time reviewing the manuscript. And thank Dr. Sukharev, an incredible compassionate figure, for his supports on my applications.

Next, I would like to thank my colleagues in my group, who are all very kind and supportive, in particular Shaon and Naoto. Without Shaon's extraordinary

ideas, and concrete thinking at the beginning, there would be no first part of this thesis. And without the constantly supports of Naoto at the end, it would be extremely difficult for me to complete the thesis. These two men pick me up when I was most desperate, patiently working with my limitations, and has become more than only my colleagues, but my brothers, my best friends. Words cannot express the gratitude I owe them. Thank Mike who is incredibly smart and supports me on my mathematical models, and who I have always wished that I could have had more time to work with. Thanks, Yonathan, for fruitful discussions and helping me in my work on the coarse-grained model (though they are not presented here in this thesis). Thanks, Mauro, for being so supportive, and patiently reading through many of my writings. To all other members in the group that I could not mention all here: Thanks for sharing your, both research and life, experiences, with me. I might have been too shy to talk to you, but I, indeed, have great admired to you all.

My thanks are extended to the staff members in UMD. Caricia, Star, and Souad are incredibly supportive coordinators. Special gratitudes to Caricia, for her long friendship. And thank Maritza and Sokhan for all the paper works.

I gratefully acknowledge the funds received from the Vietnam Education Foundation (VEF) and the program Physics of Living Systems (PoLS) during my Ph.D. program. In particular, VEF has given me, not only financial supports but also incredible networks and inspiring experiences. I was known to have friends everywhere in the world. And most of the connections I have had are from The Vietnam Education Foundation Fellows and Scholars Association. Some of them have be-

come my best friends forever. My life in the States would have been less meaningful without community projects I shared with them.

I owe thanks most of all to my parents. Ph.D. life is not only the time for me to reflect on research but also to reflect my own self, to realize where I am from, and how I get here. And it is because of them that I can be here today, though they did not tend to. I hope that I have made them proud.

Lastly, thank Cesar, Yadviga, Liz, Guoming, Jerry, Naoto for having been wonderful roommates I could possibly ask for. Thanks to my old friends in Europe (you know who you are), who understand the most my ups and downs. And it is impossible for me to reflect all the gratitudes that I have had during the course of these six years within these three pages of acknowledgments. I apologize to those I have inadvertently left out.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

θ	step function
K-1	kinesin-1
K-5	kinesin-5
CLT	Central Limit Theorem
COM	Center of Mass
NMR	Nuclear Magnetic Resonance spectroscopy
MD	Molecular Dynamics
EM	Electron Microscopy
NMA	Normal Mode Analysis
ENM	Elastic Network Model
SPM	Structural Pertubation Method
CG	Coarse-grained (simulations)
MT	Microtubule(s)
KIF	Kinesin Family
NT	Nucleotide
NA	Nucleic Acids
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
NTP	Ribonucleoside Triphosphates
dNTP	Deoxyribonucleoside Triphosphates
ATP	Adenosine Triphosphates
AAA+	ATPases Associated with various cellular Activities
s.d.	standard deviation
SF-IV	Superfamily 4
KMC	Kinetic Monte Carlo (simulations)
LH	Leading Head
TH	Trailing Head
MH	Motor Head
NL	Neck Linker
a.a.	amino acid
SW-II	Switch II

Chapter 1: Introduction

1.1 Introduction to processive molecular motors

Molecular motors convert chemical energy, typically from NTP hydrolysis, into mechanical work to facilitate myriad activities in the cell, including gene replication, transcription, translation, and cell division [6, 24–26]. Despite the bewildering variations in their sequences, structures, and functions, molecular motors can be divided into two main types: distributive and processive motors. *Distributive motors* release their substrates after every one or few cycles of catalysis. An example of this type of motors is ATP synthase, which is a motor that moves protons down the electromagnetic gradient across the membranes to create the most important energy storage molecules in cells, Adenosine Triphosphates (ATP). One ATP molecule is created for every 3 (or slightly different numbers, depending on the systems) H^+ moved from the inter-membrane space to the matrix in mitochrondria [27]. The process is reversible if the electromagnetic gradient is small or reversed. In that case, each cycle of ATP hydrolysis, the motors release 3 H^+ from the matrix into the inter-membrane space. In contrast to the motors of this type, a large number of cellular motors undergo many cycles of catalysis before they detach from their substrate, and are known as *processive motors*. Some of them are capable of tak-

ing multiple steps directionally along linear tracks. Well-known in this category are kinesins [28], myosins [29] and dyneins [30] that carry vesicles and organelles along microtubules or actin filaments [6,31], and helicases that unwind nucleic acid strands while translocating on them [32,33]. This thesis focuses on the latter type, the processive molecular motors.

Fundamental insights into their functions have emerged from single molecule experiments [8,14,25,34–40] and structural studies [6,7,41]. Corresponding to those two experimental methods, there are two main theoretical approaches that have been widely applied to investigate the fundamental principles underlying this class of molecular motors: standard chemical kinetics based mathematical models and structural based computations and simulations. Firstly, mathematical models based on standard chemical kinetics have been developed to analyze the quantities that are measured in single molecule experiments [42, 42–53]. Such quantities include: velocity; run-length (or processivity) which is the distance that each motor travels before it dissociates from its track; size of each step the motors take; dwell time the motors spend in a particular binding site; and the dependence of those quantities on the NTP concentration of the solution and the applied (both resistive and assistive) forces to the motor or to the track that the motor walks on. This mathematical modeling approach, based on chemical kinetics, has been used to answer questions at the individual motor and motor-motor level (different to the types of questions that structure based theory can answer), such as: How do they move? How much energy do they consume? What are the efficiencies? How fast and processive are they? What are their collective behaviors? The other theoretical approach includes

the structure based computations and simulations. These methods have been proved to be powerful in gaining information about mechanisms of the mobility of motors, from the static pictures obtained in structural experiments [54–57]. These structure based theoretical models shed light on the dynamics inside the motors, and are not limited to issues such as: how the motors move, generate forces or function; where the ligands bind; what the allosteric regulations are. Depending on the types of the questions to be asked, the simulations or computations might be performed at amino-acid, atomic or higher or lower levels. (Other methods that are also important in the study of molecular motors are sequence-based computations [58,59]. However, they are not considered here in this thesis.)

Among the mathematical models that have been used, there are 2 most successful and widely used methods: discrete stochastic (or chemical kinetics) models [42–47] and Brownian ratchet (or continuum ratchet) model [42,48–51] (Fig. 1.1). These two methods are not unrelated to each other. In fact, a simple ratchet model was analyzed and proved to be a simple case of the discrete stochastic model [60]. In addition, a two-state model of a stochastic model is indeed a ratchet model [51]. These models have focused on physical quantities such as mean velocity, mean run-length, and the dwell-time distribution [39,61], and were able to reproduce the single molecule experimental data at varying external forces and ATP concentrations. Given the inherently stochastic nature of the motor cycle, velocity fluctuations must play an important role in motor dynamics. However, up to now analytical tools to interpret the fluctuation data, readily available from experiments, do not exist. The *first half of this thesis* seeks to address this gap, providing universal closed-form

expressions for velocity and run-length distributions valid for any processive motor, using stochastic models.

Structural studies, such as from X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and electron microscopy (EM), are powerful in providing snapshots of biological systems. However, they lack direct dynamic information needed to describe the stepping of motors. Especially, for molecular motors, structural details of the stepping mechanisms are important in order to understand how they operate. Since the first paper [54] in 1977, molecular dynamics (MD) simulations have played important roles in revealing the dynamic pictures of many biological molecules by providing individual molecular motions as functions of time. In turn, MD can also assist to refine X-ray or EM data, to obtain a better static structure. Depending on types of questions and systems, different dynamics with different levels of description needed (Table 1.1). Another structural based method that complements molecular dynamics simulations is Normal Mode Analysis (NMA) [55–57]. This method considers the interactions between individual particles as harmonic potentials, with the assumption that the crystal strutures is at the lowest local minimum. This allows one to calculate the motions of the lowest modes, related to the flexibility of the molecule around equilibrium. The most popular method to calculate the normal modes of a molecule is the Elastic Network Models (ENM), where one single spring constant is used in the harmonic potentials between all pairs of beads that are in contact with each other within a cut-off distance in the crystal structures [62, 63]. The *second half of this thesis* (and some continuing future work which are not included in this thesis) focuses on using these powerful tools to

address questions on the docking process of kinesin-1 (and stepping mechanisms of helicases).

To summary, processive molecular motors are ubiquitous inside cells. Hence, it is very important to understand their functions and characteristics. This dissertation is an extension from the current theoretical studies of such motors, with an emphasis on developing models/simulations to analyze and interpret current experimental data, in order to discover new general principles of the workings of those biological motors. The theories are demonstrated here with two illustrations: kinesins and helicases, but they are general to the wide class of processive motors. Brief introductions about kinesins and helicases are in the next sections.

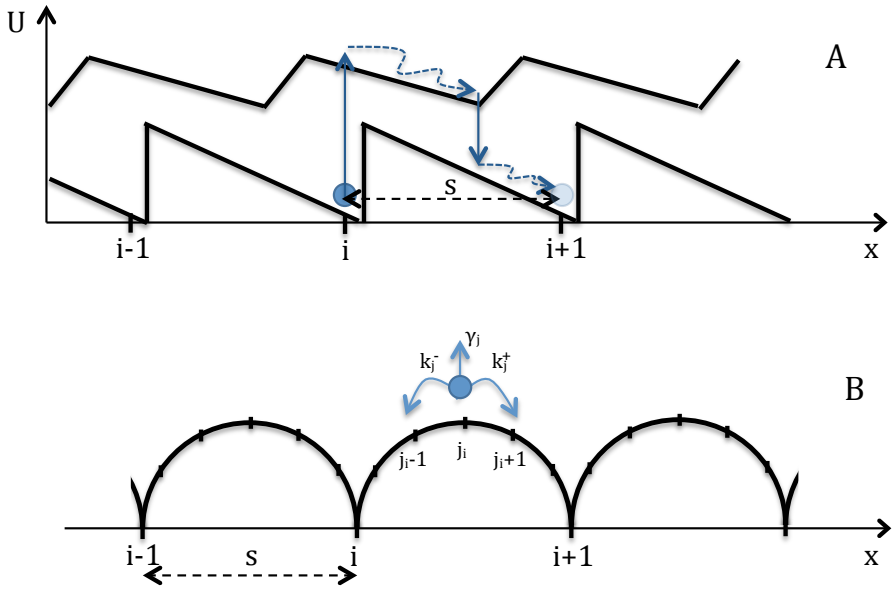


Figure 1.1: Illustrations of the two mathematical models for processive motors: (**A**) Brownian ratchet model and (**B**) stochastic discrete model. **A:** A Brownian ratchet model is normally illustrated with 2 periodic potential levels. A motor (blue circle) at the binding site i utilizes the energy from the ATP hydrolysis to jump to the higher energy (long vertical solid arrow). It undergoes a Brownian diffusion in the higher energy surface (dashed arrow), jumps down to the lower energy (short vertical solid arrow), and diffuses to the next binding site (dashed arrow), thus, completing a step. With the two potential levels and the chemical energy to pump the motor to the higher level, the motor can "diffuse" toward one direction. **B:** In a stochastic discrete kinetic model, two adjacent binding sites i and $i+1$ are separated by intermediate states, and at each state, the motor can hop to the next state (forward or backward) or dissociate from the track with different stochastic rates (blue arrows). For a review, see [42].

Table 1.1: Different dynamics in molecular dynamics simulations. For details, see [1].

	Type of dynamics	Motion equation	Notes
(a)	Newtonian	$m_i\ddot{r}_i = -\Delta_i U(\mathbf{r})$	The solvent might be treated implicitly or explicitly. Normally used in all-atom simulations.
(b)	Langevin	$m_i\ddot{r}_i = -\Delta_i U(\mathbf{r}) - m_i\eta_i\dot{r}_i(t) + R_i(t)$	Including the two latter terms representing to the interactions with the solvent. Normally used in coarse-grained (CG) simulations to access long timescales.
(c)	Brownian	$\dot{r}_i(t) = \frac{-\Delta_i U(\mathbf{r}) + R_i(t)}{m_i\eta_i}$	In the over-damped solvent regime, and also normally accomplished with CG simulations.

1.2 Introduction to kinesins

Kinesins are one of the smallest molecular motors inside cells. Most of them "walk" as transporters on microtubules (mostly towards the plus end), which classified them as processive motors. Since the initial discovery of a small force-generating motor in 1985 [64], many different families of kinesins have been found in different oligomer states, and have been associated with different functions inside cells (Table 1.2). In the genome of mouse alone, there are already 45 different kinesin superfamilies (KIF) [5]. The most standardized kinesin nomenclature divides them into 14 families. The first family in this nomenclature, kinesin-1 (K-1 for abbreviation in this thesis), is the family of the conventional kinesins that were first discovered in the squid axons [2]. Structurally, each K-1 motor contains two heavy chains forming the main motor subunits, which bind to two other small chains. The two heavy chains (KHC) bind a cargo with the binding domains at their C terminals, twist around each other to form a "coiled coil stalk" that connect to two motor heads at the N terminals (Fig. 4.1). The connections between the heads and the stalk are small segments named "neck linker" (NL), which is believed to be a critical element in force generation.

Kinetic analysis of the single molecule experimental data has shown that the motors use motor heads to walk on the MT in the manner called "hand over hand": one head binds, one head moves at a time, and then they switch [65, 66]. With this mechanism, they are able to take about 100 steps processively on a single protofilament, towards the plus end of the microtubule (MT). Each step-size is

exactly 8.2 nm, which is the mean distance between the two adjacent binding sites on the MT, which is polymer consisting of $\alpha\beta$ monomers. The motors hydrolyze one ATP molecule per step.

During the ATP cycle, the motor head changes its structure, alternates its MT affinity, and generates force through the NL to move the other head forward and bind to the next binding site. In particular, the motor heads are believed to change their conformations so that they have higher MT affinity in the ATP-bound or no-nucleotide states than in the ADP states (Fig. 1.2). When the motor head binds MT and ATP, the NL binds and interacts with the core domain of the motor, adopting an "ordered" conformation. However, at the no-nucleotide state, the NL unbinds from the core domain and becomes a flexible, "disordered" coil. Interestingly, in the ADP bound states, the NL has been observed with both ordered (2KIN [67], 3KIN [68], 1MKJ [69]) and disordered (1BG2 - [17]) conformations. It was suggested that the distinct NL conformations are only observed when the motor is in complex with MT [70].

Much of the detailed structural changes during the ATP cycles have been resolved, thanks to cryo-EM imaging and crystallographic works performed for over 30 years [72]. Most importantly, structures of MT-K-1 complex have finally been solved using both methods, almost within the same year [19,71,73,74]. The allosteric regulations between the NT binding, the MT binding, and the conformation of the NL has been proposed leading to good agreements with a "seesaw" model [41, 75] (Fig. 1.2). The motor head can be divided into 3 main domains: SWI/II domain, P loop domain and tubulin binding domain; with 3 clefts created between them: a

docking pocket, a nucleotide cleft and a polymer cleft (Fig. 1.2). The NT and MT binding (or unbinding) close (or open) one or two clefts on one side of the motor head, trigger the open (or closure) of the pocket 15 $\mathring{A}$ far away on the other side.

As explained above, current experimental and theoretical works have revealed many insights about the dynamic mechanism of kinesins. Still, many details on mechanisms of this type of processive motors are not fully understood [76]. Typically, some important questions for transporting motors are still open, such as: how forces are generated, what are their regulation mechanism (gating mechanisms), and how their behaviors depend on external loads. Also, with a large number of different sub-classes revealed, questions especially related to the commonality among the many different families, and how the variations and similarities contribute to functions inside the cell have become more important. The theories explained in this thesis contribute to answer some of those important general questions.

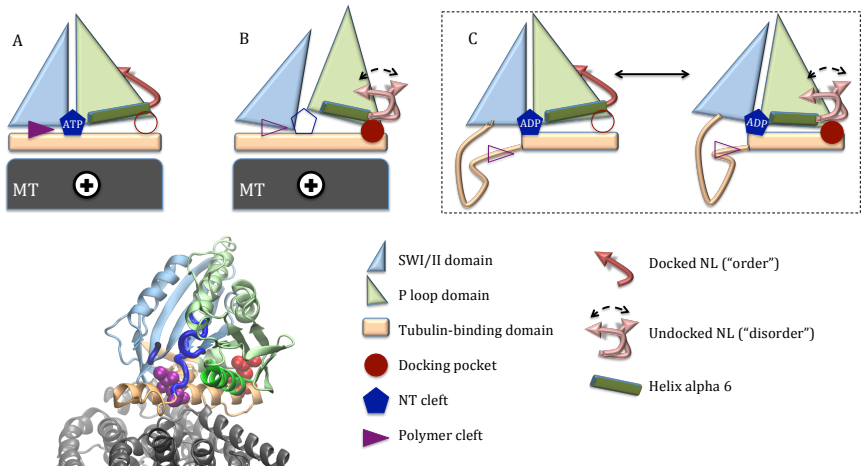


Figure 1.2: Seesaw model for kinesin head.
Kinesin head is divided into 3 main domains, forming 3 different clefts/pocket at the interfaces. **A:** ATP molecule binds to a MT-bound head, closing the NT cleft (filled pentagon). As a consequence, the polymer cleft adjacent to the NL cleft is also closed (filled triangle), opens the docking pocket at the other end of the "seesaw" (opened cicle), allowing the NL (red arrow) to interact and bind to the core domains (mainly with the P loop domain in this case). **B:** MT-bound head with no-nucleotide bound has an opened NT pocket (opened pentagon). Thus, the polymer cleft is opened (open triangle), closing the docking pocket at the other side of the motor head, thus, preventing the NL from interacting with the core domain, making the NL disorder (pink arrows). **C:** When the kinesin is unbound from the MT, the "relay helix" that forms the main MT binding interface (long orange rectangle) is shortened and creates a longer disorder adjacent loop (orange loop), making the polymer cleft impossible to close (open triangle). Thus, even when the NT cleft is closed with a nucleotide (ADP) bound to it (filled pentagon), the docking pocket is not locked at an opened or closed form. The NL is capable of transfering between order and disorder conformations. **Bottom** is the structure of the MT-bound kinesin head at the no-nucleotide bound state [71], colored corresponding to the schemes in A, B and C.

Table 1.2: Examples of kinesin families. For the full standardized nomenclature, see [2]. For reviews, see [3–6]

Standardized name	Oligomers	velocity in vivo (nm/s)	velocity in vitro (nm/s)	Found in	Function	Known as
Kinesin-1	dimer	1800	840	axons	cargo transport	conventional kinesin/ kinesin heavy chain (KHC)
Kinesin-3	monomer/dimer	690	1200	axons/ mitosis	cargo transport	monomeric kinesin (though they might actually work as dimers in vivo [77])/ Unc104/kif1
Kinesin-5	bipolar tetramer	18	60	mitosis/ meiosis	spindle formation	Eg5 (human)/ bipolar mitotic/ tetrameric/ BimC
Kinesin-14	dimer	–	-90	mitosis/ meiosis	Chromosome segregation/ cargo transport	Ncd/ C-terminal motors (the only family walks toward the minus end of the MT)

1.3 Introduction to helicases

The main function of helicases is to destabilize or unwind the helical structures of long DNA or RNA substrates. Thus, the helicase function is also normally accompanied by a translocase function, in which helicases processively walk along nucleic acids while unwinding them. These functions also require NTP/dNTP hydrolysis, as many other processive motors. The first helicase TraI was discovered in *E.coli* in 1976 [78, 79]. Since then, various helicases have been classified into 6 superfamilies based on the types and orders of their shared "signature motifs" [33, 80]. Different helicases can work with different oligomer states and have different functional mechanisms (Table 1.3 and Fig. 1.3).

Table 1.3: Six superfamilies of helicases [7].

	Oligomers	Examples	Folds	Polarity
SFI	monomer/dimer	PcrA/UvrD/TraI	RecA	SFIA 3'-5'/ SFIB 5'-3
SFII	monomer/dimer	RecG/RecQ/NS3	RecA	SFIA 3'-5'/ SFIB 5'-3
SFIII	hexamer	E1	AAA+	3'-5'
SFIV	hexamer	DnaB/T7/T4	RecA	5'-3'
SFV	hexamer	Rho	RecA	5'-3'
SFVI	hexamer	MCM	AAA+	3'-5'

Structures of the NTP/dNTP binding sites of helicases fall into the class of P-loop NTPases, which share in common two motifs, called Walker A (AxxGxGKT) and Walker B (DExx – with A, G, K, T, D, E are names of amino acids, and x is any possible amino acids) [81]. More specifically, they belong to the Additional Strand Catalytic E (ASCE) subgroup, characterized with the presence of a catalytic glutamate, and the nucleotide binding site formed at the cleft of two separated

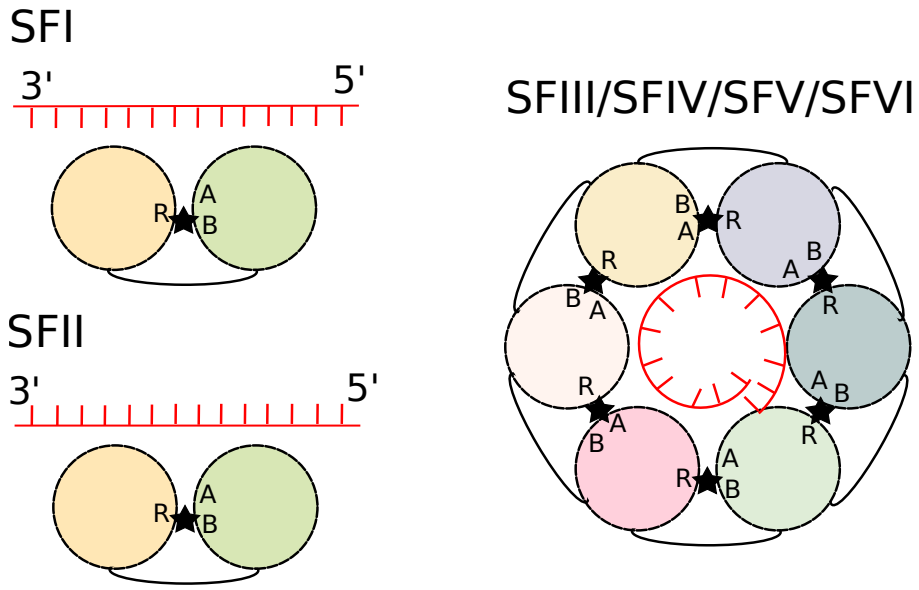


Figure 1.3: Schematics of DNA/RNA helicases from different superfamilies. SFI and SFII helicases have two main DNA/RNA domains (other than auxiliary domains) that have Rec-A like conformations, forming a NT binding site in the cleft (black star). The two domains also interact with NA strand, mainly with the bases in SFI (**left above**), and the backbone in SFII (**left below**). The other four members of the superfamilies are hexamers (**right**). 6 Rec-A-like or AAA+-like subunits form 6 NT binding sites sandwiching between them. The traslocated ssNA substrate is threaded through the central pore, forming interactions with the 6 subunits mainly via the backbone. DNA/RNA strands are in red. Black stars are NT binding sites. A, B, R are the 2 common conserved motifs (Walker A and B) and an arginine finger R in NT pockets.

domains (Fig. 1.3). (Within this ASCE subgroup, there are two smaller subdivisions known as Rec-A fold and AAA+ fold, with an addition of C-terminal α-helical domains in the AAA+ fold [7].) The NTP/dNTP binding sites normally formed by Walker A (also known as the "P-loop") and Walker B motifs on one domain, with an arginine finger (or other additional elements necessary for hydrolysis) coming from another domain. The domains separately bind the NA strand in the other sites (Fig. 1.3). Once bound, NTP/dNTP molecules are sandwiched in the clefts

formed by the two adjacent domains, changing the widths of the clefts during their hydrolysis cycles.

Based on both bulk and single molecule experiments and structural studies, it is suggested that the relative changes in the distances between the domains, coupled with the different DNA/RNA binding affinities, is the main mechanism for the motors to translocate. In the cases for monomeric helicases, this is known as the "inch-worm" model [7, 33]. In the case of hexameric helicases, it is suggested to be a sequential NTP hydrolysis model. However, details about the order of the sequential models are not known yet. The monomeric/dimeric helicases might contain additional domains or units that directly interact with the double strand junctions to destabilize the basepairs. These domains serve as wedges or "pin points" for the helicase function. The hexameric helicases, on the other hand, adopt a different mechanism. They thread one sDNA/sRNA strand through the central channel formed by the 6 subunits, and exclude the other strand of sDNA/sRNA outside of the ring, separating the two strands (Fig. 1.4) [7, 33].

One of the stochastic discreted models has been developed to analyze data from single molecule experiments of helicases more than 10 years ago and has been referred to as Betterton and Julicher's model [47]. This model has been wildly used to analyze data in several works to estimate important characteristic parameters of helicases, such as step-size, interacting range, interaction energy between the motor and the junction [15, 16]. They also proposed to classify helicases into two types: active vs passive, based on the active involvements to destabilize the duplex or passively waiting for the thermal fluctuation to open it. Since then, there are several

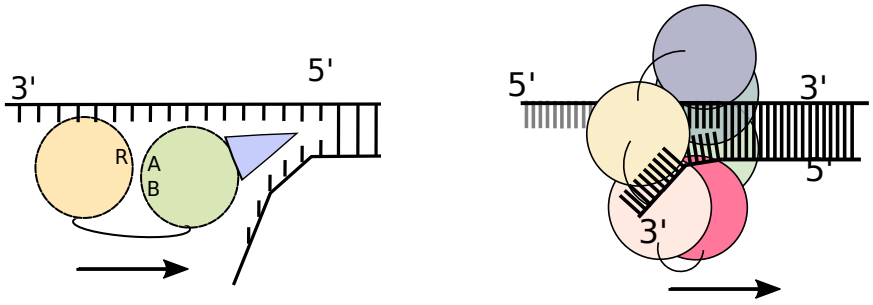


Figure 1.4: Schematics of the unwinding and translocating mechanisms in helicases.

different criteria for classifying helicases as active or passive [47,82]; although strictly speaking, all helicases are active, with some are more so than others. There are also two ways to define step-size in helicases: physical step-size (per NTP hydrolysis) vs kinetic step-size (between 2 successive rate-limiting steps of the reaction). Interestingly, it is commonly known in the field that there are inconsistencies in the estimations of the step-sizes in many helicases, due to the limitations of resolution in tracking measurements to visualize the step-size directly (compared to kinesins). Moreover, the heterogeneity nature of helicases also contribute to limit the correct estimation of parameters (including step-size) from bulk experiments [83]. Also, using Betterton and Julicher's model, different sets of parameters were able to explain the same single molecule experiments which leads to a limitation of estimating the correct parameter set [15, 16]. The last limitation of current theory has been resolved by an extended model developed in our group [23, 84]. Using their results,

the theory explained in this thesis can be used to analyze the heterogeneity nature of helicases.

1.4 Outline of the Thesis

This thesis focuses on using both analytical and computational tools in order to discover general principles of the functions and dynamics of biomolecular motors.

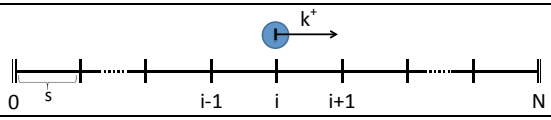
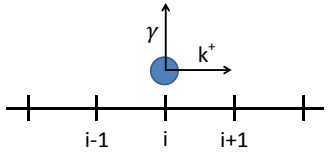
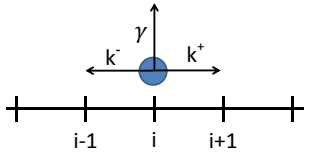
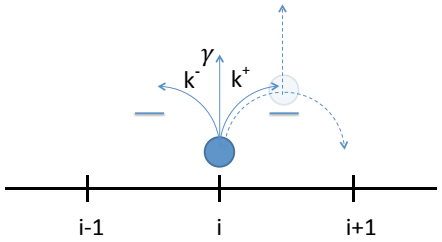
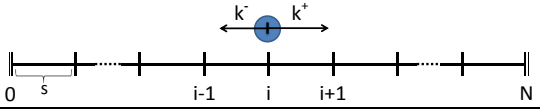
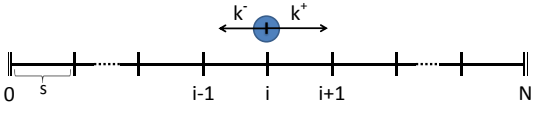
Fluctuations in the physical properties of biological machines are inextricably linked to their functions. Distributions of run-lengths and velocities of processive molecular motors, like Kinesin-1 and helicases, are readily accessible through single molecule techniques, yet the lack of a rigorous theoretical model for these probabilities hitherto has prevented their use in experimental analysis. In the first half of the thesis, I describe analytical models to investigate velocity distributions of processive motors. The models allow us to derive exact expressions for velocity distributions under various circumstances (Table 1.4). The theory quantitatively explain experimental data, allowing us to predict interesting behavior of the motors at the molecular level. In **Chap. 2**, we show that the distribution of velocities of Kinesin-1 becomes bimodal under load as a consequence of the discreteness in the step-size traversed along the MT [85]. In **Chap. 3**, we quantitatively examine the emergence of heterogeneity in a helicase by analyzing their velocity distributions, based on a model developed in the group [23, 84], which in turn is a extension of an earlier model for helicases [47]. We show that T4 and T7 helicases in SFIV exhibit heterogeneity. Based on our model, we propose further experiments to discern

between two types of heterogeneity, *dynamic disorder* vs *static disorder* [85]. The detailed derivations for the different types of velocity distributions can be found in Appendix A. Refer to Table 1.4 for the schematics of the different models.

Computational tools are useful in order to understand the structural basis of function and mechanisms of molecular motors. Different methods can be chosen to tackle different types of questions. In **Chap. 4**, I describe all-atom simulations to investigate the docking process of Kinesin. A one μs simulation was able to capture, for the first time in simulations, the docking of the neck-linker in the Kinesin-1 head, which allows us to predict possible sites that the neck-linker interacts with in the "disordered state". These binding sites correspond to the two binding sites that were obtained in crystal and EM structures of Kinesin-5, suggesting that the two types of kinesins have very similar docking mechanisms.

Appendices B and C provide supplemental materials for chapters 2, and 3, respectively.

Table 1.4: Velocity distributions have been calculated for 6 different models as below. Detaied derivations are in Appendix A.

	Notes	Schematics	Notations	Source
Model 1	Fixed runlength		$P_N(v)$	Chapter 2, 3
Model 2	Finite processivity			Chapter 2
Model 3	Finite processivity		P(v), ($P_A(v)$ for approximation)	Chapter 2
Model 4	Finite processivity with an intermediate state		$P_i(v)$	Chapter 2
Model 5	Fixed runlength		$P_N(v)$, ($P^A{}_N(v)$ for approximation)	Chapter 3
Model 6	Fixed runlength with *static disorder* in the forward rate		$P_N{}^H$ ($P_N{}^{H,A}$ for approximation)	

Chapter 2: Discrete step sizes of molecular motors lead to bimodal non-Gaussian velocity distributions under force

2.1 Overview

Let us assume a motor takes steps of size s on a polar track, leading to a net displacement ns before detaching at time t, where n is an integer proportional to the run-length. The natural definition of average velocity for the trajectory is $v = ns/t$. This chapter tackles several key questions relate to the distributions $P(v)$ and $P(n)$ for molecular motors subject to resisting force, F. A natural question is: when can $P(v)$ be represented by a Gaussian, an approximation used to analyze experiments [14, 35, 77, 86–88]? At small force, if the detachment rate from the polar track γ is negligible compared to the forward rate k^+ (Model 2 or Model 3 with $k^- << k^+$ in Table 1.4), the motor takes a large number of steps forward before detachment. In this case, $P(v)$ should approximately be a Gaussian, as expected from the Central Limit Theorem (CLT). However, what is the behavior of $P(v)$ and $P(n)$ when γ becomes comparable or even larger than the other rates, situations encountered in single-molecule experiments in the presence of external force (Fig 2.4A [25, 89])? To address these questions, we derive exact analytical

expressions for $P(n)$ and $P(v)$, using a simple but accurate kinetic model, redrawn in Fig. 2.2B), with only three rate parameters (Table 1.4 Model 3). The model has a broad scope, allowing for analyzing with potential prediction for experimental outcomes for a large class of processive motors.

The central results in this chapter are: (i) At non-zero F, $P(v)$ is non-Gaussian because γ cannot be neglected compared to k^+, resulting in the number of steps being not large enough for CLT to be applied. Even when $F = 0$ there is a discernible deviation from the Gaussian distribution. (ii) Surprisingly, when $F \neq 0$, $P(v)$ is asymmetric about $v = 0$ with a bimodal shape containing peaks, one at $v > 0$ and the other at $v < 0$. With increasing F, the peaks become symmetrically positioned with respect to $v = 0$, and is completely symmetric at the stall force F_S. (iii) As F exceeds F_S, reaching the superstall regime, the peak position at $v > 0$ ($v < 0$) moves to higher (lower) values. We show that these counter-intuitive results are consequences of the discrete nature of steps that molecular motors take on their tracks.

2.2 When will $P_N(v)$ be non-Gaussian?

In order to understand when the velocity distribution $P_N(v)$ will be Gaussian or non-Gaussian depending on the fixed run-length N, we derive an exact expression for $P_N(v)$ from the Model 1 in Table 1.4 (see Appendix A for details). This expression for $P_N(v)$ (Eq. A.8) is plotted in Fig 2.1, for increasing values of N. In each panel of the figure, a Gaussian is fit to data points generated from Eq. A.8.

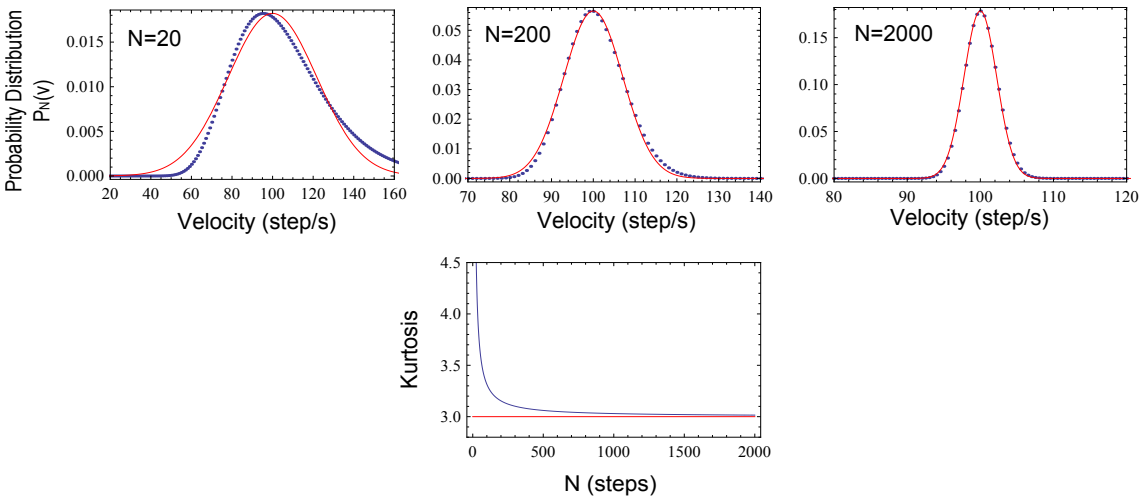


Figure 2.1: Comparision of velocity distribution with a Gaussian function for different number of steps N.
The blue dotted lines are exact results and the red solid lines are Gaussian fits. Below: Kurtosis of $P_N(v)$ depends on number of steps N (blue line). As N increases, the kurtosis approaches 3, the exact value for a Gaussian distribution.

Clearly, as N becomes larger, the Gaussian fit becomes better and at $N = 2000$, the Gaussian is indistinguishable from the theory. When $N \to \infty$, the velocity distribution approaches a delta function. This behavior is not surprising, since the distribution becomes more Gaussian-like at large N a consequence of the central limit theorem.

This simple model, however, explicitly highlights the fact that for $N \sim 100$ or less, the velocity distribution could be distinctly non-Gaussian. In reality, the processivity of any molecular motor with ($\gamma \neq 0$) will be finite, and hence N is unlikely to be large enough for CLT to be valid. For example, at zero force, Kinesin takes on an average $N = 77$ steps prior to dissociation [14]. Moreover, when a motor is under a resistive force either *in vivo* or in optical tweezer experiments in *vitro*, the detachment rates is even larger, which would make N smaller than 100. Under

these circumstances, Gaussian approximation for $P(v)$ may be inaccurate and more accurate expressions need to be derived.

2.3 $P(v)$ at zero force

We first analyze the $F = 0$ data for kinesin-1 (K-1) [14] using our theory for Model 3 (Table 1.4), in which each motor can move towards the plus/minus ends of the track or detach from any site. From the model, the run-length distribution should take the form (as in Eq.A.44 and A.45),

$$P(n \gtrless 0) = \left(\frac{2k^{\pm}}{k_T + \sqrt{k_T^2 - 4k^+k^-}}\right)^{\pm n} \frac{\gamma}{\sqrt{k_T^2 - 4k^+k^-}}, \tag{2.1}$$

and $P(n = 0) = \frac{\gamma}{\sqrt{k_T^2 - 4k^+k^-}}$. The exact expressions for $P(v \gtrless 0)$ is (Eq. A.63 and A.70)

$$P(v \gtrless 0) = \frac{\gamma}{|v|} \sum_{n=0}^{\infty} \left(\frac{n}{|v|}\right)^{n+1} \frac{1}{n!} \left(k^{\pm} e^{-\frac{k_T}{|v|}}\right)^n {}_0F_1\left(;n+1;\frac{n^2k^+k^-}{|v|^2}\right). \tag{2.2}$$

Experiments have shown that K-1 walks hand-over-hand [90,91] taking discrete steps in multiples of $s = 8.2\,\text{nm}$ with each step being almost identical [34, 37, 38, 92]. The 8.2 nm is commensurate with the α/β tubulin periodicity of a single MT protofilament, here modeled as a one dimensional lattice (Fig 2.2) [28]. For K-1 the measured $F = 0$ mean velocity is 1089 nm/s (Fig. 2.3B) [14], which implies that $k_0^+ - k_0^- = 132.8\,\text{step/s}$. The ratio $\frac{k_0^+}{k_0^-}$ is not reported in [14], which forced us to use data from other sources. This ratio is reported independently in 2 papers as $\frac{k_0^+}{k_0^-} = 221$ [8] and $\frac{k_0^+}{k_0^-} = 802$ [9]. Solving the two equations, we get k_0^+ and k_0^- corresponding to the two experimental values of the ratio (Table I (a) and (b), note

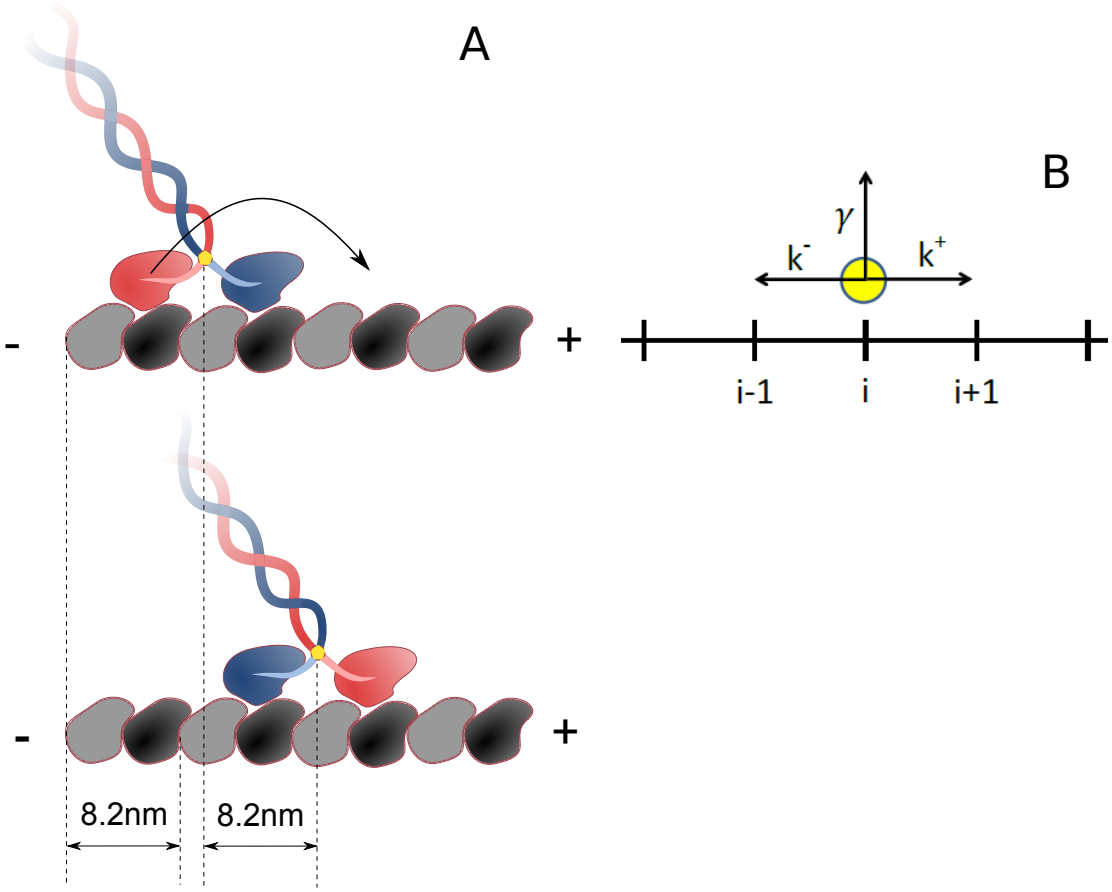


Figure 2.2: Kinetic stochastic model applied to kinesins. **A:** Schematic of a kinesin molecule walking hand-over-hand on a microtubule (MT) with a discrete step-size of 8.2 nm. **B:** Sketch of the Model 3 (Table 1.4) in which the yellow circle represents the center-of-mass (COM) of the kinesin (capturing the point in yellow in **A**). The position of the COM on the MT is denoted by i. Kinesin can step ahead, back, and detach from the microtubule with rates k^+, k^-, and γ respectively.

Table 2.1: Force-dependent rates for K-1. (a) and (b) are rates at $F = 0$ obtained from experimental data on Kin1/acetylated microtubule with ratio $\frac{k_0^+}{k_0^-} = 221$ [8] and $\frac{k_0^+}{k_0^-} = 802$ [9] respectively. The error in γ_0, obtained by simultaneously fitting $P(n)$ and $P(v)$ (Fig. 1), is estimated using the bootstrap sampling method [10]. (c)-(g) are rates at different values of load, calculated using the $F = 0$ values in (a).

	F (pN)	k^+ (s^{-1})	k^- (s^{-1})	γ (s^{-1})
(a)	0	133.4	0.6	2.3±0.3
(b)	0	133.0	0.2	2.3±0.3
(c)	3	47.1	1.8	6.2
(d)	4	33.3	2.5	8.7
(e)	5	23.5	3.6	12.1
(f)	7.6	9.5	9.3	29.0
(g)	8.5	7.0	12.8	39.1

that even though the ratios are different, k_0^+ and k_0^- are similar.) Using k_0^+ and k_0^-, we obtained the detachment rate at zero force, γ_0, by simultaneously fitting the measured [14] velocity distribution using Eq. 2.2 and run-length distribution using Eq. 2.1. The excellent fits to both data sets $P(n)$ and $P(v)$ (Fig. 2.3) using a *single* parameter shows that our theory captures the basic aspects of the K-1 motion. More importantly, the value of the only unknown parameter γ_0 is the same in Table 2.1(a) and (b) and is in a very good agreement with an independent way of obtaining the detachment rate (see Appendix B). While analyzing the experimental data, we convert velocity in step/s to nm/s (or vice versa) by multiplying (dividing) by $s = 8.2\,\text{nm}$.

2.4 $P(v)$ under resistive force

The distribution $P(v)$ has the same form as in Eq. 2.2 when the motor is subjected to an external force, F, (Fig. 2.4A) except $k^+(F)$ and $k^-(F)$ are depen-

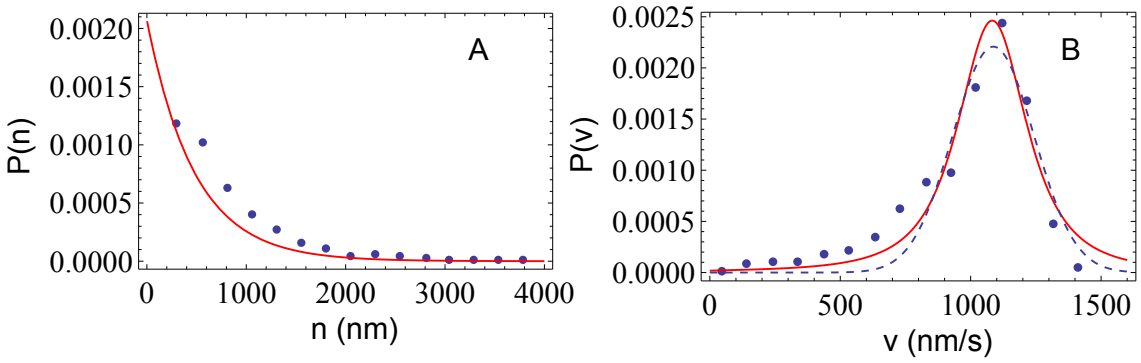


Figure 2.3: Simultaneous fits (red lines) of zero force K-1 data (blue dots) [14] for run-length (**A**: $P(n)$–Eq. 2.1) and velocity (**B**: $P(v)$–Eq. 2.2) distributions.
The dashed line in **B** is a Gaussian fit. It should be stressed that the results in **(A)** and **(B)** were fit using a single parameter, γ_0, with the extracted zero force values for k_0^+ and k_0^- in Table 2.1(a).

dent on F. We model these rates using the Bell model, $k^{\pm}(F) = k_0^{\pm} e^{-\frac{F_{\parallel} d_{\parallel}^{\pm}}{kT}}$ where $F_{\parallel}$ is the component of the force parallel to the microtubule, k_0^+ and k_0^- are the forward and backward rates at $F = 0$, and the transition state distances $d_{\parallel}^+$ and $d_{\parallel}^-$ are defined in Fig. 2.4B. We can rewrite the arguments of the exponentials as $k^{\pm}(F) = k_0^{\pm} e^{-\frac{F d^{\pm}}{kT}}$, defining effective distances $d^{\pm} = d_{\parallel}^{\pm} |F_{\parallel}/F|$.

We obtained $|d^+| = 1.4$ nm and $|d^-| = 1.6$ nm (Fig 2.9), by fitting the average velocity as a function of force [8] with $\overline{v}(F) = (k^+(F) - k^-(F))$ subject to the constraint $|d^+| + |d^-| = 2.9$ nm [8], assuming that $d^{\pm}$ are independent of F. Note that $|d^+| + |d^-| = 2.9$ nm is different from the mean step-size (8.2 nm) of K-1, because the force transmitted to the kinesin heads is not parallel to the direction of the motor movement (Fig 2.4A), so $d^{\pm} = d_{\parallel}^{\pm} |F_{\parallel}/F| < d_{\parallel}^{\pm}$. An alternate mechanism for $|d^+| + |d^-|$ being less than 8.2 nm has been proposed elsewhere [93]. Similarly, the F-dependent detachment rate is taken to be $\gamma(F) = \gamma_0 \exp\left(\frac{F_{\perp} d_{\gamma}}{kT}\right)$, which we

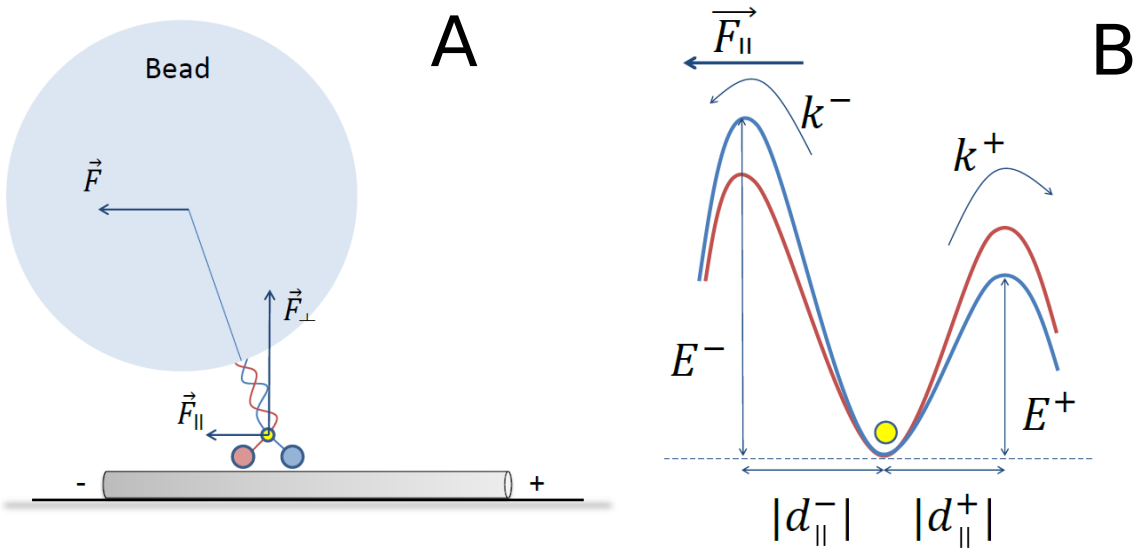


Figure 2.4: Model for force-dependent of kinetic rates.
A: Decomposition of the resistive optical trap force F applied to the bead attached to the coiled-coil, along components parallel ($F_\parallel$) and perpendicular ($F_\perp$) to the MT axis [25]. **B:** Energy landscape for forward and backward rates with the blue and the red curves corresponding to zero and non-zero F respectively. $F_\parallel$ increases the backward rate and decreases the forward rate; $F_\perp$ increases the detachment rate.

rewrite as $\gamma(F) = \gamma_0 \exp\left(\frac{|F|}{F_d}\right)$ where $F_d = \frac{|F|kT}{F_\perp d_\gamma}$ ($\approx$ 3 pN) is the force at which the two-headed kinesin disengages from the microtubule [89]. At distances greater than the transition state distance d_γ the motor is unbound from the MT. Table 2.1 (c)–(g), listing the three rates at several values of F, shows that $\gamma(F)$ is appreciable relative to $k^+(F)$ at $F > 3$ pN, which is profound consequences on $P(v)$, as we show below.

2.4.1 Bimodality in $P(v)$ from Model 3

The normalized $P(v \neq 0)$ distributions for different F values are plotted in Fig. 2.5A using Eq. 2.2, showing distinctly the non-Gaussian behavior, in sharp con-

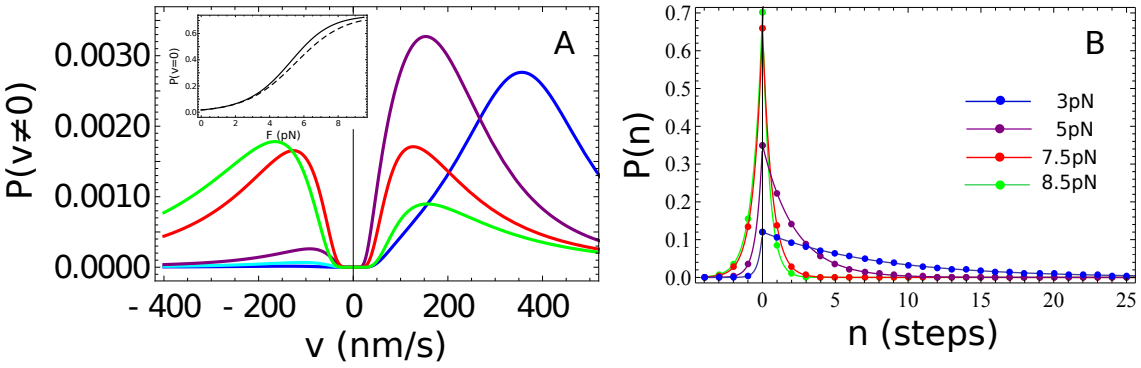


Figure 2.5: Predictions of the normalized velocity distributions ($(P(v>0)+P(v<0))/(1-P(v=0))$) (**A**) and run-length distributions (**B**) at different force values.
The inset shows $P(v=0)$ as a function of F. The dashed line is the probability (γ/k_T) that the motor detaches without taking a step. The solid line is the cumulative probability, $P(n=0)=P(v=0)=\frac{\gamma}{\sqrt{k_T^2-4k^+k^-}}$, the total number of motors detaching with zero average velocity and zero net displacement, which includes both motors that detach without stepping, and those that return to their starting location.

trast to the approximately Gaussian distribution at $F=0$ (Fig. 2.3B). By stringent standards even at $F=0$, $P(v)$ is not a Gaussian [94] but the extent of the deviation from a Gaussian increases dramatically as F increases. This happens because γ increases and eventually becomes larger than the other rates (Table 2.1), thus decreasing the processivity (Fig. 2.5B). As a result, CLT does not hold resulting in $P(v)$ to exhibit non-Gaussian behavior.

More unexpectedly, the predicted F-dependent $P(v)$s are bimodal (Fig. 2.5A). As F increases, the peak at $v<0$ becomes higher and reaches the same height as the one at $v>0$ at the stall force F_S. For forces below $F_S=7.63$ pN [8], the location of the peak of the $P(v>0)$ curves, v_P, shifts to lower velocity values as F increases, but then moves to higher ones at $F>F_S$ (v_P at $F=8.5$ pN (green curve) is larger than v_P at $F_S=7.6$ pN (red curve)).

These two counter-intuitive results are direct consequences of the discrete steps that kinesin takes on the MT. For large values of $\gamma(F)$, corresponding to large forces (where the bimodal structure is most prominent), the time the motor spends on the microtubule is necessarily small ($t \sim 1/\gamma(F)$). Since ns has to be an integer multiple of 8.2 nm, it cannot be less than 8.2 nm, implying that velocities close to zero (both positive and negative) are improbable, giving us a full explanation of the two-peak structure in $P(v)$. In addition, for $F \leq 5$ pN, v_P can be estimated using $8.2(k^+(F) - k^-(F))$ nm/s. As F increases, $k^+(F)$ decreases while $k^-(F)$ increases (Table 2.1, Fig. 2.4B), leading to a decrease of v_P with F. However, v_P cannot shift to arbitrarily low values of the velocity due to the discreteness of the step-size. As the force increases beyond F_S, for most of the trajectories that contribute to the $v > 0$ peak, the motor falls off after taking just one step (smallest n) (Fig. 2.5B), and at the same time the detachment time continuously decreases, shifting v_P to larger velocities ($v \sim \frac{1}{t}$).

The discrete nature of the stepping kinetics is less significant when a large number of n terms contribute to $P(v)$. At $F = 0$, the motor takes in excess of 50 steps of net displacement before detaching. It is then reasonable to replace the summation in Eq. 2.2 by an integral. An *ansatz* for the approximate velocity distribution, $P_A(v)$, which is highly accurate for $F = 0$ where $\gamma \ll k^+$ (Fig. 2.3B), is given by (same as Eq. A.77 in Appendix A)

$$P_A(v > 0) = \frac{\gamma}{v^2}(1+4a)^{-1/4}\left[2\ln\left(\frac{1+\sqrt{1+4a}}{2}\cdot\frac{v}{k^+}\right) - 2\sqrt{1+4a} + 2\frac{k_T}{v}\right]^{-3/2}, \tag{2.3}$$

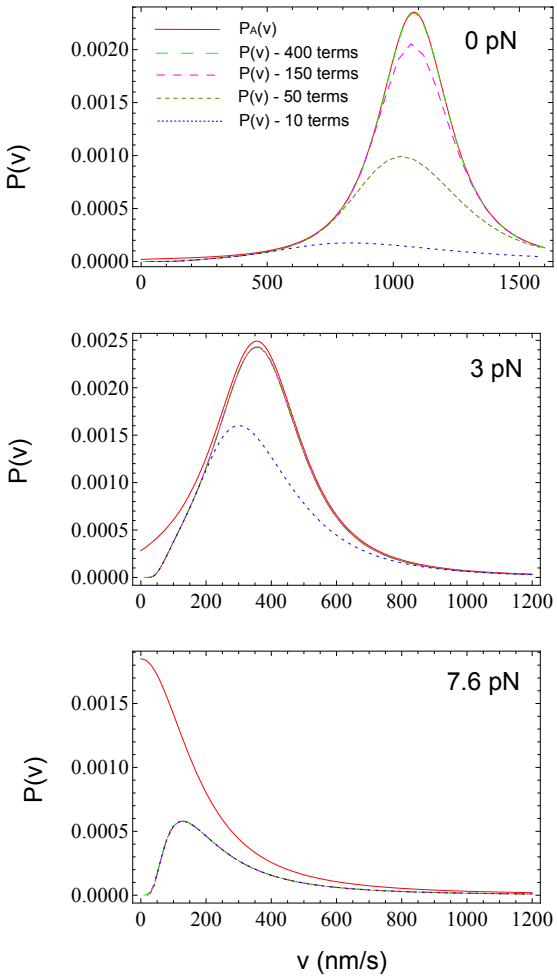


Figure 2.6: Comparison of the exact $P(v > 0)$ (Eq. 2.2 or A.63 plotted as dashed curves) and approximate $P_A(v > 0)$ (Eq. 2.3 or A.77 plotted in red) for three values of F.
As a test of convergence, $P(v)$ is plotted as a function of the number of terms used in the summation in Eq. 2.2, showing that the number of terms required to converge the sum is F-dependent. For $F = 0$ in excess of 400 terms are needed whereas at higher F converged results are obtained using very few terms. Comparison of the results for $P(v)$ and $P_A(v)$ shows agreement only for $F = 0$. There are qualitative differences between $P(v)$ and $P_A(v)$ as F increases.

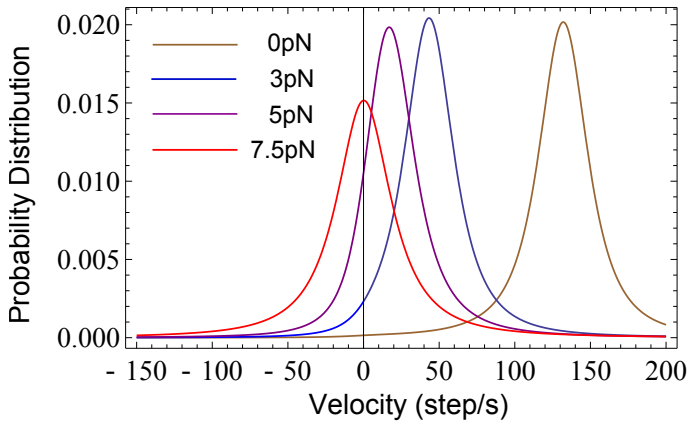


Figure 2.7: Predictions of velocity distributions at different values of loads using the approximation Eq. 2.3 or A.77.
The continuum approximation washes out the bimodal structure in the exact $P(v)$ (see Fig. 2.5A).

where $a = \frac{k^+k^-}{v^2}$. However, for $F \neq 0$ the values of γ are such that only few terms in Eq. 2.2 are non-negligible, thus making the approximate expression (Eq. 2.3) invalid (Fig. 2.6). As F increases, the average number of forward steps decreases dramatically. At $F = 3$ pN the probability of K-1 taking in net displacement for more than 10 steps is small (Fig. 2.5B). This results in qualitative differences between $P_A(v)$ (continuum steps) and $P(v)$ (discrete steps), which is dramatically illustrated by comparing Fig. 2.5A and Fig. 2.7. Thus, the discreteness of the motor step must be taken explicitly into account when analyzing data, especially at non-zero values of the resistive force.

For $F \neq 0$ the motor would take only a small number of steps because $\gamma(F)$ is an increasing function of F, raising the possibility that the predictions in Fig. 2.5A may not be measurable. To account for potential experimental limitations we calculated the conditional probability $P(v|n > n_0)$ where the first n_0 terms in Eq. 2.2 are neglected. Fig. 2.8 shows that even with this restriction the bimodal distribution

persists especially near the stall force where it is prominent.

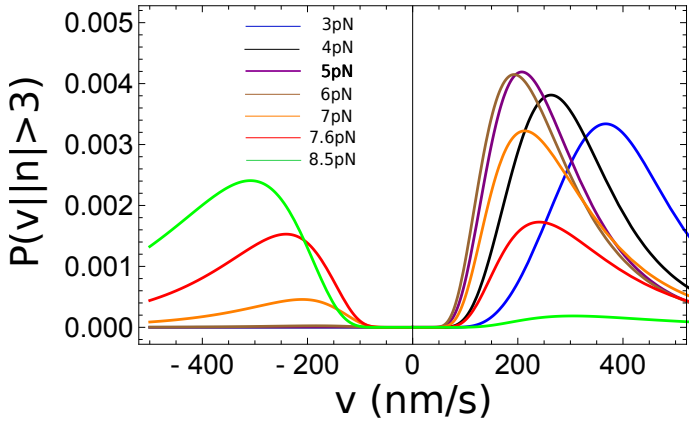


Figure 2.8: Normalized velocity distribution obtained from motors that take more than three steps.
This restriction is imposed to assess if the bimodal distribution in $P(v)$ can be measured at $F \neq 0$. The main conclusions are robust.

2.4.2 Bimodality of $P(v)$ with different force velocity curves

There are substantial variations in force velocity (F-v) curves at saturated ATP concentration for K-1 in different experiments. Even the shape of the F-v curve in [8] does not agree with the results in [9, 95]. In order to assess their impact on our predictions, we analyzed the F-v curves reported in [9, 95]. First, Fig. 2.9A shows that with a single parameter (d^+) we can quantitatively fit the three $F - v$ curves. Using this value, d^+, along with k_0^+, k_0^-, and γ_0 we calculated the dependence of $P(v)$ on F using the $F-v$ curve from [95] (Fig. 2.9B) and the one from [9] (Fig. 2.9C). The results show that, just as those in Fig. 2.5, the predicted bimodality in $P(v)$ remains regardless of the differences in the shapes of the F-v curves among different experiments.

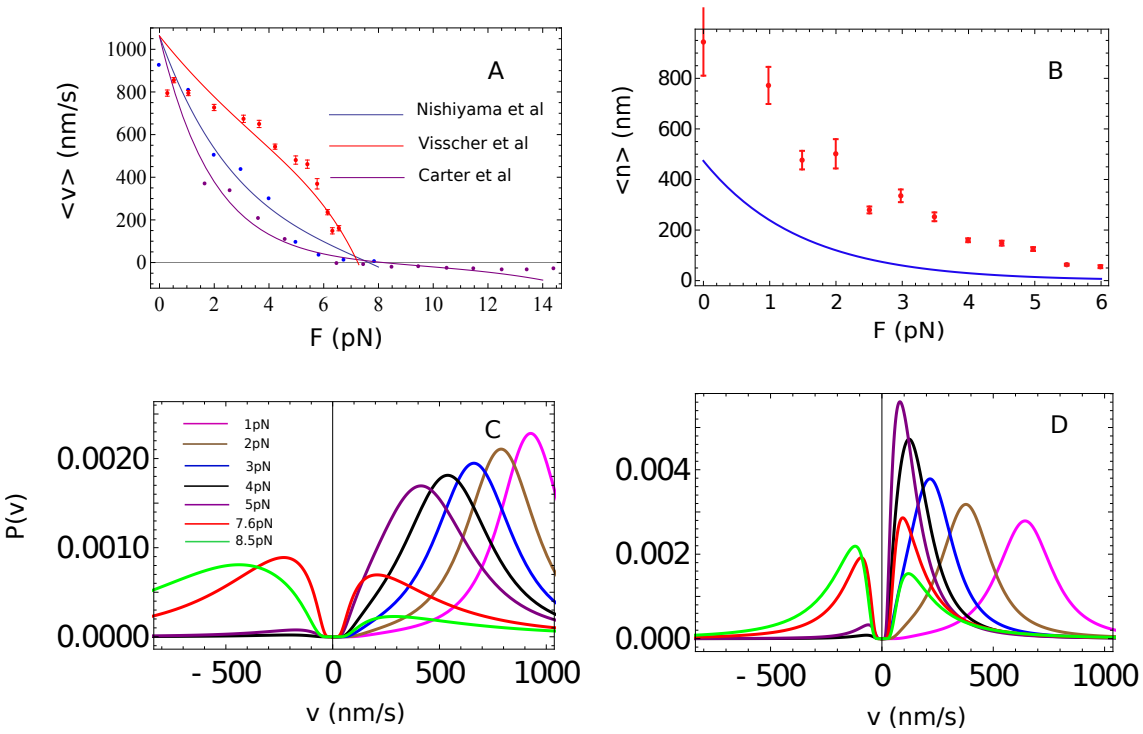


Figure 2.9: Bimodality of $P(v)$ (**C** and **D**) with different force velocity curves (**A**).
A: Force velocity curves. The dots are experimental data from [8] (Blue), [95] (Red), [9] (Purple), while the lines are fits to the data, using Eq. $\overline{v}(F) = s(k_0^+ \mathbf{e}^{-\frac{Fd^+}{kT}} - k_0^- \mathbf{e}^{-\frac{Fd^-}{kT}})$. The characteristic distances extracted from the fit are in the Table 2.2. **B:** Force dependence of averaged run-length. The solid line is the averaged positive runlength estimated from our Model 3. The deviation in the values with experimental data (red dots) comes from the fact that the analyzed data in the model [14] is from a different system to the available force-runlength data from experiments [40]. **C (D):** Predictions for $P(v)$ at different values of loads with parameters listed in the second (third) row of Table 2.2.

2.4.3 Bimodality of $P(v)$ with variability in the step-size

Although we assumed that stepping occurs in integer multiples of 8.2 nm, experiments show that the step-size distribution has a finite but small width for K-1 [38], which could possibly affect the bimodal structure in $P(v)$ at $F \neq 0$. Since inclusion of the distribution in the step-sizes makes the model analytically

Table 2.2: Characteristic distances d^+ and d^- for the fitting lines with different data sets in Fig. 2.9A.

	k_0^+	k_0^-	$\lvert d^+ \rvert$	$\lvert d^+ - d^- \rvert$
[8]	$133.4\,\mathrm{s}^{-1}$	$0.6\,\mathrm{s}^{-1}$	1.4 nm	2.9 nm
[95]	$133.4\,\mathrm{s}^{-1}$	$0.6\,\mathrm{s}^{-1}$	0.6 nm	3.0 nm
[9]	$133.0\,\mathrm{s}^{-1}$	$0.2\,\mathrm{s}^{-1}$	2.1 nm	3.3 nm

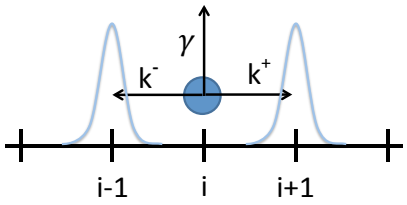


Figure 2.10: Kinetic model extended from Model 3 with a Gaussian step-size distribution.

intractable, we performed Kinetic Monte Carlo (KMC) simulations [96]. In this set of simulations, every time kinesin jumps forward or backward, s is sampled from the Gaussian distribution $N(s, \sigma)$, with mean $s = 8.2$ nm, and standard deviation σ varied from 0 nm to 8 nm (Fig. 2.10). As σ increases, the two-peak structure gets washed out (Fig. 2.11). However, for the experimentally measured value of $\sigma = 1.6$ nm, the two-peak velocity distribution is present (Fig. 2.11C) (we converted the standard error $SE = 0.03$ nm from the measurements (see Fig. 2 in [38]) on wild type kinesin LpK to $\sigma = SE\sqrt{N} = 0.03\sqrt{2993} = 1.6$ nm). We conclude that the key predictions in Fig. 2.5A are robust with respect to inclusion of a physically meaningful step-size distribution and experiments should be able to discern the two-peak structure in $P(v)$ at $F \neq 0$. The predictions for $P(v)$ at various forces with $\sigma = 1.6\,\mathrm{nm}$ are displayed in Fig. 2.12.

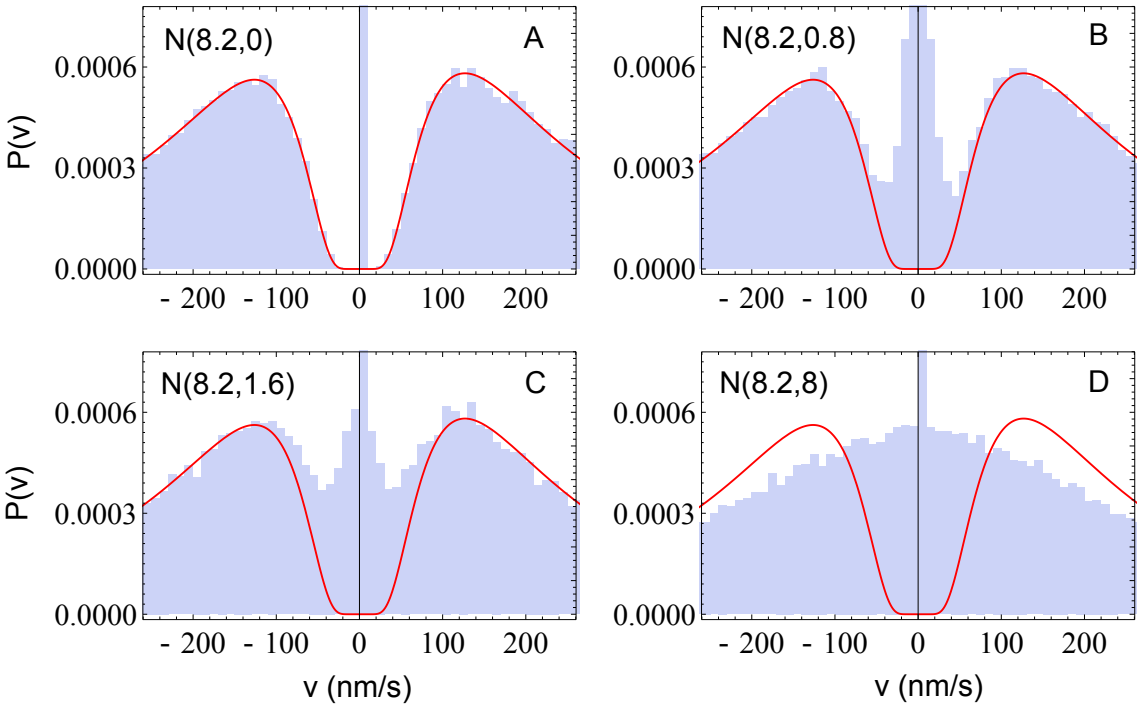


Figure 2.11: Kinetic Monte Carlo simulations (blue histograms) of the velocity distribution at 7.6 pN load, with a Gaussian step-size distribution [N(s,σ) in the left corner of each panel].
Progressive loss of bimodal nature of the velocity distribution as the width of the step-size distribution increases. At the physical step-size distribution (**C**), although inclusion of distribution in the step-size leads to minor deviations in $P(v)$, the predicted bimodality persists. The solid lines are the exact $P(v)$–Eq. 2.2 obtained by assuming that all motors take identical 8.2 nm step.

2.4.4 Bimodality of $P(v)$ with an intermediate state – Model 4

It is natural to wonder if the predicted bimodality in $P(v)$ at $F \neq 0$ is a consequence of the simplicity of the model. We now show that the presence of a chemical intermediate state, which has been invoked to analyze experiments [61], does not alter the bimodal structure in $P(v)$. We generalized a more complicated model to include a chemical intermediate (Model 4 in Table 1.4 and Fig. A.7) and computed $P(v)$ exactly (see Appendix A). Eqs. A.81 and A.82 (blue lines) are plotted

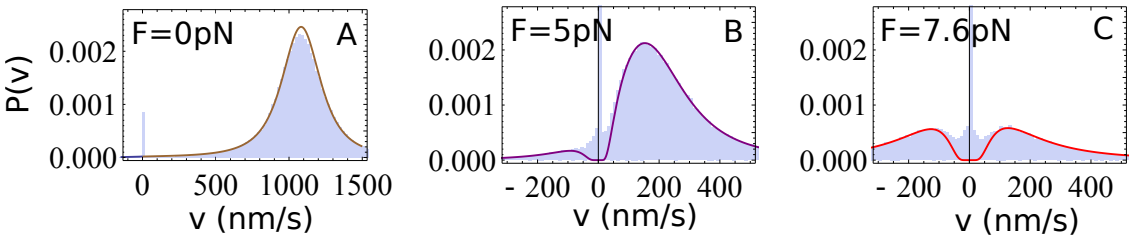


Figure 2.12: Velocity distribution simulations with step-size distribution N(8.2,1.6) with increasing load (blue histograms).
The solid lines, which are the same exact velocity distribution functions (Eq. 2.2) with no step-size distribution, fit the simulation results quantitatively.

in Fig. 2.13 to compare with KMC simulations of the same model (blue histograms). The velocity distributions Eq. 2.2 (red lines) for the Model 3 in the Fig. 2.5A are also plotted for comparison. In the intermediate state model (Model 4), the motor takes two sub-steps with rates k_s^+, k_s^-, to complete a full step, which in the Model 3 occurs with k^+ and k^- rates. Thence, for K-1 at $F = 7.5pN$ we chose $k_s^+ = 2k^+$, $k_s^- = 2k^-$ and $\gamma = 2.3\,\mathrm{s}^{-1}$ in Eqs. A.81 and A.82. The presence of the intermediate state merely makes the velocity distribution narrower without altering the bimodality in $P(v)$. We found that the important bimodal feature is still preserved, thus further establishing the robustness of our conclusions.

2.4.5 Randomness – parameter measuring the fluctuation effects

Other than $P(v)$ and $P(n)$, fluctuation effects in the reaction cycle of motors and more generally in any enzyme reaction has been succinctly captured using the randomness parameter (r) introduced by Schnitzer and Block [43]. Hence, why do we have to calculate $P(v)$ and $P(n)$ and make our predictions for $P(v)$ under force?

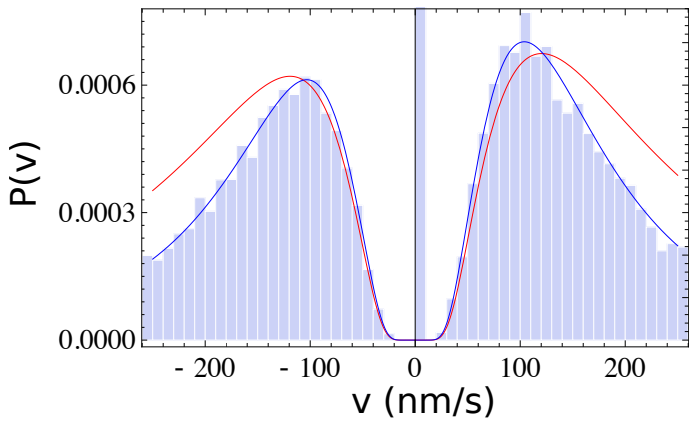


Figure 2.13: Velocity distribution simulations with intermediate state from KMC simulations (blue histograms) and the exact function Eqs. A.81 and A.82 (blue line) at $F = 7.5$ pN.
For comparison we also show the exact velocity distribution function for the Model 3 (Eq. 2.2) with no step-size distribution (red line). Although the velocity distribution becomes narrower when an intermediate state is present, the two-peak structure still remains.

For motors, r can be measured and computed from the overall motor displacement $x(t)$ using [25, 97, 98]

$$r = lim_{t\to\infty} \frac{< x^2(t) > - < x(t) >^2}{s < x(t) >} = \frac{2D}{s < v >}. \tag{2.4}$$

where D is an effective diffusion constant. If there is no dispersion in the step-size s, r can be used to obtain the second moment ($r \equiv \frac{\sigma^2(\tau)}{<\tau>^2}$) of the dwell-time distribution $P(\tau)$ with $1/r$ setting the lower bound for the number of intermediate states needed to characterize the reaction cycle [43]. If the step size is non-uniform then $r = \frac{\sigma^2(s)}{<s>^2} + \frac{\sigma^2(\tau)}{<\tau>^2}$ where $\sigma^2(s)$ is the dispersion in s [45]. Although the use of randomness parameter is not without problems [43, 98] the ability to measure and analyze r has given considerable insights into motor function. However this approach has some limitations. The parameter r only gives us information about second moment, not the full distribution. It is conceivable that different distributions could have the

same value of r, the interpretation of the results may be difficult [43, 98]. Indeed, in the oversimplified model considered in Model 3, the dwell-time distribution is $P(\tau) = k_T e^{-k_T \tau}$. Thus, r is unity. Even though $r = 1$ at all forces, the $P(v)$ distribution are highly F-dependent suggesting that r alone is not indicator of the non-Gaussian shape and bimodal structure of $P(v)$ at large forces. It should be borne in mind that measurement of r (Eq. 2.4) requires the enzymes to be processive, allowing data to be collected for a long sufficient of time. This makes randomness to be inappropriate to investigate for fluctuation of Kinesin at around and above stall forces.

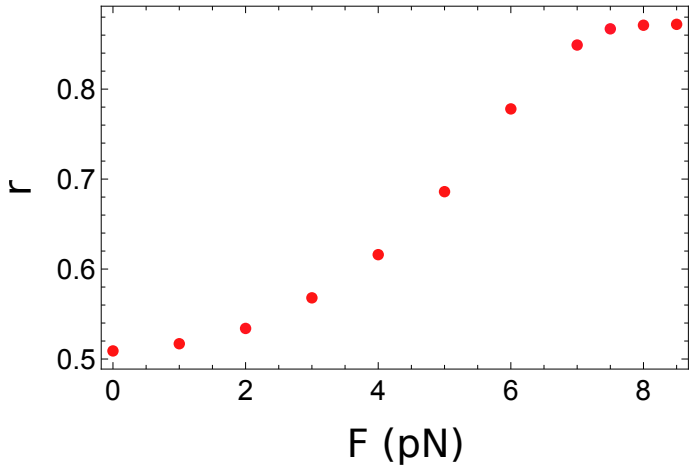


Figure 2.14: The dependence of the randomness parameter on F for the model described in Fig. S6 A.

Nevertheless, $r = 1$ for the simple model does raise a question. Is the central result is a consequence of the over simplification in Model 3? Fig. 2.13 already shows that $P(v)$ is bimodal when an intermediate is considered. We calculated the F-dependent r parameter for the more realistic model in Fig. 2.13, Model 4. The

results in Fig. 2.14 show that r increases from about 0.5 at small forces to unity as F approaches F_s. The values of r are in qualitative agreement with experiments [95] and is satisfactory given the difficulty in computing r accurately even with models with multiple intermediate states [61]. The combined results in Figs. 2.13 and 2.14 show that the central predictions in the Model 3 hold even using models that qualitatively capture the F-dependent r values.

2.4.6 Bimodality of distribution of "instantaneous" velocity v_{inst}

Our approach also allows us to compare $P(v)$ with the distribution of $P(v_{\rm inst})$ of "instantaneous velocity", an alternative measure of motor dynamics. If the dwell-time of a motor at a site is τ, then $v_{\rm inst} = s/\tau$. The distribution $P(v_{\rm inst})$ can, in principle, be computed from $P(\tau)$, the distribution of dwell-times. $P(\tau)$ has been establised generally in several important studies [99–101] of sequential models with multiple intermediate states (between 4 and 6) connecting two successive target binding states. From the numerical solution of an appropriate master equation, the F-dependent instantaneous mean velocity has been reported for K-1 without explicitly considering detachment [99]. Good agreement with experiments is found using the six state sequential model.

The measured $P(\tau)$ decays exponentially at all F (see Fig. S2 in [9]). This finding is a consequence of the absence of correlation between successive steps (dwell time >> jump time) that the motors take on the underlying lattice. If $P(\tau)$ decays exponentially, then for our model, which quantitatively describes all the experi-

mental $F - v$ curves (Fig. 2.9A using the kinetic scheme given in Model 3), the distribution of τ can be written as $P(\tau) = k_T e^{-k_T \tau}$. The dwell time distributions for forward and backward steps are, $P^+(\tau) = k^+ e^{-k_T \tau}$, and $P^-(\tau) = k^- e^{-k_T \tau}$ where, in our simple model, $k_T = k^+ + k^- + \gamma$ (all the rates are F-dependent.) This implies that the total dwell time distribution alone cannot be used to separately calculate the distribution of dwell times when the motor takes predominantly forward or backward steps without an underlying model. The distribution of instantaneous velocities ($v_{inst} > 0$) and ($v_{inst} < 0$) can be readily calculated by evaluating the integral $P(v_{inst} > 0) = \int_0^\infty P^+(\tau)\delta(v_{inst} - \frac{1}{\tau})d\tau$, leading to

$$P(v_{inst} > 0) \quad = \quad \frac{k^+}{v_{inst}^2} e^{-k_T \frac{1}{v_{inst}}}. \tag{2.5}$$

A similar expression holds for $P(v_{inst} < 0)$. Note that for convenience we measure v in unit of step/s. We can convert the unit of velocity to nm/s by multiplying it with the step-size s, which is equal to 8.2 nm for K-1. The F-dependent $P(v_{inst})$ is shown in Fig. 2.15 exhibits the bimodality found in the distribution of the physical velocity discussed in the Model 3. However, $P(v_{inst})$ and $P(v)$ deviate qualitatively as F decreases. In particular, $P(v_{inst})$ differs significantly from the expected Gaussian distribution at $F = 0$ (Fig. 2.15).

From this analysis we draw two important conclusions: (1) From the measured F-dependent dwell time distributions alone, as reported in Fig. S2 in [9], one cannot separate $P(v_{inst})$ into $P(v_{inst} > 0)$ and $P(v_{inst} < 0)$ without a tractable model that includes the possibility that the motor can detach from the track. The more complicated models [99–101] cannot produce analytic expressions for $P(v_{inst} > 0)$

and $P(v_{inst} < 0)$, which can be used to analyze experimental data. In this context, the simple but accurate model that we have proposed (Model 3), is ideally suited to interpret available data and also make testable predictions. (2) At zero F where k^- and γ are negligible the velocity distribution (dominated by positive velocities) is well-approximated by a Gaussian (see Fig. 2.3). This is not the case for $P(v_{inst})$, suggesting that v_{inst} may not be the correct measure of the motor velocity.

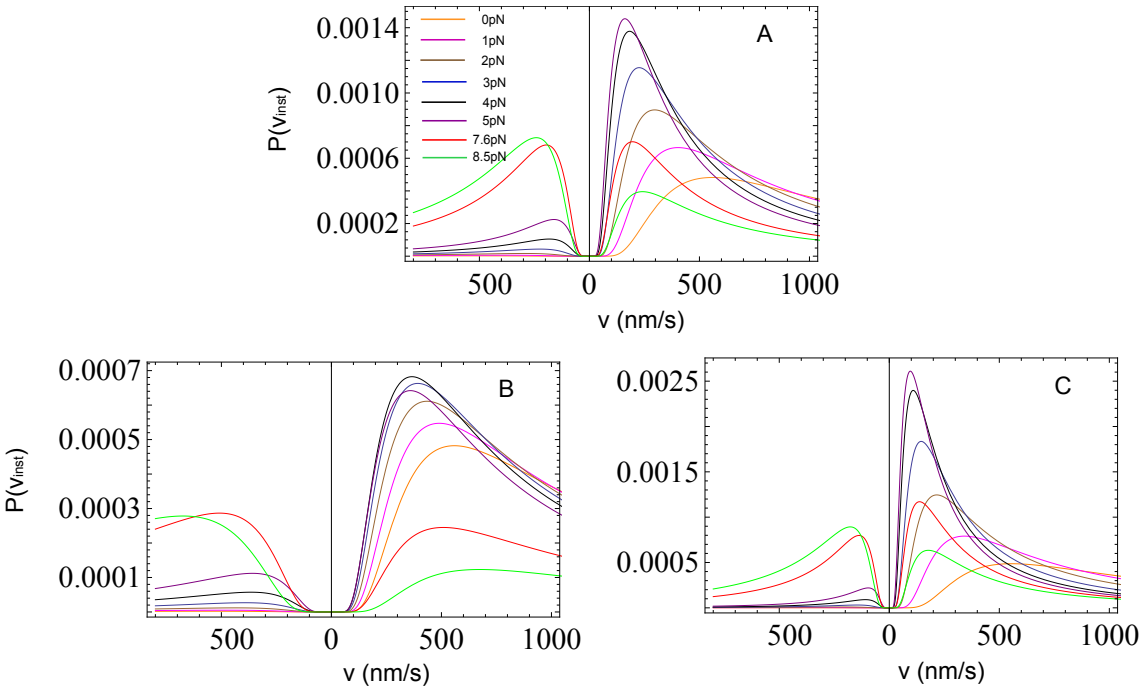


Figure 2.15: Predictions for $P(v_{inst})$ at different values of F with the parameters listed in the first row (**A**), second row (**B**), and third row (**C**) in the Table 2.2.

2.5 Conclusions

In summary, an exact theoretical analysis using a simple model for motor motility quantitatively explains the zero force velocity and run-length distributions

simultaneously for K-1, with just one physically reasonable fitting parameter. With an average run-length of ∼632 nm = 77 steps [14], we expect the velocity distribution of K-1 to deviate from a Gaussian (albeit slightly) even at zero force [94]. Based on the analysis of the zero force data, we calculated the load dependence of the velocity distribution of K-1 and discovered that the discrete nature of kinesin's steps leads to an unexpected bimodal structure in the velocity distribution under load. This surprising result can be tested in single molecule experiments, most readily accessed near the stall force where the motor has equal probability of moving forward or backward. An example of such a trajectory may be found in Fig. 1 of [9]. It remains to be seen if our predictions can be readily tested within the precision of single molecule experiments. Although set in the context of K-1, our general theory can be used to analyze experimental data for any molecular motor for which the Model 3 in Table 1.4 is deemed appropriate. Thus, our major results should hold for any finitely processive motor that takes discrete steps.

Chapter 3: Broad velocity distributions in SFIV helicases is a consequence of heterogeneity

3.1 Introduction and overview

Heterogeneity is a common and accepted theme in many complex systems, ranging from ecosystems [102], to landscapes [103], to organisms [104] and cells [105]. For an instance, within tumor, cell-to-cell variation in numerous genotypic and phenotypic traits has led to the challenging problem of drug-resistance in the field of cancer biology and cancer medicine [106,107]. Though the presence of heterogeneity is wide-spread, it is surprisingly poorly explored and established at the level of single proteins [108,109]. Perhaps this was historically a result of the 'uniqueness principle' of protein folding expounded in the seminal works of Anfinsen, which states that the functional (native) state of a protein is unique in its structure [110]. However, with the advent of high resolution techniques that allow us to probe proteins at the single-molecule level [111], we are poised at an exciting stage where we can hope to clarify the emergence of heterogenety. A number of studies have already shown a high degree of variability in the functional response of proteins, for instance in single molecule enzymes [112], in Holliday junctions [113], or Single Molecule

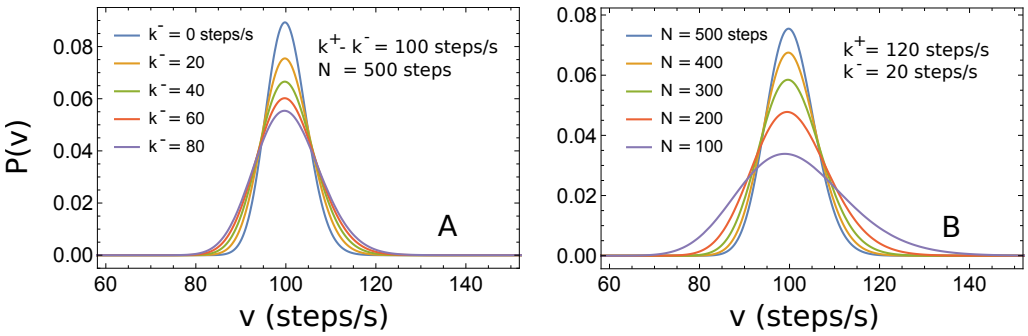


Figure 3.1: The width of the velocity distribution is a function of a number of parameters.
(**A**) Effect of changing the forward and backward rates while maintaining a fixed average velocity ($k^+ - k^- = 100$). (**B**) Effect of changing the number of base pairs used to measure the velocity distribution.

Pulling Experiments [108]. This highly variable functional response is surprising, and cannot be easily reconciled with a 'unique' native state picture of the protein energy landscape (Fig. 3.4A).

Recent experiments have suggested that there is a wide variation in the velocities of apparently identical helicases [15, 16, 83, 114–117]. Among many other physiological functions, helicases use chemical energy from NTP to translocate along nucleic acid strands and unwind them [33]. Using single molecule techniques such as fluorescence, optical and magnetic tweezers and flow-induced DNA stretching, the velocities of these motors can be tracked with unprecedented precision, providing the first glimpses of how varied the velocities actually are among supposedly identical helicase molecules [15, 16, 83, 114–117]. For instance, by measuring the time it takes for individual motors to traverse a certain length of ssDNA, Johnson et al measured the translocation velocity of T7 helicase at the single molecule level [15]. By measuring the velocity for a number of molecules, the distribution of velocity

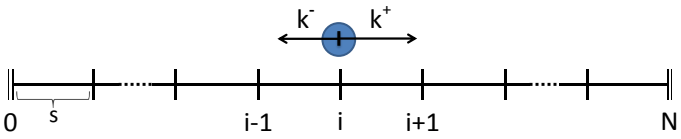


Figure 3.2: Schematic of the model.
The forward and backward rates of the helicase are denoted by k^+ and k^- respectively. The reason for neglecting any detachment from the track is explained in the text.

$P_N(v)$ was obtained. The resulting $P_N(v)$ was broad, prompting the authors to hypothesize the role of heterogeneity as a plausible explanation of the width. However, this experiment only hinted at the presence of heterogeneity, since no quantitative measure was developed for the expected width of velocity distributions of homogeneous versus heterogeneous motors. A quantitative model is required to decouple the effects of inherent stochasticity, DNA length, forward and back-stepping rate – all of which contribute to the width of the velocity distribution (see Fig. 3.1). In this chapter, we develop a theoretical model to quantitatively establish the presence of heterogeneity in two ring helicases T7 and T4 belonging to the SF-IV superfamily.

3.2 Methods

3.2.1 Kinetic model for *homogeneous* motors

In a single molecule experiment [15,118], the time t each helicase takes to traverse a fixed length of ssDNA (N_{DNA}), is recorded for each trajectory. Translocation velocity is obtained as $v_{data} = N_{DNA}/t$. Based on this definition of velocity, we now propose a simple model for the velocity distribution of helicases. As shown in Fig 3.2 (Model 5 in Table 1.4), starting from the binding site 0, the helicase can move

forward (backward) along a one-dimensional track with kinetic rates k^+ (k^-). The helicase trajectories have an absorbing state at the binding site N, corresponding to a fixed distance $N_{DNA} = N\,s$, where s is the step-size. The waiting time for backward or forward steps is assumed to be exponentially distributed. The velocity of the motor in a particular trajectory is defined as $v = N/t$, where in accord with [15], t is the time taken to travel the fixed length on the nucleic acid. (For simplicity, we measure v in unit of step/s in our calculations. When dealing with real data, all the equations are converted back to the real unit bp/s with the equation $v_{data} = v\,s$). Note that this definition of velocity assumes that all the motors can travel the fixed distance without dissociation. This is a reasonable assumption for the helicases we analyze here (T7 and T4), since the average run-length of both these helicases are orders of magnitude larger than the N_{DNA} used in the experiments. Note that unlike this model, the detachment rate was explicitly considered in our previous work on Kinesin in Chap. 2 [85]. Finally, we also note in passing that in this model there is no difference in the end rates as opposed to the "n-step" model [119–121]. This is because the helicase does not actually stop or dissociate after traveling the distance N binding sites in the experiments [15], but is simply a point where the measurement detects the presence of the helicase.

The exact analytical equation for N-step velocity distribution of homogeneous motors is given by the following expression (see Appendix A for detailed derivations):

$$P_N(v) = \left(\frac{N}{v}\right)^{N+1} \left(k^+ \exp(-\frac{k_T}{v})\right)^N \frac{1}{N!}{}_0\mathrm{F}_1\left(;N+1;\frac{N^2}{v^2}k^+k^-\right), \tag{3.1}$$

where ${}_0\mathrm{F}_1$ is a hypergeometric function [122].

Using the approximate expression for the hypergeometric function [85] ${}_0\mathrm{F}_1(;N+1;aN^2) \approx (1+4a)^{-1/4}\exp(Nc)$ where $c = \sqrt{1+4a} - 1 - \ln\left(\frac{1+\sqrt{1+4a}}{2}\right)$ and $a = \frac{k^+k^-}{v^2}$, we arrive at the following approximate expression for the velocity distribution,

$$P_N^A(v) \approx \left(\frac{N}{v}\right)^{N+1}\left(k^+\exp(-\frac{k_T}{v})\right)^N \frac{1}{N!}(1+4a)^{-1/4}\exp(Nc). \tag{3.2}$$

As pointed out in chap. 2 and [85], we have to use the exact equation if the number of steps N is small ($N \sim 10$ steps) because the discreteness of N in that case is important. In the experimental systems that we analyze here [15, 16], helicases are likely to travel $N \sim 100$ steps and hence Eq. 3.2 is a very good approximation for the exact Eq. 3.1.

In a recent work, Chakrabarti *et al.* [23] (CJT) developed a method based on Betterton and Jülicher's models [47, 51], to analyze the force dependence of the mean velocity and mean runlength of helicases. By analyzing both force - unwinding velocity and force - runlength curves simultaneously, they were able to accurately extract kinetic parameters such as step-size, interaction range, and forward/backward rates at saturating ATP conditions (See Appendix C and the original paper). With the kinetic parameters extracted using the CJT theory, Eqs. 3.1 and 3.2 allow us to predict the velocity distribution of the systems without heterogeneity.

3.2.2 Motors with *heterogeneous* rates

In this section, we build on the model discussed above, to allow for heterogeneity or variation in the forward rates of the motors. The function of a number of molecular motors demands directed motion along linear tracks, which is charac-

terized in models with much larger forward rates k^+ as compared to backward k^- or detachment rates. This is actually the common case, for instance T7 gp4 [23] or T4 gp41 helicases. It is found that k^- is roughly hundred times smaller than k^+ (see Appendix C). Therefore, although it is reasonable to think about heterogeneity in all of the parameters of the model, the dispersion in the forward rate should be the main contributor to the heterogeneity in the velocity of the motor. So we calculate the velocity distribution where the forward rate k^+ is normal distributed with a mean of μ_k (which equals to the mean forward rate estimated from the CJT theory), and a standard deviation (s.d.) of σ_k (the only fitting parameter), which accounts for static disorder in the forward rate (Model 6 in Table 1.4). With this addition, the velocity distribution becomes,

$$P_N^H(v) = \int_{-\infty}^{\infty} \mathrm{d}k^+ P_N(v) \frac{1}{\sqrt{2\pi}\sigma_k} \mathrm{e}^{-\frac{(k^+-\mu_k)^2}{2\sigma_k^2}}. \tag{3.3}$$

Here, after doing some approximations (see Appendix A for details), we obtain an analytical result $P_N^{H,A}(v)$ as

$$P_N^{H,A}(v) = \frac{\sqrt{N}}{\sqrt{2\pi v}} \frac{\mathrm{e}^{-\frac{N(k^-+v-\mu_k)^2}{2(N\sigma_k^2+v(2k^-+v))}}}{\sqrt{\frac{N\sigma_k^2+v(2k^-+v)}{(2k^-+v)}}}. \tag{3.4}$$

Comparision of this approximation function (orange line) with the numerical result (brown dots) of the exact integral, displayed in Fig. 3.3A shows that this is a very good approximation, when the step distance N is large ($\sim$ 50 steps) and $k^+ >> k^-$. When N is small or $k^- \sim k^+$, Eq. 3.3 can be numerically computed. Note that the unit of velocity in the graphs has been converted from step/s to nt/s, as usually reported in experiments.

3.2.3 A test of heterogeneity in motors with negligible back-stepping rates

We now provide a simple method to quantitatively test for heterogeneity of a motor from the measured velocity distribution, without fitting the data to a specific model. We calculate the average velocity and the width of the velocity distribution for the simplest case where the backward rate is negligible compared to the forward rate ($k^- \to 0$) (Model 1 in Table 1.4). The dependencies of the average velocity and the s.d. of $P(v)$ on the number of base pairs that the motors travel $N_{DNA} = N.s$, step size s and the *homogeneous* forward rate k^+ are,

$$< v_{data} >=< v > s = k^+ s \frac{N_{DNA}}{N_{DNA} - s}, \tag{3.5}$$

and

$$\sigma_{v_{data}} = \frac{< v_{data} >}{\sqrt{N-2}} = \frac{< v_{data} >}{\sqrt{\frac{N_{DNA}}{s} - 2}}. \tag{3.6}$$

Note that v_{data} is in bp/s and v is in step/s, s is the step-size (see Appendix A for detailed derivations).

For general processive motors with the forward rate is typically much greater than the backward and detachment rates, it is possible to measure the velocities of many motors over a fixed length specified by the travel distance, N_{DNA}. Using average velocity, $< v_{data} >$, we can calculate the expected standard deviation of the distribution $\sigma_{v_{data}}$ using Eq. 3.6, assuming that all motors are *homogeneous*. Comparison of the expected value , $\sigma_{v_{data}}$, and measured values $\sigma'_{v_{data}}$ allows us to discern the degree of heterogeneity. If $\sigma_{v_{data}}$ is smaller than $\sigma'_{v_{data}}$, then we have

heterogeneity. In other words, the stepping of each motor is distinct.

Another way to interpret Eq. 3.6 is that the coefficient of variation (ratio of the s.d. to the mean) of velocity distribution for *homogeneous* processive motors depends on the length of measuring segment as $\frac{\sigma_{v_{data}}}{<v_{data}>} = (\frac{N_{DNA}}{s} - 2)^{-1/2}$. If the experimentally measured coefficient of variation is greater than this value, there is heterogeneity.

Note that when $N_{DNA} \to \infty$, this simple test converges to the Central Limit Theorem for average of many Poisson processes ($\frac{\sigma_{v_{data}}}{<v_{data}>} \approx \frac{1}{\sqrt{N}}$). However, when a small N_{DNA} is used in an experiment (for example in FRET experiments [118]), Eq. 3.6 provides an important test of heterogeneity. If k^- is comparable to k^+, our more general theory with the exact Eqs. 3.1 and 3.3 is needed instead of the simple test.

3.3 Results

3.3.1 Heterogeneity in helicase T7 gp4

Helicase T7, which is a processive motor and rarely takes backward steps while translocating on ssDNA ($k^- \ll k^+$, see Table 3.1), is a great example to apply our model and test for heterogeneity. Applying the method for motors with negligible back-stepping rates (Section 3.2.3) to helicase T7, we found that the s.d. of $P(v)$ from the experiments $\sigma'_{v_{data}}$ is clearly larger than the predicted value $\sigma_{v_{data}}$ for the homogeneous case (Table 3.1). This easy test leads to the conclusion that helicase T7 is highly heterogeneous, which means there are motor-to-motor variations in the

stepping of T7 helicase.

A more rigorous test with the full model in Fig. 3.2, involving both the forward and backward rates (Section 3.2.1), leads to the same conclusion. Using the parameters extracted from the CJT analysis of the experimental data for helicase T7, the predicted translocation rate distributions when the system is homogeneous using Eq 3.2 (dashed green lines - $P_N^A(v)$) are narrower than the distribution observed in experiments (blue dots) (see Fig. 3.3).

To quantify the extent of heterogeneity, we fit Eq. 3.4 (orange line - $P_N^{H,A}(v)$) to the actual experimental data (blue dots) that was measured with N_{DNA}=500 nt, and extracted the s.d. of $P(k^+)$ as $\sigma_k = 32 \pm 15\,\mathrm{s}^{-1}$. The error in the fitted σ_k was obtained using the bootstrap method [10]. The authors also reported that they measured another $P(v)$ as a Gaussian $N(320, 44)$ nt/s for N_{DNA}=2800 nt, without providing the raw data [15]. Therefore, we simply fit the same equation with a generated data from the distribution that the authors reported (Fig. 3.3A inset), and ontained $\sigma_k = 22\,\mathrm{s}^{-1}$ which falls in the range for N_{DNA}=500 nt quoted above.

3.3.2 Heterogeneity in T4 gp41

We also tested for heterogeneity in the hexameric helicase, T4 gp41 [16]. Due to insufficient data for T4 gp41, we could not extract the step-size of this helicase unambiguously (see Appendix C). However, we analyzed different cases of physically reasonable step-size values and all of them show that T4 gp41 is also heterogeneous. With the quick test of $\sigma_{v_{data}}$ (Table 3.2), all of the predicted values of the standard

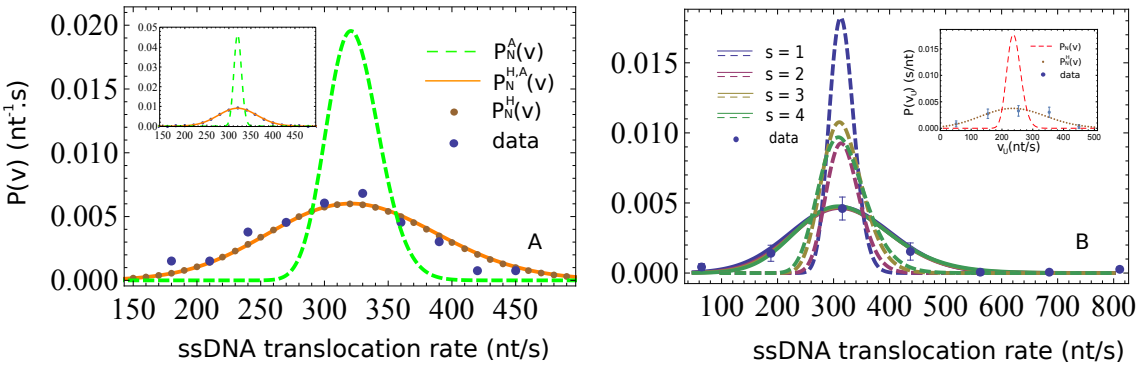


Figure 3.3: Fitting velocity distribution functions for T7 gp4 [15] (**A**) and T4 gp41 [16] (**B**) helicases's ssDNA translocation rates.
A: The rates were measured over a distance of $\sim$ 500 nt for the main graph and $\sim$ 2800 nt for the inset. The blue dots are the data extracted from the papers. (The blue dots in the inset are generated from a distribution that the authors reported as 320 ± 44 nt/s.) All the dashed lines are predicted $P_N^A(V)$s if the system is homogeneous, drawn from Eq. 3.2 with different sets of parameters in Table 3.1 (**A**) and Table 3.2 (**B**). The solid lines are the fitted to the data of velocity distribution with heterogeneity ($P_N^{H,A}(v)$ - Eq. 3.4). **B inset**: Fitting velocity distribution functions for unwinding velocity of T4 gp41 [16] using exact equations for homogeneous ($P_N(v)$ - Eq. 3.1) and fitting heterogeneous ($P_N^H(v)$ - Eq. 3.3) velocity distributions.

deviation of the velocity are smaller than the corresponding measured values. Note that we did not apply this test to the first data set when $s = 1$ because, the backward rate in that case is non-negligible, $k^- = 20\,\mathrm{s}^{-1}$. Further test with the predicted homogeneous translocation rate distributions ($P_N^A(v)$ - dashed lines) are clearly narrower than in experiments (Fig. 3.3B). The heterogeneous property is quantified as the s.d. of the forward rate σ_k in the Table 3.2.

In case the backward rate is not negligible, we can use our exact equations (Eqs. 3.1 and 3.3) in the analysis for heterogeneity We demonstrate an example by analyzing the unwinding velocity of T4 helicases, and show in the inset of Fig. 3.3B. Chakrabarti *et al.* showed that the backward rate is not negligible when the dsDNA is present. They estimated that the backward rate increases to as much as $\sim 26\%$ of the total rates when T7 unwinds dsDNA under zero force conditions [23]. With this information, and the observation that difference of the two rates is equal to the averaged velocity ($k^+ - k^- = 238\,\mathrm{nt/s}$), we estimated the forward and backward rates of unwinding helicases reported in Fig. S10A of [16] as in Table 3.2e). With these parameters, we computed the velocity distribution. If the system is homogeneous, then $P_N(v)$ (Eq. 3.1) is predicted to be the dashed line, which is narrower than the real data (blue dots). Fitting the exact Eq. 3.3 to the experimental data in this case give us a larger value of σ_k compared to the case of translocation velocity (Table 3.2e). This indicates that the presence of DNA increases the disorder in the systems, which has been reported in experiments with UvrD helicases [114].

Table 3.1: Helicase parameters for T7 helicase.

	N_{DNA} (nt)	s [23] (nt/step)	$< v_{data} >$ [15] (nt/s)	$\sigma'_{v_{data}}$ [15] (nt/s)	$\sigma_{v_{data}}$ Eq. 3.6 (nt/s)	k^+ [23] (s^{-1})	k^- [23] (s^{-1})	σ_k (s^{-1})
a)	500	2	322	62	20	161.6	0.6	32 ± 15
b)	2800	2	320	44	8.6	161.6	0.6	22

Table 3.2: Helicase T4 gp41's parameters. (a-d) are estimated parameters for translocation velocity analysis – Fig. 3.3B and (e) is estimated parameters for unwinding velocity analysis – Fig. 3.3B inset.

	N_{DNA} (nt)	s Appen. C (nt/step)	$< v_{data} >$ [16] (nt/s)	$\sigma'_{v_{data}}$ [16] (nt/s)	$\sigma_{v_{data}}$ Eq. 3.6 (nt/s)	k^+ Appen. C (s^{-1})	k^- Appen. C (s^{-1})	σ_k (s^{-1})
a)	231	1	314	83	–	334	20	87 ± 14
b)	231	2	314	83	29	157.7	0.7	41 ± 7
c)	231	3	314	83	36	104.73	0.06	26 ± 4
d)	231	4	314	83	42	78.505	0.005	18 ± 3
e)	231	1	238	21	–	366	128	103 ± 16

3.4 Discussion

Based on a simple kinetic model, we have shown that the wide translocation-velocity distributions of T7 and T4 helicases cannot arise from purely homogeneous motors in which all motors have the same velocity with a small dispersion. We built a theory based on the assumption that small differences in certain critical components in the helicase molecules might cause the large dispersion in the velocity. We modeled the heterogeneity by choosing the forward rates of different helicases from a Gaussian distribution. The mean of the Gaussian was chosen to be the forward rate obtained using analysis given in [23], and the s.d. σ_k required to explain the data gives a measure of the heterogeneity. We also proposed a method to test for heterogeneity in motors, which involves measuring the coefficient of variation of the measured velocity distribution. It is a quick test that can be readily applied to motors that rarely take backward steps.

3.4.1 Possible role of ATP hydrolysis in helicase heterogeneity

Mechanical movement of motors along a preferred direction is a consequence of structural changes that occur as a result of ATP consumption. As a result, the forward rates are larger than the backward rates in the absence of a resistive force. Possibily that heterogeneous ATP hydrolysis rates creates the wide velocity distributions observed in experiments. Indeed, we show that allowing for heterogeneity in the forward rate of our simple model (which is linked to the effect of ATP hydrolysis) is the only way of reconciling our theory with experimental data. Heterogeneity in

the diffusion (essentially backward) rate does not contribute to broadening of the velocity distribution, as shown in the Appendix C. Our theory therefore suggests that the mechanisms of NTP/dNTP binding to the motors and hydrolysis possibly play a significant role in the heterogeneity of helicases, which is supported by experimental observations bellow. There is evidence that ATP-Mg binds to and sets the initial state of RecBCD, which in turn controls the unwinding rate of the helicase [115]. The unwinding rate only changes if the ATP concentration is reset, which might change the nature of ATP-Mg binding to the helicase. The idea that heterogeneity originates from the variability of NTP/dNTP hydrolysis rates is also supported by the chemo-mechanical kinetics of hexameric helicases like T7 and T4. There is evidence to support a model where all six subunits of the helicase T7 coordinate their chemo-mechanical activities and DNA binding. Only one subunit at a time can accept the incoming nucleotide, while the rest are nucleotide-bound and coordinated with each other [123]. However, while the subunits are coordinated to translocate along ssDNA, there are many possible kinetic pathways that the subunits might follow depending on the initial conditions [124], each with its own rate. Increasing the concentration of dTTP shifts the kinetic fluxes to the high-occupancy states, and also increases the number of pathways [124]. This suggests that at saturating dTTP, as in the experiments we analyzed, these multiple pathways might lead to overall heterogeneity observed in the mobility of helicases. With some different initial conformations, or initial occupancy states, the motors can pass through different mechanochemical pathways, leading to different stepping rates. A model with three non-catalytic units in coarse-grained simulations for hexameric helicases [125] also

shows the chance the motor can take backward steps, increasing the diversity in the trajectories. Therefore, there is compelling evidence supporting our observation that a model with only heterogeneity in forward rates is sufficient to explain the wide velocity distributions of T7 and T4 helicases.

3.4.2 Static and dynamic heterogeneity and analogy to glasses

Though we have established that the behavior of T7 and T4 helicases as revealed in their velocities must be heterogeneous, we have not discussed different forms of heterogeneity. To illustrate the different forms of heterogeneity, we adopt an energy landscape perspective. Examples of the expected trajectories of a *homogeneous* system, with a single dominant basion of attraction, is shown in Fig. 3.4a. Fig. 3.4b–d shows trajectories from simulations to illustrate two fundamentally different forms of heterogeneity, both of which can lead to the broadening of the velocity distributions. As shown in Fig. 3.4c, the slope of the position-time curve of the motor might change within a single trajectory, indicating that the forward rate has changed within the course of one trajectory. This phenomenon is classified as *dynamic disorder*, when heterogeneity is recorded on the time scale of a single trajectory [112]. In the extreme case when the forward rate changes very rapidly, the overall behavior is averaged over the whole trajectory and results in *annealed disorder*, which leads to the behavior similar to the *homogeneous* case shown in Fig. 3.4b. In the other extreme limit, when the rate changes so slowly that within the time scale of the experiment, each motor stays the same over a given trajectory,

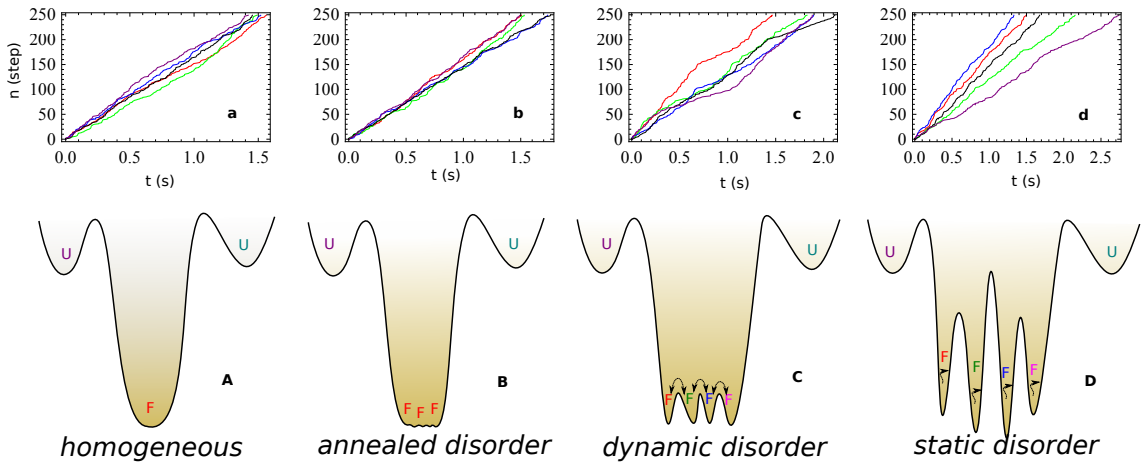


Figure 3.4: **a**–**d**: Examples of trajectories generated from KMC simulations in different energy landscapes. **A**–**D**: Schematic of the folding funnel for different scenarios in **a**–**d**.
a: *homogeneous* scenario – motors in each trajectories have the same fixed stepping rates, corresponding to a common picture of energy landscape with a single smooth basin of attraction (**A**). **b**: *annealed disorder* (similar behavior in trajectories to homogeneous scenario) – motors in each trajectories changes the forward stepping rates in every 1 binding site. The rates changes so fast corresponding to small wiggles with the barriers at the bottom of the funnel being small compared to k_BT (**B**). Each motor can pass through different states so fast, resulting in an annealing effect. **c**: *dynamic disorder* – within a single trajectory, each motor changes the stepping rate several times in the experimental time course. In this case, the energy landscape has several local minima, (**C**) corresponding to several different functional folded states (different conformations with different stepping rates). The barriers between the minima are higher than the previous case but still low enough that the system can go through several of them within the time scale of experiments. **d**: *quenched disorder* (also known as *static disorder*) – each motor has fixed stepping rates within each trajectory, but different rates between different trajectories. The corresponding energy landscape is (**D**) with several different local minima separated by high barriers corresponding to several different functional folded states (with different conformations). Within the experimental time scale, each motor is trapped in one single minimum. In this picture the motor cannot change to another functional state, unless we dramatically change the external condition (for example: temperature, ATP concentration, etc) and "quench" the system again to another functional folded state.

the heterogeneity arises purely out of different motors having intrinsically different rates. This results in different trajectories with different slopes as in Fig. 3.4d. This is an example of *quenched* or *static disorder*.

To obtain a better understanding of heterogeneity in proteins, analogies have been drawn between heterogeneity in single protein molecules and the classical dynamic heterogeneity in the theory of glasses [113, 126, 127]. In this description, the energy landscape of the protein is considered to have multiple minima. In some cases, when the barrier between these minima are small compared to thermal energy, the folded protein essentially sees just one minimum in that energy funnel and we have a *homogeneous* system with every motor functioning identically. This case is very similar to the "classic" picture in Fig. 3.4A. Even if each minima (corresponding to slightly different functions) have low barriers, the system can easily passes through those different states and we have *annealed disorder* in the system. This scenario is also very close to the homogeneous case as the function is averaged over the detailed conformations (simulations where motors have rates changing after every binding site is shown in Fig. 3.4b). In other scenarios, the barriers between the local minima are not negligible, and each folded state in each minimum has different functional parameters (for example speed of translocation/unwinding of a helicase). However, these barriers are not high enough and hence within the time scale of a single experiment, the system is able to go through several different functional states (Fig. 3.4C). In this case, we have *dynamic disorder* in the system (Fig. 3.4c). Finally, if the barriers between the conformations are high enough so that the system is not able to jump back and forth to find different conformations with different

functions (Fig. 3.4D), we have *static* (or *quenched*) *disorder* (Fig. 3.4d), similar to a glassy system [126].

3.4.3 Dynamic disorder with changing rates and broad rate distribution can also explain a broad $P(v)$

We performed several Kinetic Monte Carlo (KMC) simulations in Fig. 3.5B–D to illustrate the effects of dynamic disorder, the case when the forward stepping rate of the system varies within the trajectories. Some trajectories in these simulations are plotted in Fig. 3.4b–d. These are very simple models for dynamic disorder, and did not account for all possible cases happen in reality, but they are capable of telling us some important consequences of the dynamic disorder.

Firstly, we show that a wide velocity distribution of $P(v)$ can be explained not only by a *static disorder*, but also by a *dynamic disorder*. A simplified *dynamic disorder* simulation in Fig. 3.5B has a similar distribution with the same width (blue histogram) as we got in the *static disorder* simulation in Fig. 3.5A. The only differences in the former compared to the latter are: the forward rate is changed after the first $N_d = 200$ steps within each trajectory of $N = 250$ steps, and a wider rate distribution ($\sigma_k = 38\,\mathrm{s}^{-1}$) $> 32\,\mathrm{s}^{-1}$). This suggests that simply looking at the width of $P(v)$, one cannot be able to distinguish between *static* and *dynamic disorder*.

Secondly, in simulations reported in Fig. 3.5C, motors with the same parameters as in Table. 3.1a for T7 was performed. However, the forward rate changes at

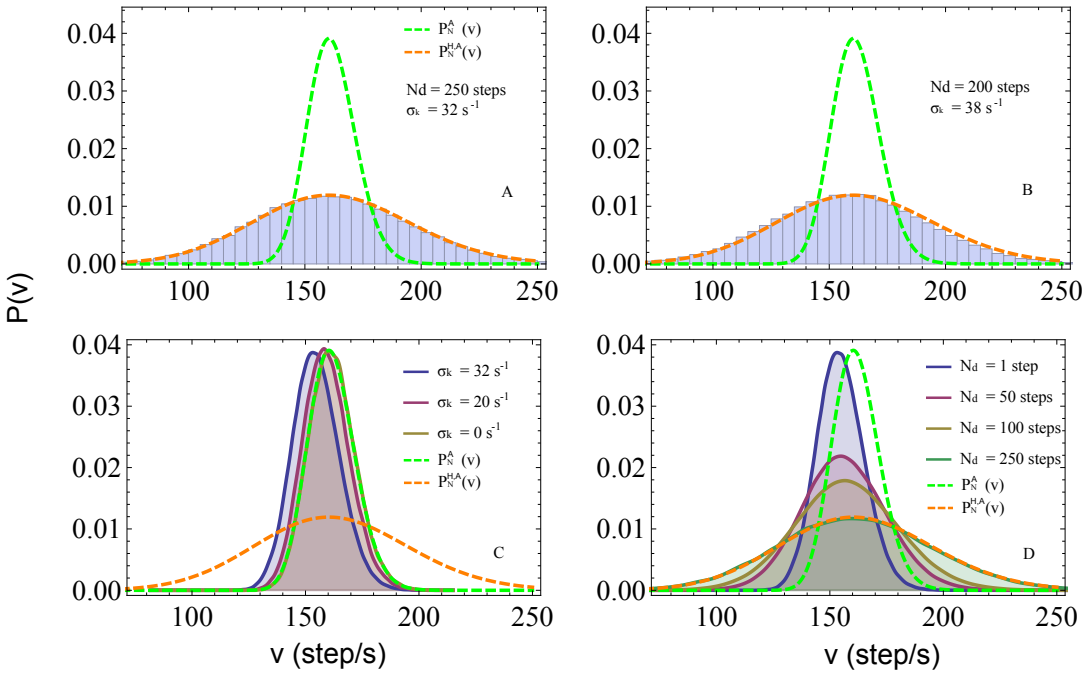


Figure 3.5: *Static disorder* vs *dynamic disorder*.
A: Comparision of approximate velocity distribution for the *static disorder* in the forward rate Eq. 3.4 (orange line) with a Kinetic Monte Carlo simulation (blue histogram). The parameters are taken from T7 data (Table 3.1a)), with the same s.d. for forward rate $\sigma_k = 32\,\mathrm{s}^{-1}$ and the forward rate does not change within each trajectory of $N = 250$ steps. **B**: A same variance of $P(v)$ can also be explained with another simplified *dynamic disorder*. The blue histogram shows a system with similar parameters of T7 from the same Table 3.1a, but after 200 first binding sites ($N_d = 200\,\mathrm{steps}$), the forward rate changes and the s.d. of forward rates increases to $\sigma_k = 38\,\mathrm{s}^{-1}$. **C**: Motors with *annealed disorder* behave similar to the *homogeneous* case. The simulations performed for the same system as in helicase T7, when the forward rate in each trajectory changes every binding site. Although the s.d. of forward rate increases, the width of $P(v)$ stays the same. The only effect is that the moves of the peak to a lower velocity. This is a consequence of the deviation from Central Limit Theorem ($N = 250\,\mathrm{steps} \neq \infty$). **D**: The more "dynamic" the system is, the narrower $P(v)$ is. KMC simulations (solid lines) of simplified *dynamic disorder* model with $\sigma_k = 32\,\mathrm{s}^{-1}$. The forward rate changes in each trajectory after every N_d binding sites. When the system changes the rate more frequently in each trajectory, $P(v)$ becomes narrower and reaches the *annealed* case (dark blue line) where there is no difference in the width with the *homogeneous* system (dashed green line).

every binding sites. The rate is drawn from a Gaussian distribution with different s.d. σ_k (same mean $k^+ = 161.6\,\mathrm{s}^{-1}$ as in T7 case). Here we can see that, as the s.d. σ_k increases, the width of the distributions stays the same. The reason is that in this case, the rates changes fast within each trajectory. As a result, there is only one average rate, as the disorder is *annealed*. The only effect that we can see is that the average velocity slightly moves to lower velocity, when the s.d. is very high.

Lastly, the main result is that even though with the same dispersion of stepping rates ($\sigma_k = 32\,\mathrm{s}^{-1}$), the width of $P(v)$ increases as the number of binding sites that the motor can take without changing the rate, N_d, increases. In the *annealed disorder* limit when the rate changes very frequently (every binding site - $N_d = 1$ step), shown in Fig. 3.5C and dark blue line in D, the width of $P(v)$ is very similar to the width of $P(v)$ in the homogeneous case (the dashed green line). As N_d increases, the width of $P(v)$ increases (Fig. 3.5D). If the rate does not change $N_d = 250$ sites, which is the measuring length of each trajectory $N = \frac{N_{DNA}}{s} = 250$ steps (dark green line) in the experiment [15], we have a histogram with a similar width as for static disorder $P_N^{H,A}(v)$ (Fig. 3.5A) (dashed orange line).

All in all, we come to a conclusion that: not only a *static disorder* model, but also a *dynamic disorder* model can account for the same dispersion in $P(v)$. However, the dispersion of the rates, σ_k, obtained with the *static disorder* model is always smaller than obtained with the *dynamic disorder*. Or, in other words, if we assume a particular system is *static disorder* (which might not be true), the value σ_k obtained using our model in this paper is the lowest bound of the true dispersion of the forward rate. The σ_k value in this case might solve as an effective dispersion

parameter of the system. This suggests to us an experimental set-up that can be used to discern between static and dynamic disorder in molecular motors in the next subsection.

3.4.4 A method to distinguish between static and dynamic heterogeneity in molecular motors from experimental data

We have demonstrated that there is heterogeneity in the translocation velocities of helicases T7 gp4 and T4 gp41, which cannot be accounted for by standard fluctuation effects. In order to quantify the extent of heterogeneity using a single parameter σ_k, we assumed that there is *static disorder* in the helicases. The value of σ_k under the assumption of static disorder has an important physical meaning. It is the "effective" dispersion of the forward as if the stepping rate of the motor does not change within each molecule. Our estimation of σ_k is an under-estimate of the actual dispersion of the rates. A large value of σ_k estimate in Eq. 3.3 would therefore indicate the presence of heterogeneity.

However, as shown in the previous section (Fig. 3.5B), a model with dynamic disorder could also explain the experimental $P(v)$. As noted before, this requires a single motor to change the stepping rate. In addition, it becomes necessary to choose a large value of σ_k. Therefore, if $P(v)$ is measured at only one value of N_{DNA}, it is not possible to distinguish between static and dynamic disorder. Here, we suggest a new extension of currently used experimental procedures, to distinguish between these two types of disorders in any motor.

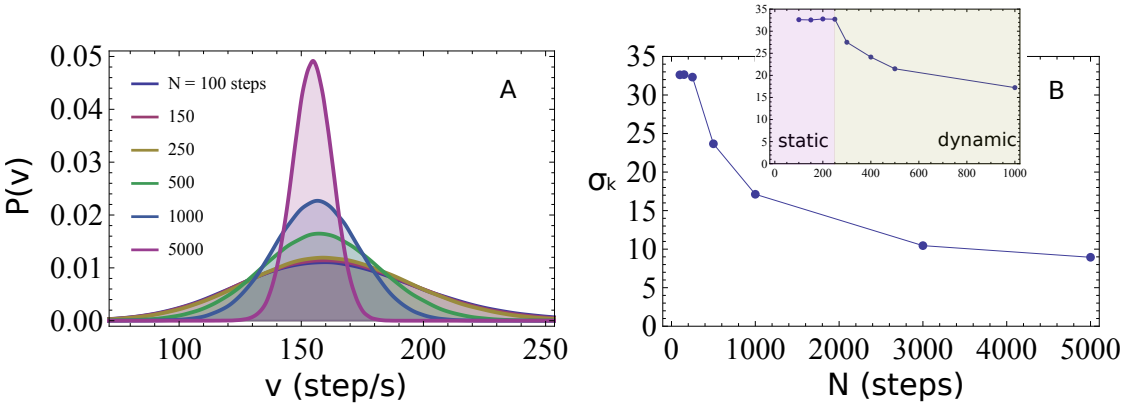


Figure 3.6: A method to distinguish between static and dynamic heterogeneities in helicases.
(**A**) KMC simulations for velocity distributions, $P(v)$, measured as a funtion of traversed binding sites, $N = \frac{N_{DNA}}{s}$. In these simulations, the forward stepping rate changes every 250 binding sites. The forward rates are drawn from a Gaussian distribution N(161.6,32) s^{-1}. Backward rate is $0.6\,s^{-1}$, which is the same as in T7 (Table 3.1A). (**B**) The variation of the effective s.d. of forward rate, σ_k^E, on N. Inset: Same graph but with a smaller scale for a better visualization of the static disorder regime.

The velocity distribution of a motor should be measured on different lengths of the same type of DNA or RNA (different $N = \frac{N_{DNA}}{s}$). Each velocity distribution $P_N(v)$ can subsequently be analyzed using our model with the assumption that there is only static disorder, and the values of σ_k can be extracted for each distribution. If a particular system has static disorder in reality, the values of $\sigma_k \equiv \sigma_k^*$ (σ_k^* – true dispersion) should not change with length N. However, if dynamic disorder in any form is in play, the value of σ_k should be smaller than the actual value $\sigma_k < \sigma_k^*$. As a result, as N increases, the disorder is likely to be dynamic, resulting in smaller σ_k values. Hence, by plotting σ_k versus N, it should be possible to distinguish static from dynamic disorder. To illustrate this method, Fig. 3.6A shows histograms of KMC simulations of a simplified motor exhibiting dynamic disorder. All the

parameters are obtained from T7 as listed in Table 3.1a), but the forward stepping rates in each trajectory can change every time the motors travels N_d =250 steps. Each velocity distribution is derived for different N values. We can see that for $N = 100, 150, 250$ steps, within each trajectory the motors does not change the stepping rate, so we have static disorder. As expected, the extracted σ_k values are the same in all 3 cases and close to the actual $\sigma_k^* = 32\,\mathrm{s}^{-1}$ used to simulate the distribution (purple regime in Fig. 3.6B inset). Once the number of steps in each trajectory increases, the dynamic disorder increases. Although using the same generated $\sigma_k^* = 32\,\mathrm{s}^{-1}$ in the simulations, the effective σ_k gets progressively smaller, indicating a transition to the dynamic disorder regime (light green regime in Fig. 3.6B inset). We expect to see the static regime in real helicase systems of $P_N(v)$ for small N and dynamic regime if N is large. Such a transition assumes that the forward stepping rate changes stochastically, an assumption that requires experimental support.

For T7 helicases a bootstrap analysis shows that $\sigma_k = 22\,\mathrm{s}^{-1}$ for $P(v)$ of $N_{DNA} = 28\,000\,\mathrm{nt}$ trajectories falls within the error range of $\sigma_k = 32 \pm 15\,\mathrm{s}^{-1}$ for $P(v)$ of $N_{DNA} = 500\,\mathrm{nt}$ trajectories (Table 3.1). This suggests that T7 helicases exhibit *static disorder*. However, we could draw a more concrete conclusion about the type of disorder if we could estimate the error in σ_k more percisely. In addition, measurement of $P(v)$ for different N_{DNA} values is a way to verify our predictions. We hope that such experiments are performed in the future, in order to better understand the nature of heterogeneity in helicases.

Chapter 4: Atomic Interactions in Kinesin Docking Process Reveal Similarity in the Neck Linker Docking in Different Kinesin Families

4.1 Introduction and overview

Conventional kinesin motors, discovered in 1985 [64] and classified as Kinesin-1 [2]), are possibly the smallest motors. They function as dimers carrying cargo by walking on microtubules (MT). Each protein consists of ∼800 amino acids including a head domain with ∼340 amino acids [128]. Two motor heads, labeled as LH (leading head) and TH (trailing head) in Fig. 4.1, are connected by a coiled coil made of 2 α-helical stalks (Fig. 4.1A). These motors transport cargos bound to the tip of the coiled coil helix as they walk predominantly on a single protofilament of the MT from the minus to the plus end [129,130]. The structural basis of how these tiny motors utilize chemical energy from ATP to generate forces has been extensively studied, although fundamental questions still remain. The key region of a 13-amino-acid connecting the motor head (MH) and the α-helix stalk, referred to as the neck linker (NL), plays a vital role in the mobility of kinesin. The structural change of the NL during the ATP cycle shows that, upon ATP binding to the leading head, the

NL goes a transition from a disordered (freely moving) state to an ordered (bound to the MH) state [70]. This docking process of the NL is sometimes refered to as a "power stroke", which helps to propel the rear head forward, initiates the stepping process (first stage in Fig. 4.1B). Furthermore, the disorder-order transition of the NL is a key factor in lowering the possibility that kinesin takes side steps [130].

Hwang, Lang and Karplus provided insights into the docking process of the NL [133] using Molecular Dynamics (MD) simulations. They showed that the NL tends to form a β-sheet with the 9-amino-acid cover strand (CS) in the N-terminal of the head (colored in red and yellow in crystal structures in Fig. 4.3). They suggested that the resulting Cover-Neck Bundle (CNB) induces a conformational bias that forces the NL into the binding pocket of the core motor head by a hinge-like action. This conformational bias might stablize the CNB by a free energy gain greater than the amount of 3 kJ/mol [134] compared to the NL docking alone. The additional gain in stablility could partially account for the estimate of the required energy of $\sim$25 kJ/mol that kinesin needs in tranversing a single step [76]. Subsequently, several experimental papers supported the importance of the CNB formation in different kinesin families [21, 22, 135–138].

Hwang *et al.* also proposed a model for structural changes that occur upon NL docking as a consequence ATP binding to the LH. At the beginning of the docking process, the CS is free. After ATP binding to the LH, the helix α6, which connects directly with NL, and α4, which belongs to the SWII cluster and lies in the MT binding interface, change their positions. They suggest that before ATP binding, the α4 is too closed to the α6, blocks right at the end of the α6, thus unwinds a part

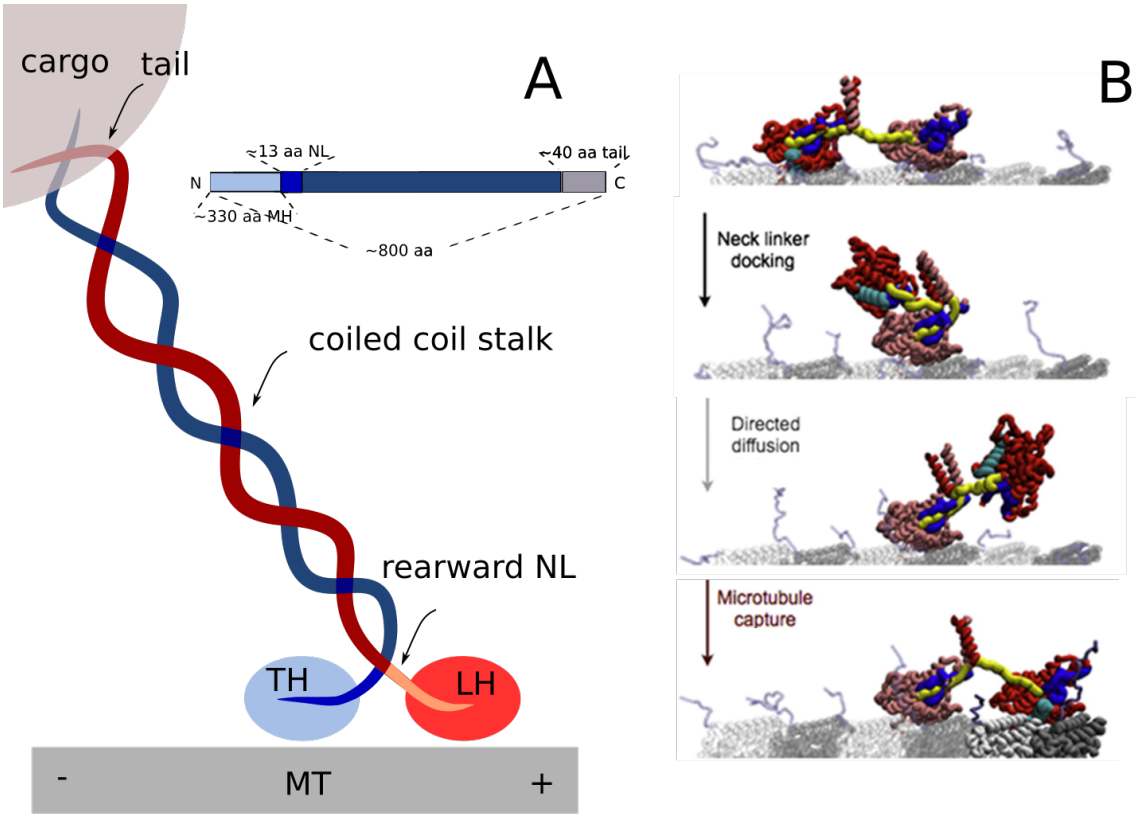


Figure 4.1: Kinesin-1 structure (**A**) and its stepping process (**B**). **A**: Model for the structure of Kinesin heavy chain (KHC) known as conventional kinesin, consisting of 2 proteins (red and blue) with a tail connected to a cargo. The two heads are bound to a microtuble at a stage right before docking. The NL of the LH points backwards towards the minus end of the MT in this state. The figure is not to scale. **B**: A kinesin's step is believed to start with a disorder-order transition of the NL (dark black arrow), followed by a diffusive search of the TH (grey arrow), and finished by a microtubule capture of the TH to the next binding site (red arrow). During the whole process, the LH is tightly bound to the MT [130, 131]. Figure is adopted from Zhang *et al* [130] (with permission) with the NLs in yellow. A single step is completed with a life time of $50\,\mu$s [132].

of $\alpha 6$ which connects directly to the NL, inhibiting the NL to be forward. Upon ATP binding, the relative positions of $\alpha 6$ and $\alpha 4$ change: the $\alpha 6$ raised up so that $\alpha 4$ is out of its end and $\alpha 6$ can form an extra turn and bring the NL to interact with the core and dock to the front. This opens the "docking pocket", bringing the NL in contact with the CS, enabling it to dock back to the binding site. Thus, the power stroke is guided by this mechanism refered to as "ATP gating" (see Chap. 1). At the atomic level, the "docking pocket" lies between Ile9 at the end of the CS and Leu268 at the C terminal of the $\alpha 4$ helix. The pocket has to be opened to create a cavity, allowing the first residue of the NL (Ile325) to be buried, bringing the NL in contact with the CS [19, 41, 71, 73] (In Fig. 4.2, the alignment between 1BG2 structure and our structure at the beginning of our simulations shows that, $\alpha 4$ at the beginning of our simulation (dark blue helix) is clearly lower than in 1BG2 structure(light blue helix). Thus, residue Ile325 of $\alpha 6$ is allowed to be buried between residues Ile9 of the CS and Leu268 of the $\alpha 4$ (right box Fig. 4.2), whereas part of the $\alpha 6$ is unwound and the residue Ile325 is protruded away outside of the closed cleft between Ile9 and Leu268 (left box Fig. 4.2).) Although there are agreements about the link between NL dockings and ATP gatings, a clear picture of the NL docking process with the CNB formation at the molecular level has not been well established.

Detailed conformational changes that the NL undergoes from a disordered to an ordered state are not yet fully established at the molecular level. The NL is small (only 13 amino acids), which means atomically MD detailed simulations could be useful in yielding insights into some aspects of the docking process. The processivity of Kinesin-1 ($\sim$100 steps before detaching from the MT [14, 139]) is

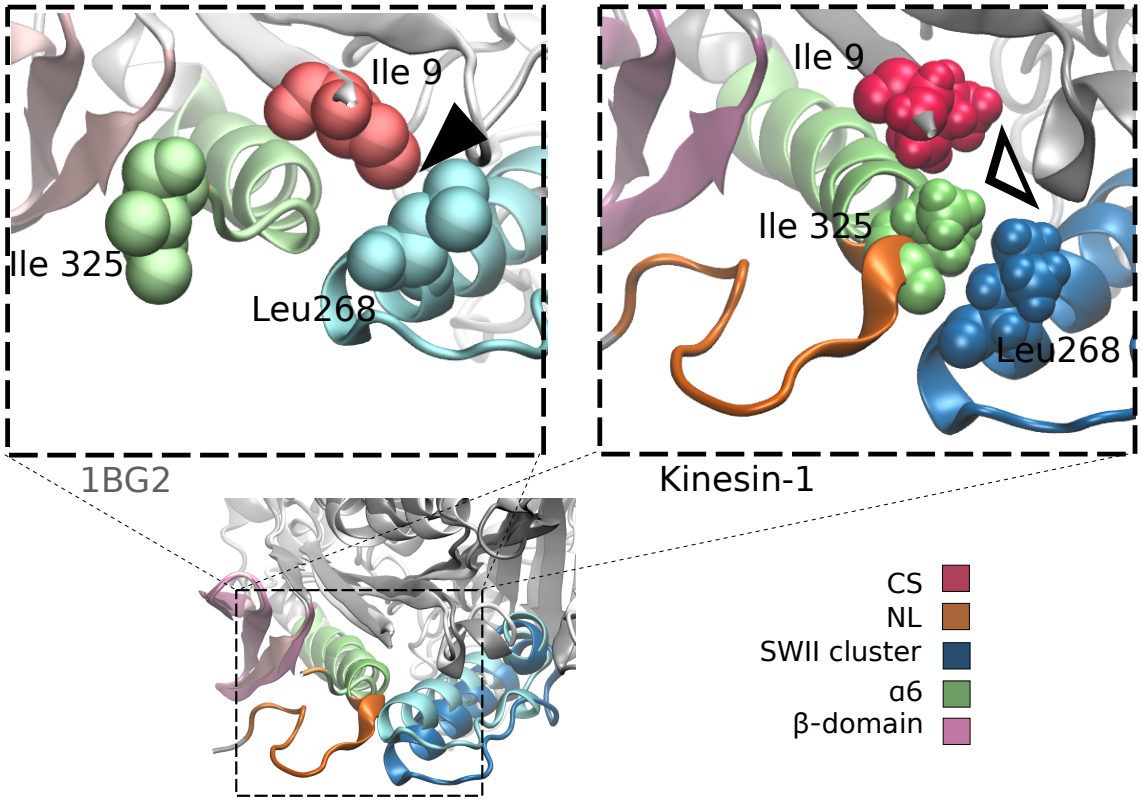


Figure 4.2: Closed (**left**) vs open (**open**) "docking pocket" of a Kinesin-1 in crystal structure 1BG2 with a "disorder" NL [17] (**left**) vs at the beginning of our simulation (**right**).
Kinesin-1 in our simulation is in darker colors, compared to 1BG2 structure. Residues Ile 9 (CS), Ile325 (NL), and Leu 268 ($\alpha 4$) are represented as balls. **Bellow:** Alignment of the two structures showing the difference of $\alpha 4$ (blue, lighter in 1BG2, darker in our simulation) and $\alpha 6$ (green, again, lighter in 1BG2, darker in our simulation) relative positions. In all figures, parts of the CSs of both structures has been truncated for better visualizations.

accomplished by a very short stepping time, compared to the dwell time. Thus, the actual time one of the motor heads remains detached from the MT is very short compared to the time the two heads are bound. In particular, Kinesin-1 takes ∼100 steps every second [8, 9, 140] under zero load, leading to a dwell time in the range of 10 ms. However, the life time for the detached state of the TH during one step is $\sim 50\,\mu$s [132]. Thus, the docking process is even faster ($\sim 20\,\mu$s [141]). Unfortunately, with the time window of a few μs, the docking process is too fast to be directly measured in experiments, and too slow to be simulated in atomically detailed MD simulation.

In this study, we provide, for the first time to our knowledge, a comprehensive atomically detailed view of the docking process in kinesin in order to gain insights into the movements of the NL and the CS during the process. In order to do so, we need to start with a conformation of a motor head as it is in the leading head (LH) position right before the kinesin takes a step (the red head in Fig. 4.1A). In such a conformation, the NL is undocked and pointed backwards (light pink) as it connects to the coiled coil stalk and the NL of the trailing head (TH – blue) attached to the MT. However, crystal structures describing such a waiting state are not available. All the crystal structures with a clear ordered NL were captured in an "ATP-like state", which is believed to be the state of LH after docking. Note that, such conformations were observed with either ATP analogues or ADP bound to the head when it is not bound to the MT [69, 70, 142].

To create the conformation of the LH right before docking occurs, we pulled the NL of an ATP-like state structure (pdb 1MKJ [69]) completely backwards by

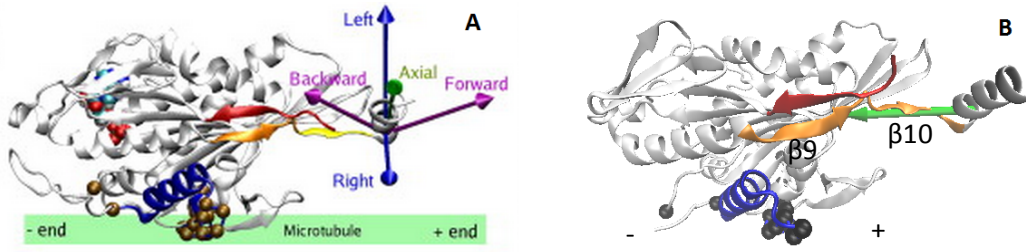


Figure 4.3: X-ray crystal structure of kinesin with pulling directions of the neck-pulling simulations.
A: 2KIN crystal structure with five pulling directions performed in Hwang's paper [133] (The figure is adopted from their paper with permissions from the journal) **B**: 1MKJ crystal structure with a backward pulling direction (green arrow) we performed to get a more realistic starting conformation for docking simulation. CS (red), NL (orange), $\alpha 4$ (blue) and anchored atoms (brown sphere) in MT binding domains are colored similarly to those in the reference paper.

180 °. Note that the direction of the applied force here is different from the upward or partially upward applied forces used previously (Fig. 4.3). Starting from an ATP-like state ensures the "docking pocket" is open at the beginning of the simulation, *open* the "ATP gate" (Fig. 4.2). The pulled conformations were allowed to relax, in order to observe whether or not the docking process spontaneously occurs, as in experiments (for details, see the Methods section).

With the help of the resources made available through the Anton supercomputer, we were able to partially overcome the time scale limitation, allowing us to observe the NL docking process in one μs simulation. This allowed us to analyze the interactions of the NL and CS with the other residues in the head. The main results are: (i) the NL interacts with the domains in the back at the beginning of the docking process by interactions between hydrophobic residues; (ii) interactions

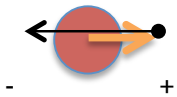
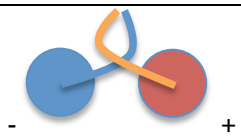
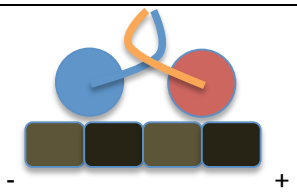
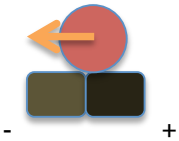
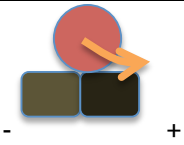
with the CS enable the NL to dock; (iii) the NL can form an extra turn and become an extended alpha helix of α6, resulting in shortening of the NL when the tension is low; (iv) NL docks to a new binding site adjacent to the MT-binding site, which is distinct from the conformation observed in crystal structures. Using these observation, we suggest some additional roles of NL during the stepping of kinesin (Fig. 4.1B).

4.2 Methods

4.2.1 All atomic molecular simulations

We deleted the coordinates of the α7, which is the N-terminus of the coiled coil connected with β10 of the NL, from the ATP-like crystal structure 1MKJ of human Kinesin-1 motor domain [69] to reduce the system size. We added the coordinates of the L11 loop, which is a missing loop for the MT binding pocket. The system is equilibrated and solvated in explicit water. We then performed pulling simulations to prepare an initial conformation (Table 4.1a)). A constant force (460 pN) was applied to the last atom of the NL, and we adopted the same constraints applied to C_α atoms in the MT binding domain (AA 142-145, AA 273-281, AA 238 and AA 255) as in the previous study [133]. The pulling direction for the NL was chosen to be the direction from N338 C_α to T334 C_α, pointing backward, along the direction of β10 of the NL (Fig. 4.3). The simulations were performed using NAMD version 2.9 [143] with the CHARMM27 (CHARMM22+CMAP) force field [144] at 310K until the NL was completely undocked and pointing backward, as expected before

Table 4.1: Setups of the simulations done in the project. The MHs are represented as circles. The NLs are lines. The tubulins are rectangles. The arrow shows the relative direction of the applied force.

	Starting conformation	Simulation time	Results	Note
a)		Until the NL completely rearward	Got the initial conformation for b) and c)	Fig. 4.2
b)		1μs	See main results in the main text	Fig.s 4.5, 4.6, 4.7, 4.8, 4.10, 4.11, 4.12
c)		10ns	Confirm the NL_3 conformation during the stage I	Data not shown
d)		10ns	Confirm the NL_3 conformation during the stage I	Data not shown
e)		10ns	Confirm the NL_3 conformation during the stage I	Data not shown
f)		1μs	NL becomes an extension of α6	Fig. 4.9
g)		10ns	NL_2 conformation is stable within the time scale	Fig.s 4.13, 4.14

the docking process.

Starting with the NL that is completely pointed rearward, we solvated it in a 96 Å x 96 Å x 96 Å explicit water box, added salt NaCl to a concentration of 0.15 mol/L and neutralized the system. The system has a total of 83377 atoms. We observed the time evolution for 1 μs (Table 4.1b)) after releasing the force. Thus, our model probes unbiased dynamics of the docking process. The main results

from the simulations are presented in this chapter. Similar set-ups have been run also on clusters (Deepthought1 in UMD), each for a much shorter time scale: 10 ns (Table 4.1c))

When we carried out this work, the Kinesin-MT complex structure was not available. So we created the MT-Kin structure by aligning the main head with the bovine tubulin crystal structure 1JFF [145] using the complex structure 2P4N [146]. For the 1μs simulation with the rearward pointing NL, after fully solvating the system with the added salt, the system with 116031 atoms was simulated on the Anton supercomputer (Table 4.1f)) and other supercomputing clusters (Table 4.1g)). We also adopted Kinesin dimer structures and kinesin dimer-bound MT structures that had been built in a previous work in the group [147] to conduct several 10 ns simulations on clusters (Table 4.1d) and e)).

4.2.2 Data analysis

Correlation functions (Fig. 4.5A) were used to analyze the NL motion and orientation. Running averages of various quantities were followed to show the evolution of the process over time. We defined a unit vector $\widehat{\vec{u}}$ along the direction of the NL, from C_α of residue 337 to C_α atom of residue 325, throughout the simulation,

$$\widehat{\vec{u}} = \frac{\vec{r}_{C_\alpha^{325}} - \vec{r}_{C_\alpha^{337}}}{|\vec{r}_{C_\alpha^{325}} - \vec{r}_{C_\alpha^{337}}|}. \tag{4.1}$$

Let $\widehat{\vec{u_0}}$ be the same unit vector in the docked state from the crystal structure (See diagram in Fig. 4.4). If the NL is fully docked, the product

$$f_{NL(t)-NL_0} = \widehat{\vec{u}}.\widehat{\vec{u_0}} \tag{4.2}$$

is equal to one and when the NL is pointing rearward as in the starting conformation of our simulation, $\widehat{\vec{u}}(t=0).\widehat{\vec{u_0}} = -1$. Ideally, f_{NL} should be averaged over multiple trajectories to draw statistically meaningful conclusions.

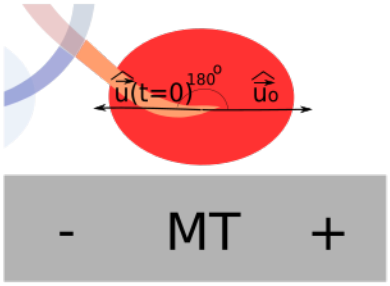


Figure 4.4: A diagram of unit vectors that used to quantify the orientation of the head.
Here the unit vector of NL $\widehat{\vec{u}}(t=0)$ is opposite to the unit vector of NL $\widehat{\vec{u_0}}$ in the crystal structure. The LH is shown in pink and NL is in orange as in Fig. 4.1.

The running time average function (Fig. 4.5B) is defined as

$$\bar{f}_{NL(t)-NL_0} = \frac{1}{t}\int_0^t f_{NL(s)-NL_0} ds. \tag{4.3}$$

Similar functions $f_{CS(t)-CS_0}$ and $\bar{f}_{CS(t)-CS_0}$ were also calculated for CS with a unit vector $\widehat{\vec{v}} = \frac{\vec{r}_{C_\alpha^1} - \vec{r}_{C_\alpha^9}}{|\vec{r}_{C_\alpha^1} - \vec{r}_{C_\alpha^9}|}$. We also quantitate the relative orientation between the NL and the CS at a particular time by calculating $f_{NL(t)-CS(t)} = \widehat{\vec{u}}.\widehat{\vec{v}}$ and $\bar{f}_{NL(t)-CS(t)} = \frac{1}{t}\int_0^t f_{NL(s)-CS(s)} ds$.

Fig. 4.8 shows the main interactions that the NL and CS have during the three different stages. Hydrogen bond interactions are assigned between residues if the distance between the donor atom and the acceptor atom is smaller than 3 Å and the angle D-H-A is smaller than 20 degrees. Hydrophobic interactions are assigned to any two hydrophobic residues that are within a distance cutoff of 3 Å.

Table 4.2 shows the list of PDB structures used in this work. The structure alignments between different structures were performed using MultiSeq [148] within VMD [149]. Graphs are rendered using Tachyon [150].

Table 4.2: Table of crystal structures used in this work.

	pdb ID	NL conformation	appeared in	reference	note	NT state
K1	1BG2	disorder	Fig. 4.2	Kull1996 [17]	crystal structure	ADP bound
K1	2KIN	NL1	Fig. 4.3	Sack1997 [67]	crystal structure	ADP bound
K1	1MKJ	NL1	Fig. 4.7	sindelar2002 [69]	crystal structure	ADP bound
K5	1II6	NL3	Fig. 4.15	Turner2001 [20]	crystal structure	ADP bound
K5-MT	4AQW	NL3	Fig. 4.16	Goulet2012 [21]	Fitting of 1II6 and 3HQD with EM	rigor (no NT)
K5-MT	4CK7	NL2	Fig. 4.17	Goulet2014 [22]	pseudo atomic structure fitting with EM	ADP.AlFX bound

4.3 Results

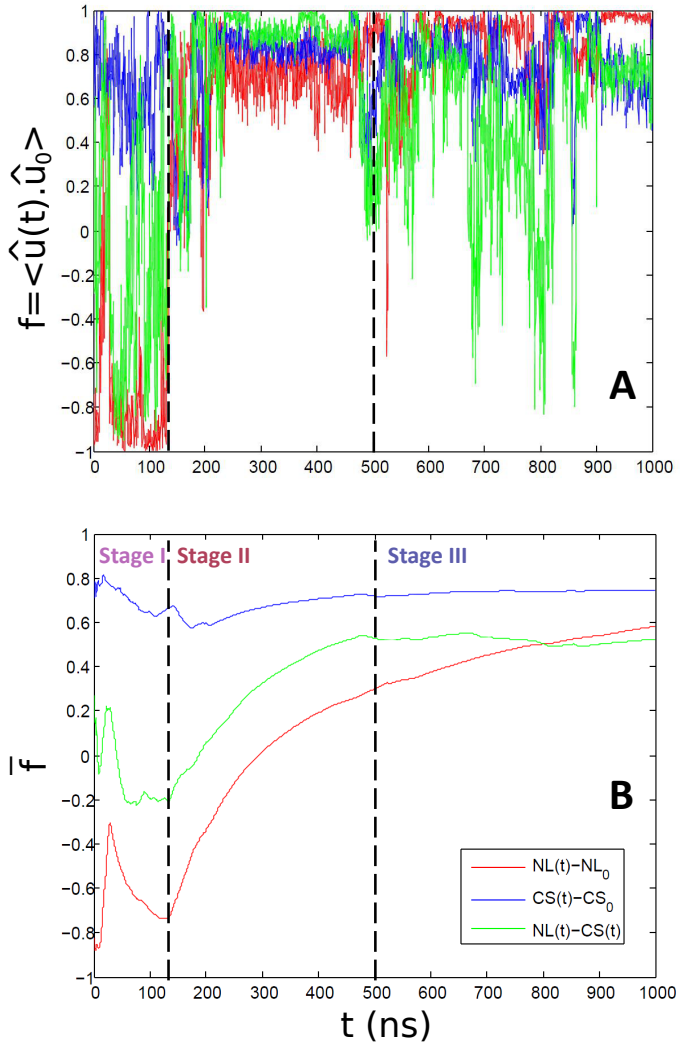


Figure 4.5: Angle correlation functions (**A** – Eq. 4.2) and running average functions (**B** – Eq. 4.3) of directions of NL (red line), CS (blue line) ($u(t)$) vs the docked directions in the original crystal structure (u_0) and between NL and CS (green line) during 1 μs simulation.
Changing of NL's functions (red line) from around -1 to +1 shows the docking process. The close to 1 value of the correlation between the NL and CS (green line) in the middle of the simulation shows the CNB did form for a short time. Dashed lines divide the trajectory into 3 main stages. See the text for details.

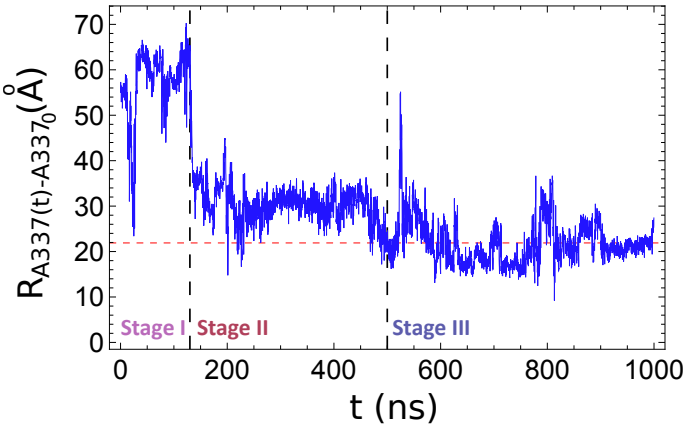


Figure 4.6: Distance between the residue $A337_0$ at the crystal structure (1MKJ) and the same residue $A337(t)$ along the simulation ($A337$ is the position where the force was applied at during the initial simulation – see Fig. 4.3).
The vertical gray lines separate the three main stages as in Fig. 4.5. The red dashed line is the averaged distance during the stage III (21.9 Å). The distance does not reach zero at the end of our simulation, indicating that the NL does not come back to the binding pocket as observed in the crystal structure.

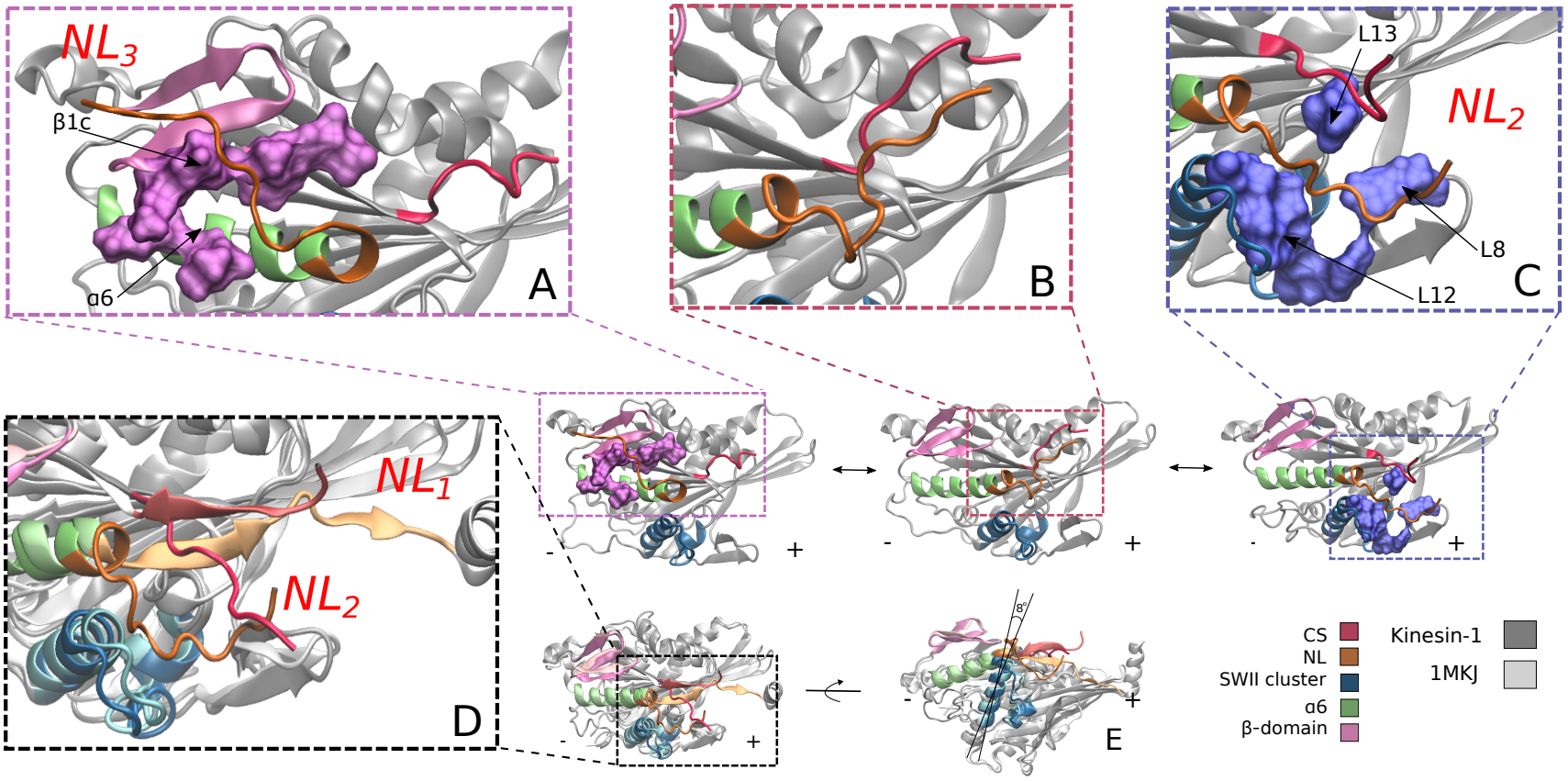


Figure 4.7: Three snapshots from the 1 μs simulation representing the three main stages.

Continuing caption of Fig. 4.7 **A**: Stage I, where the NL mainly interacts with the α6 and the β1-domain in the back region of the head while the CS is "free". The residues that interact with the NL during this stage (Stage I in Fig. 4.8) are shown as a purple surf surface. *Caption is continue on the next page.* **B**: In stage II, the CS interacts with the NL, facilitating its motion forward. The CNB is formed between the β10 and the CS, and the docking process starts. The NL starts to form an extra turn at the end of the α6. **C**: In stage III, the CNB is disengaged. The NL points forward but moves down to interact with the residues adjacent to the MT-binding site. The residues that interact with the NL during this stage (Stage III in Fig. 4.8) are shown as a blue surf surface. **D**: Comparision of a snapshot in stage III with the crystal structure (1MKJ – lighter colors) shows the difference between the two docking conformations of NL and CS. We call the NL-CS conformations in the crystal structure NL_1, in the stage III of our simulation NL_2, and in the stage I of our simulation NL_3. **E**: Same as **D** but rotated 90°, showing an 8° tilted of α4 at the end of the simulation compared to the 1MKJ structure.

We successfully observed the docking process of the NL, which starting from a rearward direction, orients itself toward the plus end direction of the MT within the duration of the simulation (1μs – Table 4.1b)). Fig. 4.5A shows the correlations between the directions of the NL and the CS with the original directions and the correlations between each other (see the Method sections for details). Fig. 4.5B shows the average of the correlations over the trajectory. The quantity of $f_{NL(t)-NL_0}$ change from -1 at the beginning of the simulation to 1 at the end, showing that the orientation of the NL changes from the rearward direction towards the plus end direction during the whole course of the simulation. However, the NL does not come back to the binding pocket as observed in the original crystal structure (Fig. 4.6). (We refer to this NL conformation observed in crystal structures as NL_1, to distinguish with other conformations that we see in our simulation.) Based on the evolution of the functions and the distance along the trajectory, we divide the

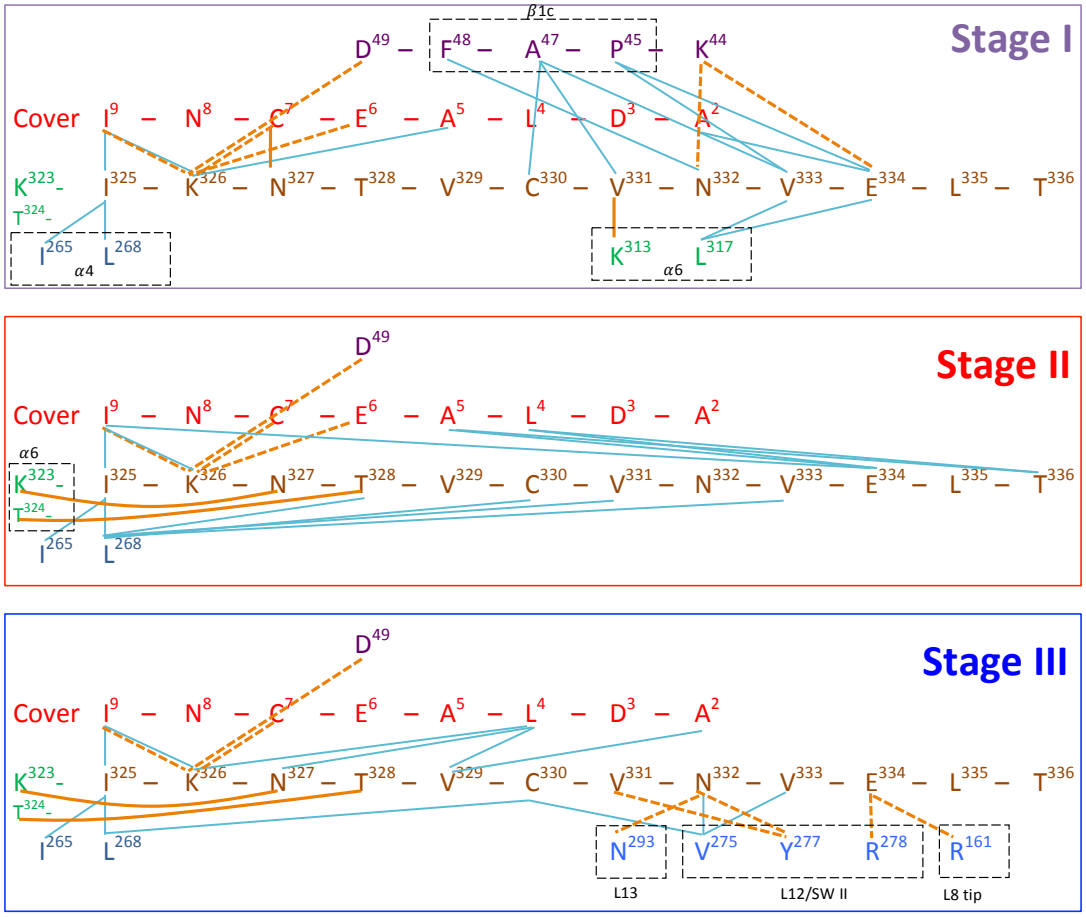


Figure 4.8: Interactions of the NL and CS highlighting the key residues are in contact during the three corresponding docking stages in Fig. 4.5 and 4.7.
Residues on NL (orange), CS (red), α6 (green), α4 (blue), β1-domain (purple) are colored similarly as in Fig. 4.7. Thick orange lines: backbone hydrogen bonds. Dashed orange lines: salt bridge between polar side chains, or hydrogen bond between polar side chain and backbone hydrogen or oxygen. Blue lines: hydrophobic contact between side chains.

docking process of the NL into 3 stages (dashed lines in Fig. 4.5 and Fig. 4.7) and analyze the main interactions with the NL during each stage using the contact-map (Fig. 4.8).

Stage I: NL is "trapped rearward" by β1-domain and α6. In the crystal structure of kinesin, other than the eight-stranded antiparallel β-sheet located at the center of the head, there is a smaller antiparallel β-sheet composed of 3 short strands (β1a, β1b and β1c, closed to the N-terminal) located at the rear of the head named β1-domain (Fig. 4.7). Interestingly, at the start of the simulation, the NL docking does not happen immediately. Instead, it is"trapped" by this β1-domain for about 130ns (Fig. 4.7A). We refer to this as stage I and the conformation of the NL as NL_3. The residues that trap the NL by strong interactions, thus preventing it from directly redocking, are located in the β1-domain and α6. They are shown in the interaction-map (Stage I in Fig. 4.8) and presented as a purple surf pocket in Fig. 4.7A. During this stage, the CS is completely disordered and does not interact with the NL.

We performed several other simulations starting from the same conformation for a shorter time scale (10 ns – Table 4.1c)). Within that time scale, none of the simulations show the NL moving forward, but the interactions between the NL and the β1-domain and the α6 are stable (data not shown). These interactions also exist in our other all-atom simulations for 10 ns with two heads and two heads-MT (Table 4.1d) and e)). Experiments indicate that the flexible NL interacts with the β1-domain during its undocked state in Kinesin-1 and Kinesin-3 [74], and Kinesin-5 (discuss further in the Discussions), thus suggesting that this is a common mecha-

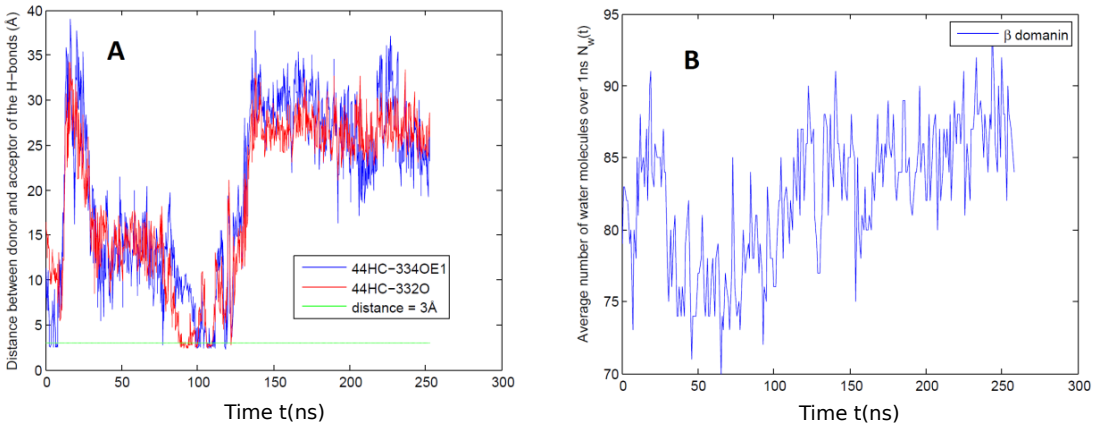


Figure 4.9: Hydrophobic interactions between the NL and the β1-domain during the first 250ns.
A: Distance between the donor and the acceptor of the key H-bonds in the interactions. **B**: Counting average number of water molecules in contact with β1-domain every 1 ns. The NL is latched by the 2 residues N332 and E334. There is a peak where NL is moving far away (at ∼20 ns) but then being trapped again (at ∼100ns) until completely apart from the β1-domain with the help of CS after 130 ns.

nism within kinesin families.

Among the interactions between the NL and the β1-domain, the most important are the H-bonds between K44 in the beta domain, and N332 and E334 in the NL. Fig. 4.9A shows that the H-bonds disrupted after some period of time at the beginning, but reform (at around 100ns) when the distance between donor and acceptor becomes less than 3 Å (the green line). The number of water molecules within 3.5 Å of the β1-domain decreased by 10 molecules during the time ∼40-70 ns (Fig. 4.9B). In other words, the NL was drawn back to the β1-domain by the hydrophobic effect, and then formed H-bonds, during the first stage of our simulation.

Stage II: The CS helps to disrupt NL – β1-domain interactions and α6 forms an extra turn. After the interactions of the NL with the β1-domain are

disrupted by fluctuations, the NL is oriented forwardly with the Cover-Neck-Bundle formation (Fig. 4.5). There are a number of hydrophobic interactions between the NL and CS. However, all of them are between the CS and the β10 (Fig. 4.7B), instead of the β9 as appeared in crystal structures (Fig. 4.3B). These contacts are not strong and the stability of the CNB is compromised.

In the simulation, the CS interacts with the β10, instead of the β9 because the NL was shortened when some hydrogen bonds (orange solid lines in Stage II in Fig. 4.8) started to form between the residues at the N-terminal of the neck with the other residues at the end of α6 (Fig. 4.7B). This indicates that α6 has an extra turn. Moreover, the NL is even shorter when the other C-terminal of the neck (T336, E334) interacts with the middle residues of the CS (A5, L4) and the other residues in the middle (C330, V331, V333) interact with the first end residue of α4 (L268) (see Stage II in Fig. 4.8).

In another 1 μs simulation that we perform in presence of α and β tubulin units, we observe the whole NL completely formed an extensive alpha helix of α6 (Table 4.1f) and Fig. 4.10). However, CHARMM27 (CHARMM22+CMAP) force field and water TIP3P model are known to stabilize the helices [151]. So, in order to testify the observation of the NL forming extra turn to α6, we perform a secondary structure prediction using Jpred4 [18]. The result also shows the possibility that the first half of the NL can form extra helix turns with the α6 (Fig. 4.11). Moreover, based on EM and crystal structures, it is also shown that the α6 is extended when the α4 change the position, opening the"docking pocket", allowing the NL to dock [19,75]. Thus we do not exclude the possibility of what we observe in our simulation

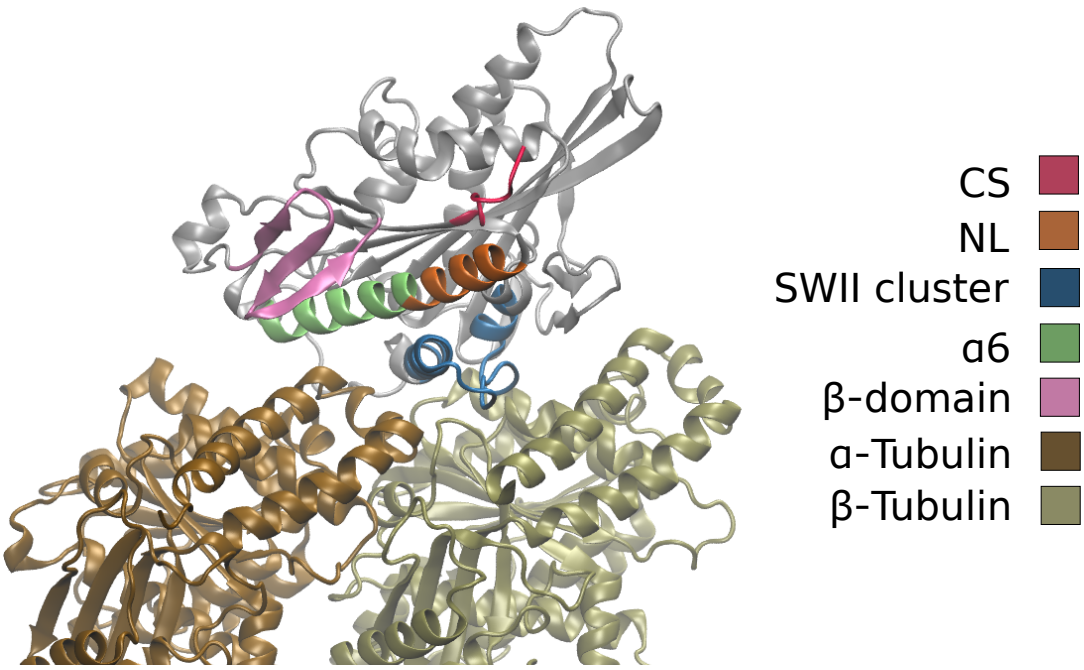


Figure 4.10: A snapshot of another 1 μs simulation for MT-Kinesin complex starting with a rearward pointing NL conformation (Table 4.1f)), showing the NL quickly becomes the extension of the $\alpha 6$ helix. The CS interacts with the NL during the time.

here showing that the $\alpha 6$ can be further extended in solutions if the interhead tension is low.

Stage III: The NL docks in a new binding site adjacent to the MT binding pocket. At $t \sim 500$ ns, the hydrophobic interactions between the NL and CS become unstable, the NL is captured near the MT-binding pocket which mainly consists of SW-II cluster (We refer to this conformation as NL_2 in Fig. 4.7C.) The two most important residues of the NL that Hwang *et al.* mentioned as a latch in the docking mechanism, E334 and N332, also contribute the most to the interactions

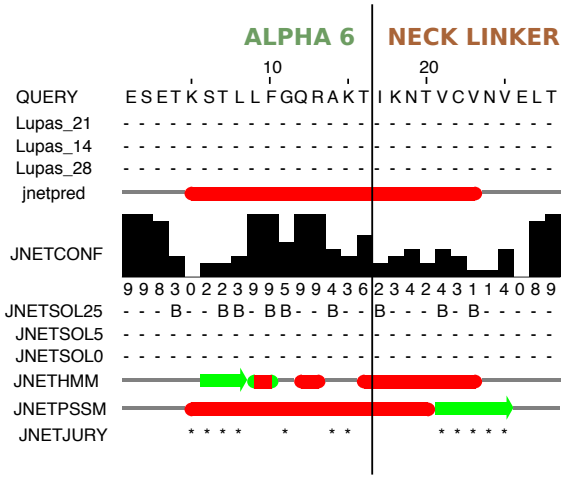


Figure 4.11: JNet secondary structure prediction for sequence of the α6 and the NL of human Kinesin-1 using Jpred4 [18].
Helices are marked as red tubes, and sheets as green arrows. "jnetpred" is the final prediction for the query, showing the prediction that the first half of the NL can form a single helix with the α6.

in this stage (Stage III in Fig. 4.8). Notice, among the two residues that trapped E334, R278 are associated with Switch II (SWII), and R161 is in the L8b loop. Both are adjacent to the MT-binding domains. Such a conformation of the NL has not been observed in any structures of Kinesin-1, but has been discovered in Kinesin-5 (see discussion section for details.)

The α4 helix, believed to be the "relay helix" of Kinesin [142], has a relative orientation with respect to the motor core links the "nucleotide cleft" and the "docking pocket" about 15 Å apart (see the "seesaw model" in Chap. 1). This is consistent with our observation that the noticeable differences between the 1MKJ structure and our structure at the end of the simulation are the NL and CS conformations (NL_1 vs NL_2 in Fig. 4.7D), coupled with an 8 ° tilt of the α4 (Fig. 4.7E).

4.4 Discussion

We have investigated the kinesin NL docking process using atomically detailed MD simulations and obtained several interesting results. The full movement of NL docking with CNB formation was captured in a 1μs simulation. Some interactions between the NL and the core were found that had not been previously observed in crystal structures. The formation of CNB played a role in bringing the NL away from the rear end, driving it toward the binding pocket. The NL can form alpha helix, providing an extra turn for α6. Despite limitations of force fields and limited simulation time scale, our simulations provided insights into the docking process, such a structural transition known to be crucial to the stepping of kinesin.

A previous study from our group [130] showed that in a step of kinesin, after NL docking, is completed by a random diffusion of the tethered head to find the correct binding site (directed diffusion in Fig. 4.1B or process driven by rate k_7 in Fig. 4.18). It is suggested that the interaction between the MT binding site and MT helps to orient the head into the correct direction. In this work, we showed that the NL interacts with the β1-domain and α6 in the back, stabilized by hydrophobic contacts in the first stage when the NL is pointing rearward. This suggests that water molecules must play an important role indriving the NL to the region in the back of the head and to orient the NL to the rearward position(orange box in Fig. 4.18). This, in addition to the MT–Kinesin interactions at the interface mentioned earlier, helps the head to orient correctly before it binds to the target binding site, thus completing the step (last phase in Fig. 4.1B).

The residues that interact with the NL during the stage III belong to L12, L13 and L8 loops (See Stage III in Fig. 4.8 for the list of residues and Fig. 4.7C to visualize their positions in the structure). Among them, L12 belong to the SWII cluster, which was observed to be directly in contact with the MT interface. Even though L8 does not interact with the tubulin [74], it is close to the interface (Fig. 4.12). Thus, there might be one possible reason that the absence of MT in the simulations made the MT-binding pocket solvent exposed and attracted the NL, preventing it from going back to the position in the crystal structure. In order to assess the role of MT, we performed two additional simulations. We aligned the conformation of Kinesin-1 at the end of the 1 μs simulation, with the crystal structure of 2 tubulin units, and relaxed the system for 10ns (Table 4.1 g)). Even though the α4 helix is shorter and L11 is longer in the simulations compared to the crystal structure of Kinesin-1 obtained in complex with MT [19,71,73,74], the positions of SWII and L8 tips, viewed from the NL side, are similar to the actual crystal structure of Kinesin-MT complex (Fig. 4.12). The NL in the NL_2 conformation does not interact directly with the tubulins. In both of the simulations, the NL conformations are stable (flat red lines in Fig. 4.13A, B), even though the CS are disordered. This suggests that the presence/absence of the MT does not affect the NL_2 conformation.

We observed in our simulations that the CS is important in enabling the NL to move forward and dock to the LH. However, CS interacts mainly with the second strand of the NL (β10) in our simulation (Stage II in Fig. 4.8), instead of the initial strand of the NL (β9) as in the crystal structure (Fig. 4.7D). In fact, when a core head of Kinesin-1 is hybridized with a NL of Kinesin-5, the motor loses its motility.

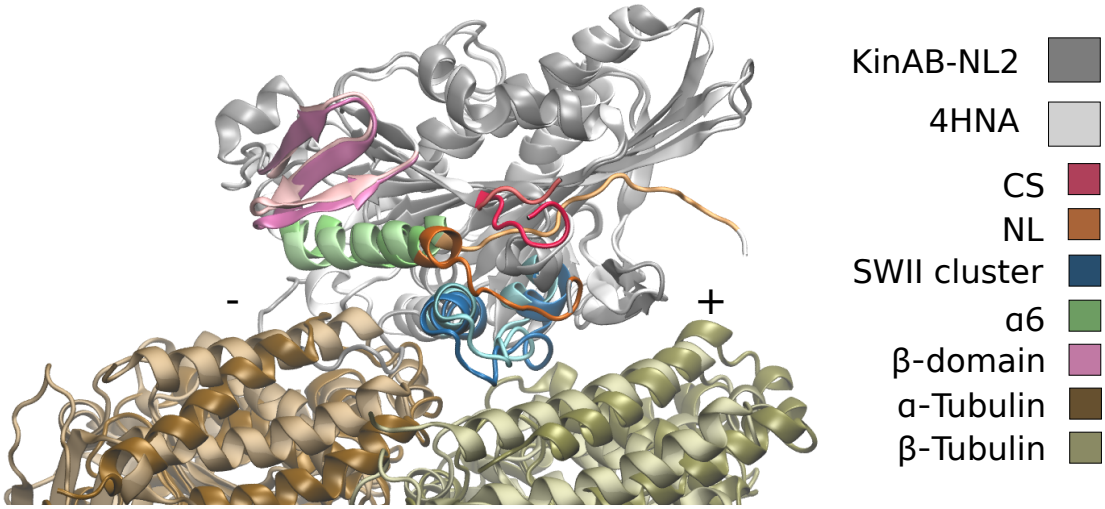


Figure 4.12: Comparing the structure of MT-Kinesin complex (dark colors) with the NL2 conformation from the end of our 1μ s simulation vs. the docked MT-Kinesin crystal structure 4HNA [19].
There are slight protrusions in of L12 and L8. However the NL does not directly interact with the tubulins even in the NL2 conformation.

However, when a wild type core head and $\beta 9$ strand of Kinesin-1 is attached to a NL of Kinesin-5, the chimera restored motility [152]. There is also evidence of sequential docking of the NL in Kinesin-3. Thus, it is crucial that the initial residues of NL and $\beta 9$ domain dock firstin order to form the "correct" docking conformation [153]. The sequential docking is also highly coupled with the chemical ATP cycle. We argue that in order for the NL to favor the docking conformation observed in the crystal structure (NL_1 conformation) the formation of interactions between the CS and the $\beta 9$ strand to form the CNB is crucial. In our simulations, we did not observe the formation of these interactions, which is the reason why we could not observe the NL_1 conformation.

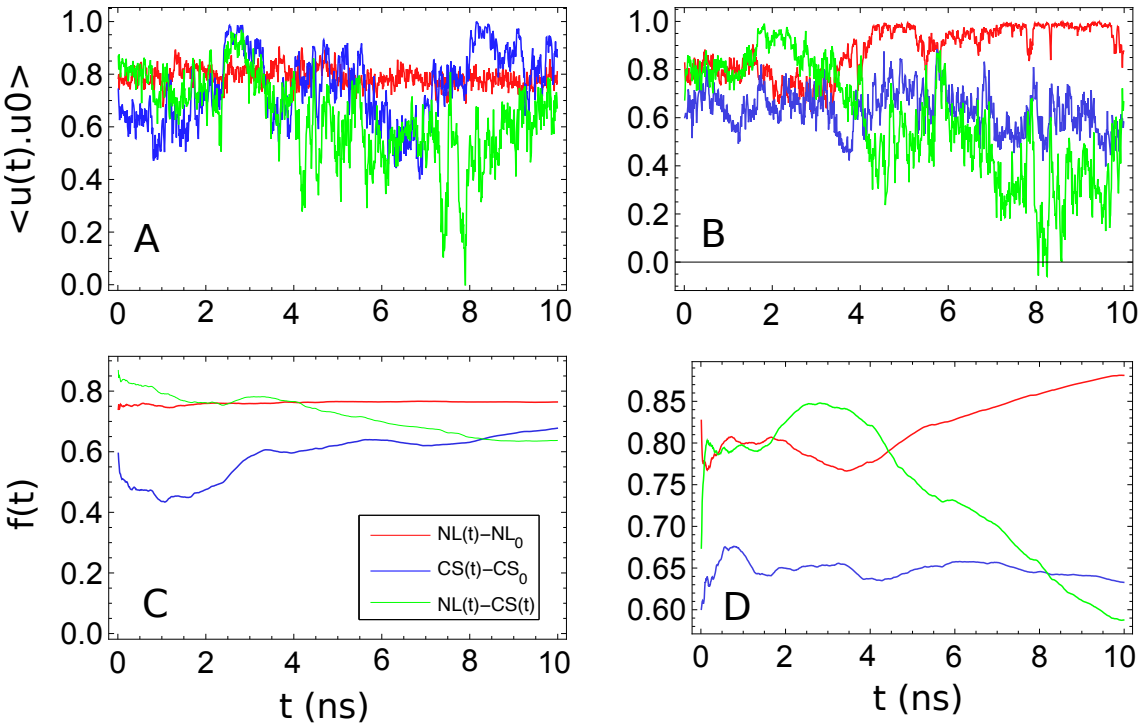


Figure 4.13: NL_2 conformation is stable within 10 ns in the MT-Kinesin complex (Fig. 4.12).
Correlation (**A** and **B**) and Average running (**C** and **D**, respectively) of the NL and CS obtained from two 10 ns-trajectory of the MT-Kinesin complex with NL_2 conformation (KinAB_NL2 in Fig. 4.12).

The tension in the NL is estimated to be 12-15 pN [154, 155], which is larger compared to the ∼6 pN stall force [25, 76]. The absence of the stalk and other trailing head protein makes the NL in our simulation more flexible than it ought to be. We believe that the enhanced flexibility of the NL is the reason it forms an extra turn at the end of $\alpha 6$ in our simulations. We expect if the tension is large, the NL could extend by stretching the $\alpha 6$ end to complete the step of 8 nm [156]. We also expect that the tensionincreases the probability of forming the CNB between the CS and the $\beta 9$, thus reducing the probability that the NL adopts NL_2-like

conformationwhen the two heads work cooperatively to walk on the track.

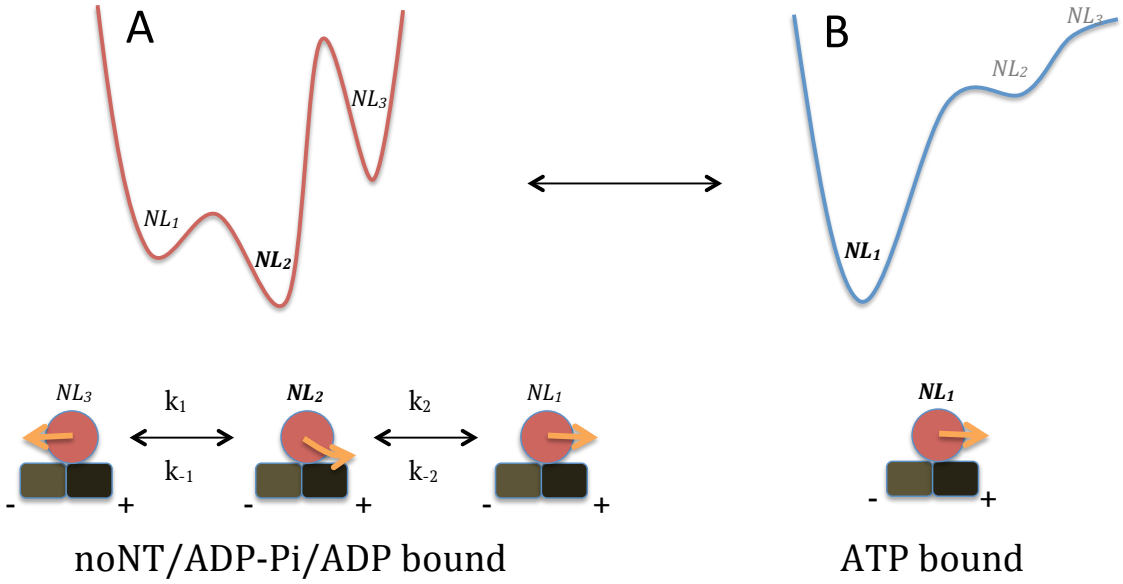


Figure 4.14: Landscape of model of NL conformations.
Kinesin has three different docking conformations. Two conformations NL_1 as observed in the crystal structures and NL_2 found in our simulations are separated with conformation NL_3 by a high barrier. **A**: In solution, NL of a nucleotide bound motor head can interchange between the conformation NL_1 with a more favored conformation NL_2. NL_3 conformation can only easily be accessed with a high "unzipping" force, or jumped out with the formation of CNB (see text for details) **B**: When ATP is bound to the MT-bound head, the landscape is shifted and only NL_1 conformation is favored.

When the motor is in a complex with MT, a distinct NL_1 conformation is only observed with bound ATP. However, without MT, it was suggested that the NL is "docked" in both the ATP analogues and ADP bound states [69, 70]. Experiments support this idea. The FRET data shows that the NL undergoes a large conformational change upon ATP binding on MT-bound Kinesin-1. However, the FRET signal suggests that the NL is not rigidly fixed in the nucleotide binding

states when the motor is in solution or weakly bound to MT [70]. The NL freely exchanges between docked and disordered conformations when MTs are absent [69]. Our work suggest that Kinesin-1 NL can also adopt at least two intermediate states (NL_3 and NL_2) beside the "docked" conformation in the crystal structure, similar to Kinesin-5 (Fig. 4.14). It is likely that Kinesin-1 and Kinesin-5 use the same basic mechanisms for force generation as shown below.

4.4.1 Similarity with Kinesin-5 mechanism

While Kinesin-1 carries cargo inside the cytoplasm of the axon of nerve cells (axoplasm), Kinesin-5 (Eg5) walks and cross-links two antiparallel microtubules, forming the bipolar spindle during mitosis [157, 158]. The latter has become one of the most studied member of the kinesin families, due to the potential to be a target for anticancer drug development [159, 160]. In fact, the number of crystal structures available for Kinesin-5 has increased dramatically and exceeded the number of structures for Kinesin-1 [72]. The mitotic motor and the conventional kinesin share 40% sequence similarity and have almost the same structure with some deviations in the lengths of the loops and α helices [20]. It has been shown that these two processive motors share several similar mechanisms. For instance, the MT bound head rotates upon ATP binding in both Kinesin-1 and Kinesin-5 [21, 71]. The central β-sheets get twisted along the nucleotide cycles in both the families [41, 73, 161]. By using our simulations, we show that they also share similarities in the NL docking processes.

The first crystal structure of ADP bound Kinesin-5 revealed a conformation of

the NL that was not observed before in Kinesin-1. Interestingly, in that structure, the NL interacts with the β1-domain in the back of the head as we observed during stage I in our simulation (Fig. 4.15). Even the two conserved residues but not in other families that mainly interact with the NL in the Kinesin-5 crystal structure (shown as black) are in the same position with the residues that mainly interact with the NL in our simulation (shown as purple). This conformation of the NL is further supported by conclusions reached in previous works [21]. The pseudo-atomic model fitting the 1II6 structure [20] into the EM data shows a similar NL conformation [21] (Fig. 4.16). Site specific gold labels in that study confirm that the NL during the rigor conformation (before ATP bound state) orients towards the minus end.

There is also evidence of the CS supporting the NL during the docking process in Kinesin-5 [21, 22, 136, 138]. This is consistent with what we see in our simulation during stage II. With the assistance of CS, the NL can escape more easily from the interactions with the β1-domain, leading to docking to the LH.

It is also suggested that the NL of Kinesin-5 is in equilibrium between two discrete orientations, one is the docked position (similar to the NL_1 conformation), and the other is adjacent to that position, before ATP binding [13, 22] (corresponding to the process driven by k_2 and k_{-2} rates in the purple box in Fig. 4.18). The kinetic results suggest that before ATP binding, equilibrium actualy favors the second conformation of the NL. Other experiments also suggest that the NL is still oriented toward the plus-end after ATP hydrolysis [21] (green box in Fig. 4.18). This implies that ATP binding simply redistributes the equilibrium between the two NL conformations to favor only the docked conformation observed in crystal

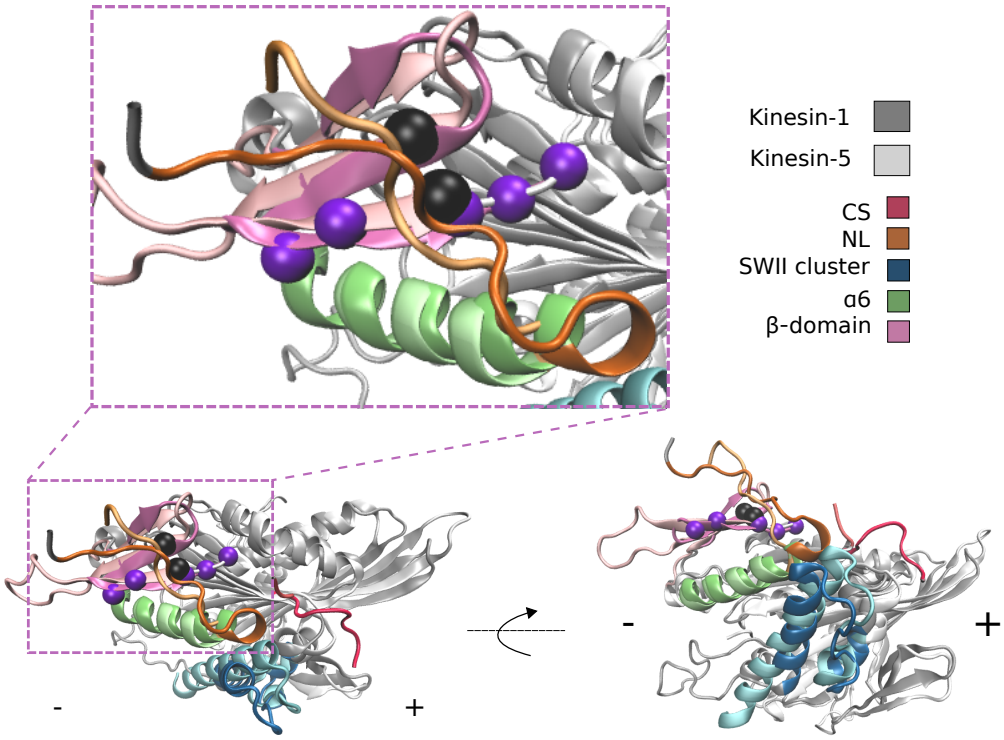


Figure 4.15: Comparing the structure of Kinesin-1 (darker colors) during stage I in our simulation (NL_3 conformation) with Kinesin-5 structure [20].
The snapshot shows similar conformations on the NLs. Two black beads are the two conserved residues found only in the Kinesin-5 family that interact with the NL in the in the 1II6 structure. The residues that interact with the NL in our simulations (Fig. 4.8 and 4.7A) during stage I are in dark purple. The main domains are colored as in Fig. 4.7.

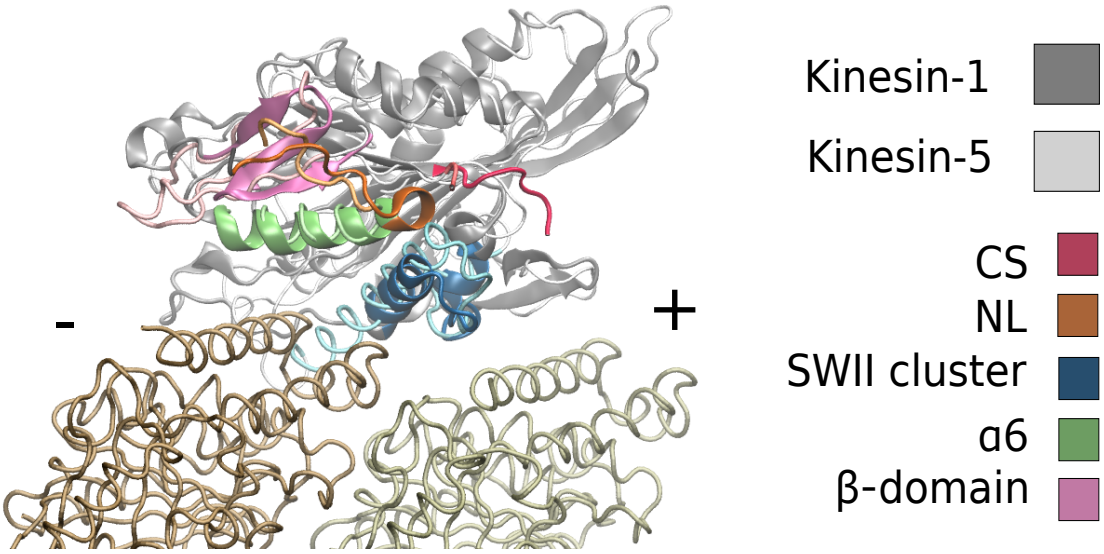


Figure 4.16: Similar NL conformations between the structure of Kinesin-1 (darker colors) during stage I in our simulation (NL_3 conformation) with Kinesin-5 pseudo-atomic model [21] obtained by fitting the 1II6 structure [20] into EM data.
The main domains are colored as in Fig. 4.7.

structures (Fig. 4.14). The second docking conformation has not been observed in crystal structures. We suggest that this conformation is the NL_2 conformation in stage III, where the NL interacts with the residues close to the MT binding site. Indeed, a snapshot of the NL_2 conformation in our simulation shows a similar arrangement with the pseudo-atomic models of disconnected NL conformation in the ATP hydrolysis transition state reported for Kinesin-5 elsewhere [22] (Fig. 4.17).

Several FRET studies provided insights into the mechanochemitry of K1 and K5 [11, 13]. In one of the experimental set-up, Kinesin-1 [11] and Kinesin-5 [13] were labeled with AEDANS at the tip of the NLs. These AEDANS probes can be excited by FRET from the MT tryptophans. The value of R_0, the Forster critical energy transfer distance between tryptophan and AEDANS has been measured to be $20.3 - 21.0\,\text{Å}$ [11]. That means, in order to obtain FRET signals, the NL that contains the FRET acceptor has to remain close within 20 Å to the tryptophan-donors in the MT most of the time. Or in other words, the NLs of both families tend to be adjacent to the MT surface as in the conformation NL_2, instead of more than 20 Å apart from the MT as in the conformation NL_1.

Moreover, another experiment measured the FRET distances between MT-bound Oregon Green 488-taxol and NL-attached QSY7 maleimide for both monomeric K-1 and K-5 (Table three in [13]). It was shown that for Kinesin-1, the mean distance between the tip of the NL and the MT is 37.7 Å in the rigor state (no NT bound) and 56.9 Å in the ATP state (AMPPNP bound). This means that once the ATP binds, the tip of the NL is moved more than 20 Å apart from the position before ATP binding, and about 20 Å farther from the MT. This is consistent with the

mean distance 21.9 Å difference between the residue A337 of the NL in the NL_2 and NL_1 conformations (Fig. 4.6). Also, the NL is more stable once the ATP binds to the head (half-width distance, related to the s.d. of the distance distribution [162], is smaller in the AMPPNP condition - 3.6 Å - than in the rigor condition - 1.2 Å) in Table three of [13]). We infer from these results that before ATP binding, the NL of a monomeric Kinesin-1 is in equilibrium between different NL conformations, and spend most of the time in the NL_2 conformation (purple box in Fig. 4.18). The ATP binding will shift the energy landscape to favor only NL_1 (2nd state in Fig. 4.18).

Before our work, it was believed that the docking process in Kinesin-5 might be very different from Kinesin-1, due to some conformations of NL observed in Kinesin-5 [20, 22], but not in Kinesin-1. We believe that, the deviations in the two families lie in the longer and less conserved NL in Kinesin-5, compared to Kinesin-1. This increases the interactions and the mean life time of stage I and stage III in Kinesin-5, making it easier to capture such conformations in crystal [20] or EM structures [21,22]. Indeed, the whole NL docking rate is reported to be more than 10 times faster in Kinesin-1 than in Kinesin-5 (Table 4.3). Thus, the fact that we have not observed such conformations in Kinesin-1 might mean that they are difficult to detect in experiments.

Our work here is the first to report such new conformations in Kinesin-1. Hence, we suggest that, other than working at different kinetic rates (Table 4.3), Kinesin-5 and Kinesin-1 have very similar docking mechanisms, including how the NL changes the conformation when it orients during the mechanochemical cycle. As

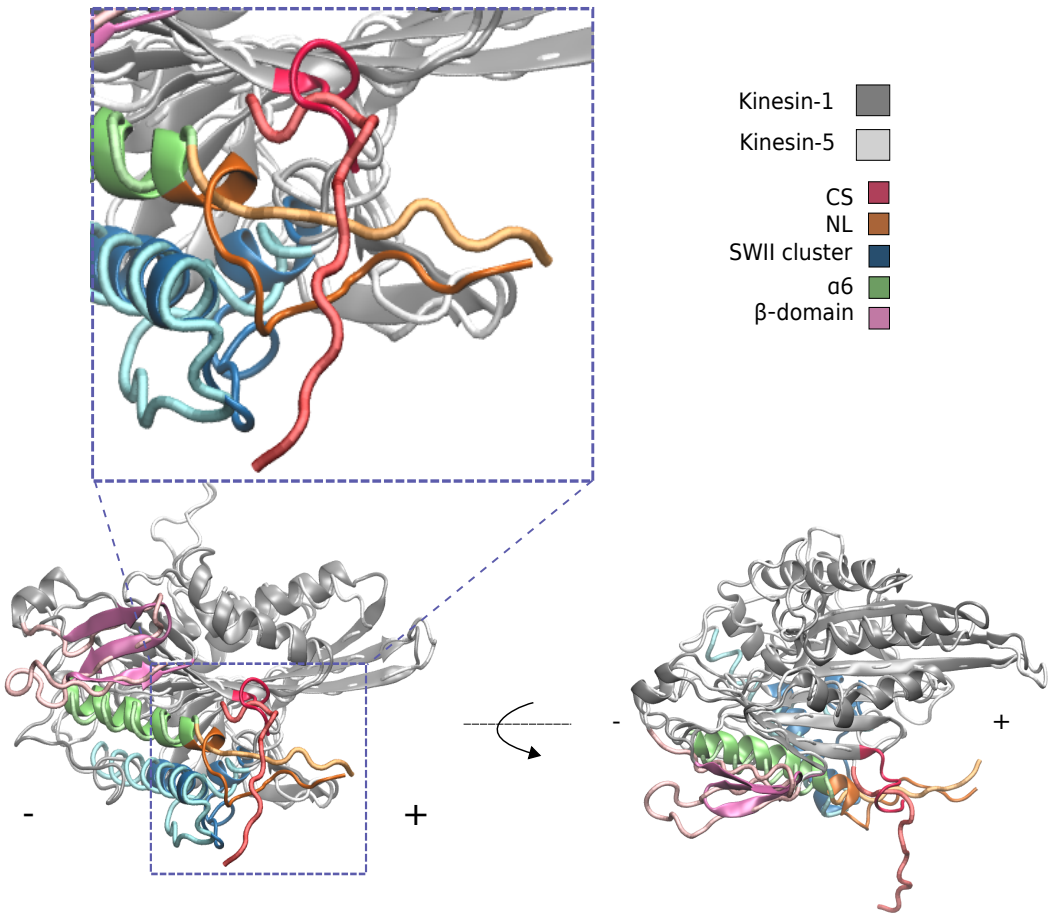


Figure 4.17: Similar NL and CS conformations between a snapshot of Kinesin-1 (darker colors) in stage III in our simulation (NL_2 conformation) and the Kinesin-5 coordination reported in [22] with a "disconnected NL".

the motors in both families walk with a hand-over-hand mechanism, the NLs can adopt 3 different conformations: interacting with the core beta sheet and pointed towards the plus end of the MT (NL_1 or often referred to as "docked" conformation), interacting with residues next to the MT-binding site (NL_2) and pointed towards the plus end of the MT, and interacting with the β1-domain and pointed towards the minus end of the MT (Fig. 4.7). As demonstrated in Fig. 4.18, the ATP cycle starts with a conformation that occurs when the partner head is dissociated from the MT, opening the "ATP gating" in the MT-bound leading head. The NL at this stage, starting from the NL_3 conformation, is in equilibrium between all 3 conformations (purple box in Fig. 4.18). With the help of the CS forming the CNB with the NL, the NL_3 conformation is less favored than the NL_2 and NL_1 conformations. Among those two latter conformations, FRET evidence suggested that the equilibrium favors the NL_2 the most, before ATP binding [13] (see again Fig. 4.14). Our main 1 μs simulation trajectory (Table 4.1b)) captures this state in the cycle. After ATP binding to the Kinesin-MT complex, the NL is "zipped" tightly to the core, only the NL_1 is most favored. At this stage, this head has become the trailing head. Following ATP hydrolysis, the NL is in equilibrium between two conformations NL_2 and NL_1, probably favoring NL_2 again (green box in Fig. 4.18) and quickly dissociate from the MT once the orthophosphate is released. The NL is still oriented towards the (+) end. After getting launched forward by the partner head, the detached head has its NL pulled backward. This NL now is trapped at the conformation NL_3, which in turn, assists the formation of the head in correct orientation to locate the next binding site on MT during the random-search-phase

Table 4.3: Some reported kinetic rates (Fig. 4.18) measured in both Kinesin-1 and Kinesin-5 at 20°C from the same lab for comparisons [11–13].

Reaction	Unit	Kinesin-1 [11, 12]	Kinesin-5 [13]
NK docking	s^{-1}	800	62 ± 17
ATP binding to K-MT (k_3)	$\mu M^{-1}\ s^{-1}$	4.9	1.1 ± 0.1
ATP dissociation from K-MT (k_{-3})	s^{-1}	105 ± 11	19 ± 9.6
ATP hydrolysis (k_4)	s^{-1}	100	10.3 ± 0.3
ATP-induced dissociation (k_5)	s^{-1}	60	6.3 ± 0.3

(orange box in Fig. 4.18). After ADP release, this head binds the MT tightly, creates the tension in the NLs that prevent the next ATP binding, until the partner head is disengaged from the MT once again, starting the new cycle. By changing between the three conformations, the NLs regulate the mechanochemistry of the heads to move processively forward. We expect both Kinesin-1 and Kinesin-5 share this mechanism.

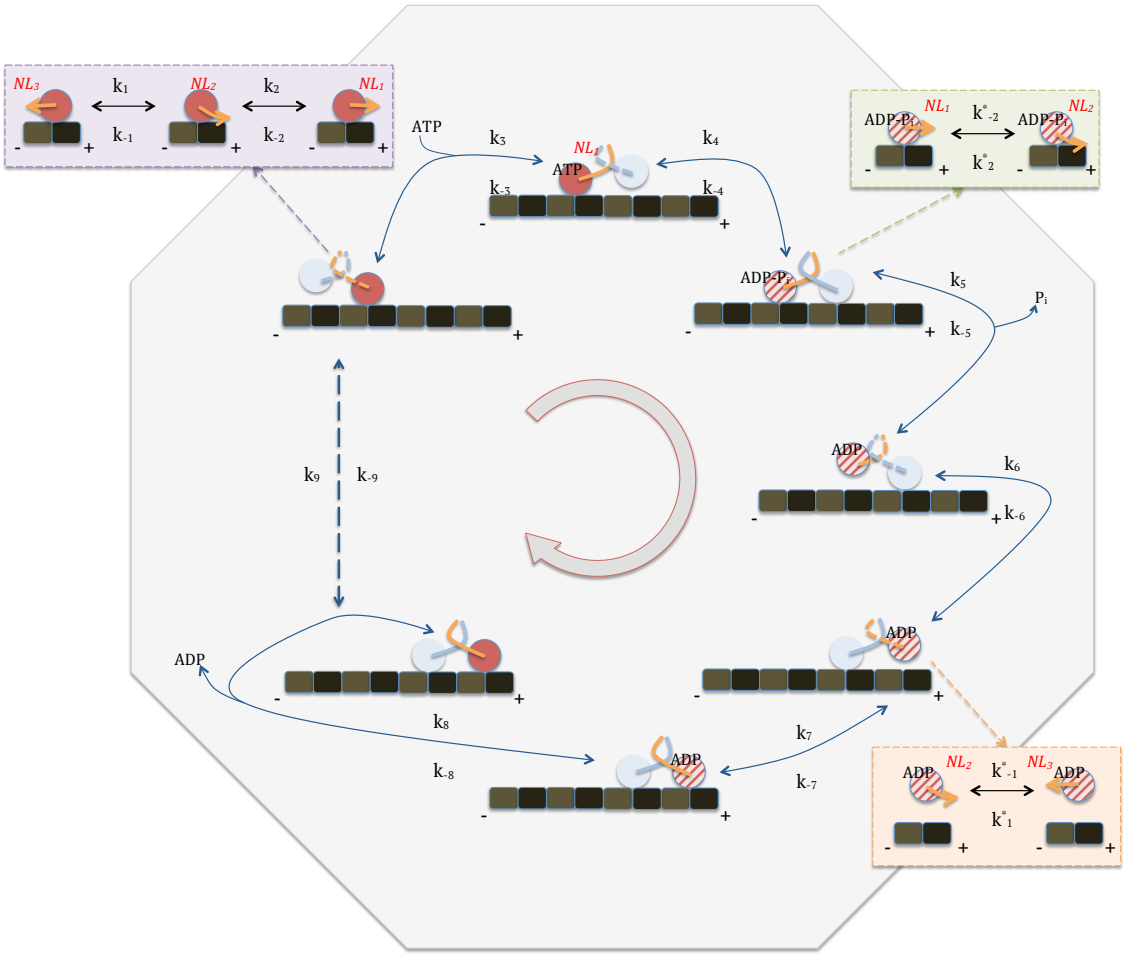


Figure 4.18: General schematic pathway for stepping in K-1 and K-5. The focused head is colored in solid red circle (striated circle) in the states having high (low) MT affinity, accompanied by the partner head colored in light blue. Tubulins are in dark grey rectangles. The NLs are thick solid (dashed) lines when they are under high (low) tensions or forces. Smaller boxes are details of the NL changing conformations in the different NL "disordered" states. The docking process is shown in the purple box. Some of the reported kinetic rates are summarized in Table 4.3.

Chapter 5: Conclusions

5.1 From stochastic discrete models for $P(v)$ to simulations for detail mechanisms of processive motors

Processive motors are a large class of molecular motors in cells which walk on linear tracks in order to execute diverse functions. Many insights about their dynamics and mechanisms have been revealed by current theories and experiments. However, most of the attention is focused on mean values of observations, such as their mean velocity and mean runlength. Dispersion in these observables are poorly described, despite the fact that fluctuations are important. This dissertation addresses that gap by providing general stochastic theories to investigate velocity and runlength distributions of processive motors, which are in pronciple measureable. The theories provide quantitative insights about the workings of kinesins and helicases, two examples of processive motors.

Our theory demonstrated that, when the processive motors can not take many steps, which is likely the case when they are subjected to large resisting loads, velocity distributions are not Gaussian. Exact equations for velocity distributions and runlength distributions are obtained for several models (Table 1.4) and can be

used to analyze data when the motors are not "processive" enough for the Central Limit Theorem to be valid. In Chap. 2, we validate the theory by fitting simultaneously our equations to the experimental data of $P(v)$ and $P(n)$ of Kinesin-1 at zero force, with a *single* fitting parameter – the detachment rate γ. The extracted value is in an agreement with the physical detachment rate of two headed Kinesin-1, estimated from the detachment rates of independent heads measured in experiments (Appendix B). The theory is used to predict $P(v)$ of Kinesin-1 under loads, which indeed deviate from Gaussian. A more surprising prediction is that, the velocity distributions are bimodal, with two peaks separated on the positive and negative velocity populations. The bimodality is a consequence of the discreteness of Kinesin-1's step-size. Our prediction awaits experimental validation.

Application of our theory to helicases in Chap. 3, we have shown that helicases in superfamily IV are heterogeneous. To our knowledge, this demonstrated model is the first quantitative test of heterogeneity. Although our theory for heterogeneous systems is weaked by assuming static disorder, the theory works as an effective model with the extracted value of the dispersion of the ATP driving forward stepping rate, σ_k as a lower bound value. Moreover, based on our theory, we suggest single molecule experiment to determine the type of disorder exist in the systems, static vs dynamic. Note again that the theories here are general and can be applied to a broad general class of processive motors.

Other than analytical theory, computations and simulations are powerful tools to investigate complex systems. Chap. 4 demonstrates one example of using all atomic simulations to investigate important process in mechanisms of kinesins: the

neck-linker docking. The NL of kinesins is important in regulating communication between the two heads. It is suggested that upon ATP binding to the leading head, the NL change its conformations in a process called NL docking to generate the required force to launch the trailing head forward. However, details of how the NL change its conformations are not well understood, due to the difficulties in directly visualizing it in experiments and the limitations of time scale in MD simulations. We have overcome this limitation by using Anton supercomputer to simulate atomic details of the NL docking process and reveal some interesting results. Firstly, the simulation confirms the role of the Cover-Neck-Bundle that assist the NL to orient forward. Secondly, in our simulations when the NL is free to move with no tension, we observe the NL forms an extra turn for $\alpha 6$ helix. And thirdly, the NL is observed to interact with different binding regions that have not been observed in Kinesin-1 before, but captured in Kinesin-5's structure. The work helps us to better understand some commonalities among different superfamilies of kinesins.

5.2 T7 and T4 helicases exhibit heterogeneity, unlike Kinesin-1

To account for different experimental setups, the analytical models that have been used to analyze data of kinesins and helicases are slightly different, but they are indeed similar in the nature. Hence, it is interesting to draw some comparasions in the biological workings of these two processive motors as the last remark of this disertation.

While most helicases (RecBCD [115], T4 [16], T7 [15], UvrD [114], PcrA [83,

116], AddAB [117]) seem to exhibit some levels of heterogeneity (quantitatively tested for T4 and T7 helicases in Chap. 3), we had shown that K-1 does not exhibit heterogeneity in Chap. 2 and [85]. A model where all K-1 molecules have identical kinetic rates in moving forward was sufficient to explain the velocity distribution at zero force. This difference is interesting, and could be a result of fundamentally different mechanisms of nucleotide binding/hydrolysis in helicases and kinesins, or could be a consequence of differences in the structure and environment of these different motors.

In particular, the linear microtubule lattice that kinesins walk on, is stiffer than the DNA/RNA track for helicases (persistence length of microtubules is 1.4 mm [163] whereas that of dsDNA is about 50 nm [164]). Moreover, microtubules are polymers made from tubulin dimers, so they are relatively homogeneous. However, DNA/RNA strands have four different types of monomer blocks, making them more heterogeneous. It was shown that motors are more likely to pause or stall on a heterogeneous track, which could explain the overall heterogeneous behavior of helicases [165].

The step-size of kinesin is also larger than in helicases. Kinesins walk in a hand-over-hand manner on microtubules with a discrete step-size of 8.2 nm [38,129], whereas the step-size of most helicases is an order of magnitude smaller (less than 1 nm [23]). Molecular heterogeneity in the track could result in large variations in the velocities of individual helicases at smaller length scales. However, the variation might simply get washed out at the larger length scale in the case of kinesins.

In addition, T7 and T4 helicases work as hexamers while conventional ki-

nesins function as dimers. Therefore, T7 and T4 helicases have more ATP binding/hydrolysis sites than kinesins, leading to more possible states than the helicases can adopt, increasing the diversity in the functions of the helicase population. While kinesins work as dimers, its gating mechanisms are very accurate, which means the two heads regulate each other very tightly. Even having two ATP binding sites in two heads, no more than one binding site (in the apo MT-bound LH) is capable to bind ATP at all time. However, the regulations among subunits of T7/T4 helicases are not well understood, and the order of ATP binding in different possible binding sites are not clear. There is evidence showing that T7 and T4 helicases can form different oligomer states, including a crystal structure of a heptamer T7 helicases [166]. Hence, it is possible that different native structures of monomers can come together to form different functional oligomer states of T7/T4 helicases. These different functional oligomer states are called morpheeins [167]. Understanding morpheeins are very important in understanding the structural basis of diseases, allosteric regulations and drug developments.

5.3 Open questions for future work

Other than some predictions and observations that await confirmation from experiments, our works here also open more questions for future studies. Some of them are: How to quantitatively demonstrate the origins of heterogeneity within helicases system? How will increasing numbers of ATP binding sites or lowering the length scale increase the heterogeneity in the system? This question can be tackled

by a discrete stochastic rate model that can accomplish also the ATP hydrolysis rates and its effects to other observable quantities. Another question can be asked is: What are the roles of heterogeneity and different morpheein forms in helicases? More mathematical models developed to analyze more future single molecule experimental data are needed to investigate this problem. Moreover, the heterogeneity property is believed to be a general property of not only helicases, but also many other molecular motors. Hence, solving this question will help us to understand an important fundamental biological phenomenon of many molecular motors.

Other questions concerning on detail structural and dynamic mechanisms of helicases are: What are the structures of those morpheein forms? What is the structural basis of different forward rates of helicases? In addition, after many decades of studies, the detail mechanisms of stepping, slipping and pausing in helicases are still not clear. The system size and time scale of these problems are too large for detail all-atom simulations. Hence, the structural based tools, such as coarse-grained simulations [**?**] and Elastic Netword Mode (or Normal Mode Analysis) [**?**] with the advantage of a wider range of parameters are suitable to investigate these questions. The current coarse-grained models are available for proteins, or DNA/RNA separately. However, helicases, and many other protein motors normally works in context with DNA/RNA substrates. Hence, these questions are await for new models for interactions between DNA/RNA and proteins to be developed. More theoretical works will continue to be important to understand the workings of these class of motors in the future.

Appendix A: Detailed derivations of different Kinetic Models for velocity distribution

A.1 Model 1 – simplest model with fixed runlength

A.1.1 Velocity distribution with negligible detachment rate

If detachment rate (γ) is much smaller than the forward stepping rate (k^+), we can neglect γ. In this simplest possible description of a processive motor (Fig A.1), the velocity of a single trajectory is:

$$v = \frac{N}{t} \tag{A.1}$$

where the trajectory is defined by a **fixed** number of steps $n = N$. Since the waiting time for each step is an exponentially distributed random variable, the time t for different trajectories will be different. Hence the velocity defined as $v = N/t$ for different trajectories will also be different. The distribution of velocities arising from

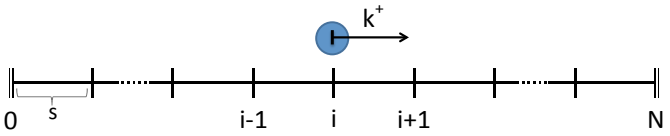


Figure A.1: Schematic of the simpliest model with fixed runlength.

an ensemble of such trajectories is defined as $P_N(v)$.

The probability density $P(n, t_1, t_2, ..., t_n)$ for the motor to jump n steps, at time $t_1, t_2, ..., t_n$ respectively, is given by

$$P(n, t_1, t_2, ..., t_n) = [k^+ e^{-k^+ t_1}][k^+ e^{-k^+(t_2-t_1)}]...[k^+ e^{-k^+(t_n-t_{n-1})}] \quad (A.2)$$

$$= (k^+)^n e^{-k^+ t_n}. \quad (A.3)$$

So the joint probability density for the motor to jump n steps in time t_n is

$$P(n, t_n) = \int_0^{t_n} P(n, t_1, t_2, ..., t_n) dt_1 \int_{t_1}^{t_n} dt_2 ... \int_{t_{n-2}}^{t_n} dt_{n-1} \quad (A.4)$$

$$= (k^+)^n e^{-k^+ t_n} \frac{(t_n)^{n-1}}{(n-1)!}. \quad (A.5)$$

This equation is widely used to analyze dwell-time distributions of molecular motors to estimate number of intermediate states within each step in the literature [90,114]. The velocity distribution as defined before, is given by:

$$P_N(v) = \int_0^{\infty} dt P(n = N, t) \delta \left(v - \frac{n}{t} \right) \quad (A.6)$$

$$= \frac{n}{v^2 (n-1)!} \left(\frac{n}{v} \right)^{n-1} (k^+)^n \exp \left(-k \frac{n}{v} \right) \Big|_{n=N} \quad (A.7)$$

$$= \left(\frac{N}{v} \right)^{N+1} \left(k^+ \exp(-\frac{k^+}{v}) \right)^N \frac{1}{N!} \quad (A.8)$$

A set of Kinetic Monte Carlo simulations [168] has been performed and compared with our analytical results (equations A.5 and A.8) in Fig A.2 to verify both the simulations agree with thetheory.

For the simple case, setting N=1 and we obtain the formula for the distribution of instantaneous velocity obtained separately in Eq. 2.5.

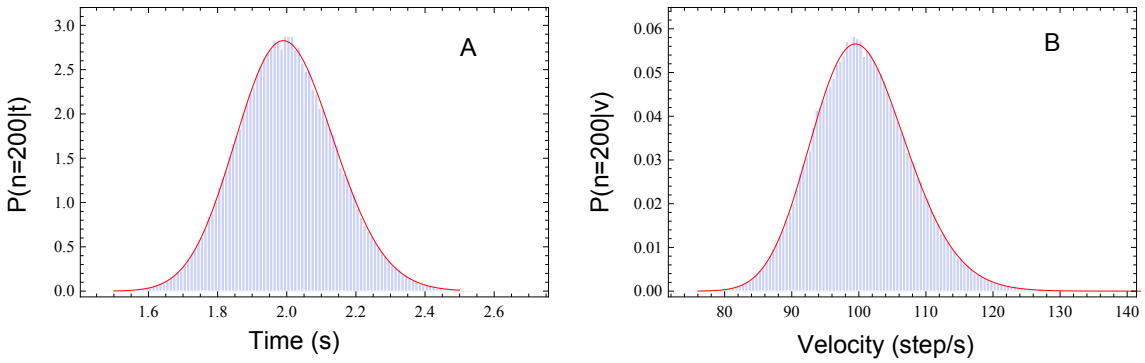


Figure A.2: Comparision of theory and Kinetic Monte Carlo simulation in the model where the motor can only take forward step with a rate of $k^+ = 100$ /s.
A: probability distribution of time for taking $N = 200$ steps and **B**: probability distribution of velocity assuming that all motors take $N = 200$ steps. Red lines are predicted functions from the analytical model.

A.1.2 Estimation of the width of $P_N(v)$ of homogeneous motors when $k^- \to 0$

Now we estimate the width of the velocity distribution for the model described above. Consider the general expected values $I_i = \int_0^\infty v^i P_N(v|k^- \to 0)dv$, we have

$$I_i = \int_0^\infty v^i P_N(v|k^- \to 0)dv \tag{A.9}$$

$$= \frac{N}{N!}\int_0^\infty v^i \mathrm{e}^{-\frac{k^+}{v}N}\left(\frac{k^+N}{v}\right)^N \frac{dv}{v} \tag{A.10}$$

Substitute $u = \frac{k^+ N}{v}$ then

$$I_i = \frac{N}{N!}\int_0^\infty \left(\frac{k^+N}{u}\right)^i \mathrm{e}^{-u}u^{N-1}du \tag{A.11}$$

$$= \frac{N}{N!}(k^+N)^i\int_0^\infty \mathrm{e}^{-u}u^{N-1-i}du \tag{A.12}$$

$$= \frac{N}{N!}(k^+N)^i\frac{1}{N-i}\int_0^\infty \mathrm{e}^{-u}u^{N-i}du \tag{A.13}$$

$$... \tag{A.14}$$

$$= \frac{N}{N!}(k^+N)^i\frac{1}{(N-i)...(N)}\int_0^\infty \mathrm{e}^{-u}u^{N}du \tag{A.15}$$

With $\Gamma(N+1) = \int_0^\infty \mathrm{e}^{-u}u^N du = N!$ is the Gamma function, we obtain:

- If $i = 0$, then $I_0 = \frac{N}{N!}(k^+N)^0\frac{1}{N}N! = 1$ which is the normalization condition for $P_N(v|k^- \to 0)$
- If $i > 1$ then

$$I_i = \frac{(k^+N)^i}{(N-i)...(N-1)} \tag{A.16}$$

From Eq. A.16, the average velocity is

$$I_1 =< v >= \frac{k^+N}{N-1}, \tag{A.17}$$

and

$$I_2 =< v^2 >= \frac{(k^+N)^2}{(N-1)(N-2)}. \tag{A.18}$$

So the standard deviation velocity distribution $P_N(v|k^- \to 0)$ is

$$\sigma_v = \sqrt{< v^2 > - < v >^2} = \frac{k^+N}{N-1}\sqrt{\frac{1}{N-2}} = \frac{< v >}{\sqrt{N-2}} \tag{A.19}$$

The validity of Eqs. A.17 and A.19 is tested in Fig. A.3 when compared with results from Monte Carlo simulations. Notice that when the number of step distance N is

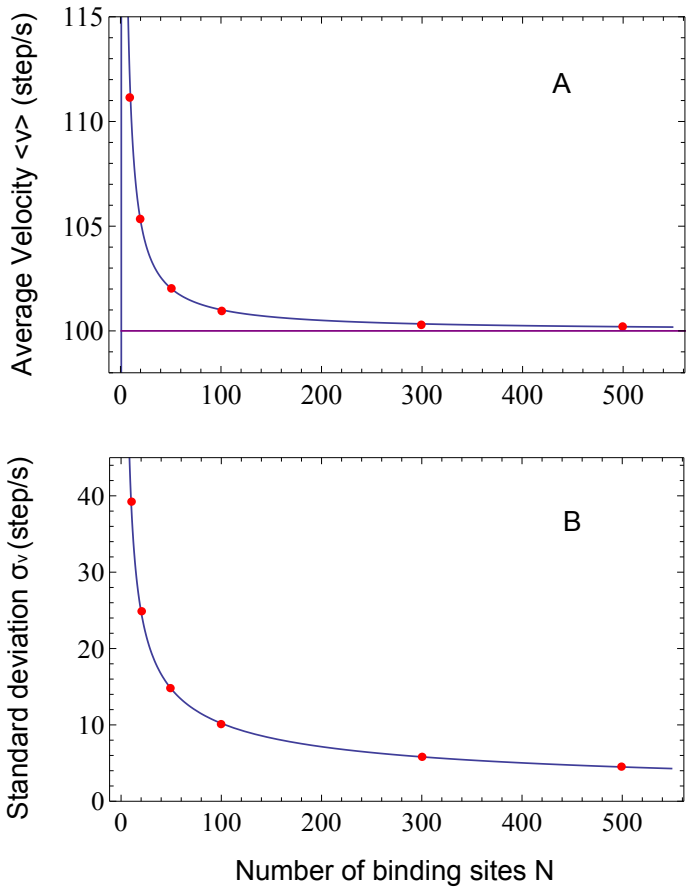


Figure A.3: Average velocity $< v >$ (**A**) and standard deviation of velocity distribution σ_v (**B**) depend on number of binding sites N that the motors travel in every velocity calculations.
The blue lines are exact analytical expressions Eqs. A.17 (**A**) and A.19 (**B**) and the red dots are from Monte Carlo simulations for the same model with $k^+ = 100/s$ which is the value of the purple line in **A**.

large, the prediction in Eq. A.19 gives is in accord wtih the Central Limit Theorem (CLT). Converting to normal real distance unit that normally used in experimental data, we have Eqs. 3.5 and 3.6 in Chap. 3.

A.2 Model 2 – model with finite processivity and zero back-stepping rate

We now consider motors with a detachment rate γ, and negligible back-stepping rate ($k^- \approx 0$). In this model, the motor has a forward rate k^+ and detachment rate γ, as shown in Fig. A.4. Since single molecule experiments can follow motors untill they detach from the track, we define the velocity v from one particular trajectory to be the runlength divided by the run-time:

$$v = \frac{n}{t}. \tag{A.20}$$

Note that unlike Eq. A.1, n is not fixed and is itself a random variable. An ensemble of such trajectories observed in the experiment therefore results in a corresponding ensemble of velocities, giving rise to the distribution $P(v)$.

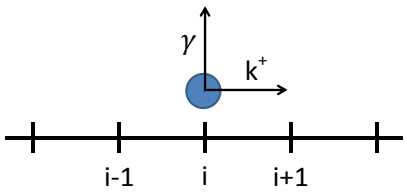


Figure A.4: Schematic of the model with no backward steps.

In this model, the distribution function of time t for a motor to that takes n steps before detaching is

$$P(t) = \gamma e^{(-\gamma t)} \tag{A.21}$$

A.2.1 Runlength distribution

Distribution function of the number of steps n that a motor takes before detaching, also known as processivity or runlength distribution, is given by

$$P(n) = \left(\frac{k^+}{k_T}\right)^n \left(\frac{\gamma}{k_T}\right), \tag{A.22}$$

where $k_T = k^+ + \gamma$ is the total rate. k^+/k_T is just the probability of taking a forward step at each site, while γ/k_T is the probability to detach from any site.

A.2.2 Velocity distribution

Let us consider a trajectory where the motor jumps n steps, at times $t_1, t_2, ..., t_n$ respectively and then detaches at time t, such that $t_1 < t_2 < ... < t_n < t$. The probability density $P(n, t_1, t_2, ..., t_n, t)$ of such a trajectory is given by:

$$\begin{aligned} P(n, t_1, t_2, ..., t_n, t) &= \left[k_T e^{-k_T t_1} \frac{k^+}{k_T}\right] \left[k_T e^{-k_T (t_2 - t_1)} \frac{k^+}{k_T}\right] \ldots \left[k_T e^{-k_T (t - t_n)} \frac{\gamma}{k_T}\right] & \text{(A.23)} \\ &= (k_T)^{n+1} e^{-k_T t} \left(\frac{k^+}{k_T}\right)^n \frac{\gamma}{k_T}. & \text{(A.24)} \end{aligned}$$

So the joint probability density for the motor to jump n steps and detach in t time is

$$\begin{aligned} P(n, t) &= \int_0^t P(n, t_1, t_2, ..., t_n, t) dt_1 \int_{t_1}^t dt_2 \ldots \int_{t_{n-1}}^t dt_n & \text{(A.25)} \\ &= (k_T)^{n+1} e^{-k_T t} \left(\frac{k^+}{k_T}\right)^n \frac{\gamma}{k_T} \frac{t^n}{n!} & \text{(A.26)} \\ &= \left(\frac{t^n}{n!}\right) (k^+)^n \gamma e^{-k_T t}. & \text{(A.27)} \end{aligned}$$

Of course, we have

$$\sum_{n=0}^{\infty} P(n, t) = P(t) \text{ and } \int_0^{\infty} P(n, t) dt = P(n), \tag{A.28}$$

where $P(t)$ is given by Eq. (A.21) and $P(n)$ is given by Eq. (A.22).

The distribution of velocities $P(v)$ is given by:

$$P(v) = \sum_{n=0}^{\infty} \int_0^{\infty} dt P(n,t) \delta\left(v - \frac{n}{t}\right) \tag{A.29}$$

$$= \sum_{n=0}^{\infty} \int_0^{\infty} dt \left(\frac{t^n}{n!}\right) (k^+)^n \gamma e^{-k_T t} \delta\left(v - \frac{n}{t}\right) \tag{A.30}$$

$$= \sum_{n=0}^{\infty} \left(\frac{n}{v}\right)^{n+1} \frac{1}{vn!} (k^+)^n \gamma e^{-k_T \frac{n}{v}} \tag{A.31}$$

$$= \frac{\gamma}{v} \sum_{n=0}^{\infty} \left(\frac{n}{v}\right)^{n+1} \frac{1}{n!} \left(k^+ e^{-\frac{k_T}{v}}\right)^n. \tag{A.32}$$

Eq. A.32 is the exact equation for the velocity distribution of a motor that does not take back-steps. For detachment rates much smaller than the forward stepping rate ($\gamma << k^+$), the small n terms in Eq. A.32 do not contribute much to the sum as the motor usually detaches only taking many steps. In that limit, we can using a particular form of Stirling's approximation $n! \approx \sqrt{2\pi n}(\frac{n}{e})^n$, to obtain an approximate result:

$$P(v) \approx \frac{\gamma}{v^2\sqrt{2\pi}} \sum_{n=0}^{\infty} \sqrt{n} A^n \tag{A.33}$$

with

$$A = \frac{k}{v} e^{1-\frac{k_T}{v}}. \tag{A.34}$$

Approximating the sum in Eq. A.33 by an integral, we obtain

$$\sum_{n=0}^{\infty} \sqrt{n} A^n \approx \int_{n=0}^{\infty} dn \sqrt{n} A^n = \frac{\sqrt{\pi}}{2(-\ln A)^{3/2}}. \tag{A.35}$$

Using the above two approximations, we obtain $P_A(v)$, the approximate version of Eq. A.32:

$$P_A(v) = \frac{\gamma}{v^2} \left(-2\ln\left(\frac{k^+}{v}\right) - 2 + 2\frac{k_T}{v}\right)^{-3/2}. \tag{A.36}$$

A.3 Model 3 – Finite processivity and non-zero back-stepping rate

We extend the previous model to a more realistic one, by adding the rate k^- for the motor to take backward steps (see Fig A.5). In this model, the distribution

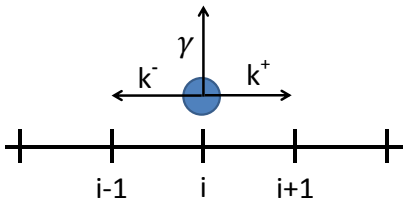


Figure A.5: Schematic of the model with finite backward rate.

function for the time of a motor to stay on the track is still the same as in the previous model (Eq. A.21).

A.3.1 Runlength distribution, $P(n)$

In the model described in Fig. A.5 the motor takes both forward and backward steps, hence we define the distribution $P(m, l)$ for the motor to take m steps forward and l steps backward before falling off the track:

$$P(m,l) = \left(\frac{k^+}{k_T}\right)^m \left(\frac{k^-}{k_T}\right)^l \left(\frac{\gamma}{k_T}\right) \frac{(m+l)!}{m!l!}, \tag{A.37}$$

where $k_T = k^+ + k^- + \gamma$ is the total rate, and the final term is just the binomial coefficient.

If we set $n = m - l$ then the processivity distribution for motors of this type

is given by

$$P(n) = \sum_{m,l=0}^{\infty} \left(\frac{k^+}{k_T}\right)^m \left(\frac{k^-}{k_T}\right)^l \left(\frac{\gamma}{k_T}\right) \frac{(m+l)!}{m!l!} \delta(n-m+l) \tag{A.38}$$

The previous equation for runlength distribution can be broken into $P(n) = P(n \geq 0) + P(n < 0)$, where

$$P(n \geq 0) = \left(\frac{k^+}{k_T}\right)^n \left(\frac{\gamma}{k_T}\right) \sum_{l=0}^{\infty} \left(\frac{k^+k^-}{k_T^2}\right)^l \frac{(2l+n)!}{(n+l)!l!} \tag{A.39}$$

$$= \left(\frac{k^+}{k_T}\right)^n \left(\frac{\gamma}{k_T}\right) {}_2F_1\left(\frac{1+n}{2}, \frac{2+n}{2}; 1+n; 4\frac{k^+k^-}{k_T^2}\right), \tag{A.40}$$

when $m \geq l$. In the above equation, ${}_2F_1$ is a Gaussian hypergeometric function, a special function represented by the hypergeometric series, defined as:

$${}_2F_1\left(\frac{1+n}{2}, \frac{2+n}{2}; 1+n; 4\frac{k^+k^-}{k_T^2}\right) = \sum_{l=0}^{\infty} \frac{\left(\frac{1+n}{2}\right)_l \left(\frac{2+n}{2}\right)_l}{(1+n)_l} \frac{\left(4\frac{k^+k^-}{k_T^2}\right)^l}{l!} \tag{A.41}$$

$$= \sum_{l=0}^{\infty} \left(\frac{k^+k^-}{k_T^2}\right)^l \frac{(2l+n)!}{(n+l)!l!} \tag{A.42}$$

Using a property of the Gauss hypergeometric series [169] page 556, we obtain

$$P(n \geq 0) = \left(\frac{k^+}{k_T}\right)^n \left(\frac{\gamma}{k_T}\right) \left(\frac{2}{1+\sqrt{1-4\frac{k^+k^-}{k_T^2}}}\right)^n \frac{1}{\sqrt{1-4\frac{k^+k^-}{k_T^2}}} \tag{A.43}$$

$$= \left(\frac{2k^+}{k_T+\sqrt{k_T^2-4k^+k^-}}\right)^n \frac{\gamma}{\sqrt{k_T^2-4k^+k^-}}, \tag{A.44}$$

and

$$P(n < 0) = \left(\frac{2k^-}{k_T+\sqrt{k_T^2-4k^+k^-}}\right)^n \frac{\gamma}{\sqrt{k_T^2-4k^+k^-}}. \tag{A.45}$$

A.3.2 Velocity distribution, $P(v)$

In order to obtain the velocity distribution function we need the joint distribution $P(m, l, t)$ for the motor to to take m steps forward, l steps backward, and

then detach. Using the same method as in the previous simpler case, we obtain

$$P(m,l,t) = \frac{t^{m+l}}{m!l!}(k^+)^m(k^-)^l\gamma\exp(-k_T t). \tag{A.46}$$

Again, by setting $n = m - l$ we have

$$\begin{aligned}
P(n,t) &= \sum_{m,l=0}^{\infty}\delta(n-m+l)P(m,l,t), && \text{(A.47)}\\
P(n>0,t) &= (k^+t)^n\gamma\exp(-k_T t)\sum_{l=0}^{\infty}\frac{t^{2l}(k^+k^-)^l}{(n+l)!l!} && \text{(A.48)}\\
&= (k^+t)^n\gamma\exp(-k_T t)(t^2k^+k^-)^{-\frac{n}{2}}I_n(2t\sqrt{k^+k^-}) && \text{(A.49)}\\
&= \left(\frac{k^+}{k^-}\right)^{\frac{n}{2}}\gamma\exp(-k_T t)I_n(2t\sqrt{k^+k^-}), && \text{(A.50)}\\
P(n<0,t) &= \left(\frac{k^-}{k^+}\right)^{\frac{n}{2}}\gamma\exp(-k_T t)I_n(2t\sqrt{k^+k^-}), && \text{(A.51)}
\end{aligned}$$

where I_n is the modified Bessel function of the first kind, defined as

$$I_n(x) = \sum_{l=0}^{\infty}\frac{1}{(n+l)!l!}\left(x\right)^{2l+n}. \tag{A.52}$$

The velocity distribution is obtained as

$$\begin{aligned}
P(v) &= \sum_{n=-\infty}^{\infty}\int_0^{\infty}dt\delta(v-\frac{n}{t})P(n,t) && \text{(A.53)}\\
&= \sum_{n=-\infty}^{0}\int_0^{\infty}dt\delta(v-\frac{n}{t})P(n,t) + \sum_{n=0}^{\infty}\int_0^{\infty}dt\delta(v-\frac{n}{t})P(n,t) && \text{(A.54)}\\
&\quad + \int_0^{\infty}dt\delta(v-\frac{n}{t})P(n,t)|_{n=0} && \text{(A.55)}\\
&= \sum_{n=0}^{\infty}\frac{n}{v^2}P(n,\frac{n}{v})\theta(v-0) + \sum_{n=-\infty}^{0}\frac{-n}{v^2}P(n,\frac{n}{v})\theta(0-v) && \text{(A.56)}\\
&\quad + P(n=0) && \text{(A.57)}\\
&= P(v>0) + P(v<0) + P(v=0). && \text{(A.58)}
\end{aligned}$$

We can break up the expression for $P(v)$ into positive and negative n since only negative n terms give negative velocities while positive n terms give positive veloc-

ities (and $P(v=0) = P(n=0)$). Now let us consider the term $P(v>0)$ first. Substitute Eq. A.50 into this equation

$$P(v>0) = \sum_{n=0}^{\infty} \frac{n}{v^2} \left(\frac{k^+}{k^-}\right)^{\frac{n}{2}} \gamma \exp(-k_T \frac{n}{v}) I_n(\frac{2n}{v}\sqrt{k^+k^-}) \quad \text{(A.59)}$$

$$= \sum_{n=0}^{\infty} B \cdot n \cdot C^n \cdot I_n(\frac{2n}{v}\sqrt{k^+k^-}), \quad \text{(A.60)}$$

where

$$B = \frac{\gamma}{v^2} \qquad \text{and} \qquad C = \sqrt{\frac{k^+}{k^-}} e^{-\frac{k_T}{v}} \quad \text{(A.61)}$$

Converting the Bessel I function into the hypergeometric function ${}_0\mathrm{F}_1$, we have

$$P(v>0) = \sum_{n=0}^{\infty} B \cdot n \cdot C^n \cdot \left(\frac{n}{v}\sqrt{k^+k^-}\right)^n \frac{1}{n!} {}_0F_1\left(;n+1;\frac{n^2}{v^2}k^+k^-\right) \quad \text{(A.62)}$$

$$= \frac{\gamma}{v} \sum_{n=0}^{\infty} \left(\frac{n}{v}\right)^{n+1} \frac{1}{n!} \left(k^+ e^{-\frac{k_T}{v}}\right)^n {}_0F_1\left(;n+1;\frac{n^2k^+k^-}{v^2}\right), \quad \text{(A.63)}$$

where the hypergeometric function is defined as,

$${}_0\mathrm{F}_1\left(;n+1;\frac{n^2k^+k^-}{v^2}\right) = \sum_{l=0}^{\infty} \frac{1}{(n+1)_l} \frac{\left(\frac{n^2k^+k^-}{v^2}\right)^l}{l!}. \quad \text{(A.64)}$$

In the case where the backward rate of the motor is small ($k^- \to 0$), the hypergeometric function ${}_0\mathrm{F}_1(;n+1;0) \to 1$, Eq. A.63 reverts back to Eq. A.32 that we obtained previously in Model 2.

For $P(v<0)$, similarly, we have

$$P(v<0) = \sum_{n=-\infty}^{0} B \cdot (-n) \cdot C^n \cdot I_n(\frac{2n}{v}\sqrt{k^+k^-}). \quad \text{(A.65)}$$

Setting $n = -h$, we obtain

$$P(v<0) = \sum_{h=\infty}^{0} B \cdot h \cdot C^{-h} \cdot I_{-h}(\frac{-2h}{v}\sqrt{k^+k^-}) \quad \text{(A.66)}$$

$$= \sum_{h=0}^{\infty} B \cdot h \cdot C^{-h} \cdot (-1)^h I_h(\frac{2h}{v}\sqrt{k^+k^-}). \quad \text{(A.67)}$$

Converting the modified Bessel I function in Eq. A.67 to the hypergeometric function ${}_0\mathrm{F}_1$, and changing the argument for the sum back to n, we have

$$P(v<0) = \sum_{n=0}^{\infty} B\cdot n\cdot C^{-n}\cdot(-1)^n\left(\frac{n}{v}\sqrt{k^+k^-}\right)^n\frac{1}{n!}{}_0F_1\left(;n+1;\frac{n^2}{v^2}k^+k^-\right) \quad (A.68)$$

$$= \sum_{n=0}^{\infty}\left(\frac{\gamma}{v^2}\right)\left(\sqrt{\frac{k^-}{k+}}\right)^n e^{n\frac{k_T}{v}}\left(-\frac{n}{v}\sqrt{k^+k^-}\right)^n\frac{n}{n!}{}_0F_1\left(;n+1;\frac{n^2}{v^2}k^+k^-\right) \quad (A.69)$$

$$= \frac{\gamma}{-v}\sum_{n=0}^{\infty}\left(\frac{n}{-v}\right)^{n+1}\frac{1}{n!}\left(k^-e^{\frac{k_T}{v}}\right)^n{}_0F_1\left(;n+1;\frac{n^2k^+k^-}{v^2}\right) \quad (A.70)$$

Eq. A.58 along with Eq. A.63 and Eq. A.70, give the exact expression for the velocity distribution function. Working out an approximate expression for the hypegeometric function, we get

$${}_0\mathrm{F}_1(;n+1;an^2)\approx(1+4a)^{-1/4}\exp(nc) \quad (A.71)$$

where

$$c=\sqrt{1+4a}-1-\ln\left(\frac{1+\sqrt{1+4a}}{2}\right) \quad (A.72)$$

and $a=\frac{k^+k^-}{v^2}$. Once again, as in the simpler model with no back-stepping, we can obtain a simple expression for the velocity distribution in the limit $\gamma << k^+$. Using the approximation for the hypergeometric function and the second order Stirling's approximation $n!=\sqrt{2\pi n}(n/e)^n$, we obtain

$$P(v>0) = \frac{\gamma}{v^2\sqrt{2\pi}}(1+4a)^{-1/4}\sum_{n=0}^{\infty}\sqrt{n}A^n \quad (A.73)$$

where the argument A is

$$A = \frac{k^+}{v}\exp\left(c+1-\frac{k_T}{v}\right) \quad (A.74)$$

$$= \frac{k^+}{v}\exp\left(\sqrt{1+4a}-\ln\left(\frac{1+\sqrt{1+4a}}{2}\right)-\frac{k_T}{v}\right). \quad (A.75)$$

Converting the sum in Eq. A.73 into an integral (see Eq. A.35), we obtain

$$P_A(v>0) = \frac{\gamma}{v^2}(1+4a)^{-1/4}(-2\ln(A))^{-3/2} \tag{A.76}$$

$$= \frac{\gamma}{v^2}(1+4a)^{-1/4}\left[2\ln\left(\frac{1+\sqrt{1+4a}}{2}\cdot\frac{v}{k^+}\right)-2\sqrt{1+4a}+2\frac{k_T}{v}\right]^{-3/2} \tag{A.77}$$

Similarly

$$P_A(v<0) = \frac{\gamma}{v^2}(1+4a)^{-\frac{1}{4}}\left[2\ln\left(\frac{1+\sqrt{1+4a}}{2}\cdot\frac{|v|}{k^-}\right)-2\sqrt{1+4a}+2\frac{k_T}{|v|}\right]^{-\frac{3}{2}} \tag{A.78}$$

Eq. A.77 is a useful approximation equation that we can use to analyze data under zero or small load. Fig A.6 shows the comparison between our analytical results with a sets of Kinetic Monte Carlo simulation [168].

A.4 Model 4 – Modeling stepping with an intermediate state

We consider a model with a chemical intermediate state (Fig. A.7). Each motor steps from site i to site $i+1$, which is physically 8.2 nm away along the track in the Kinesin case, by visiting a single chemical intermediate state. In any state, the motor can take a sub-step to the next state with the same forward rate k_s^+, backward rate k_s^-, and detachment rate γ.

The probability for the motor to take a single step forward, by passing the intermediate state and subsequently binding to the next binding site, is $P_i(1,t) = t(k_s^+)^2 e^{-k_T t}$ where $k_T = k_s^+ + k_s^- + \gamma$. The probability for the motor to take m step

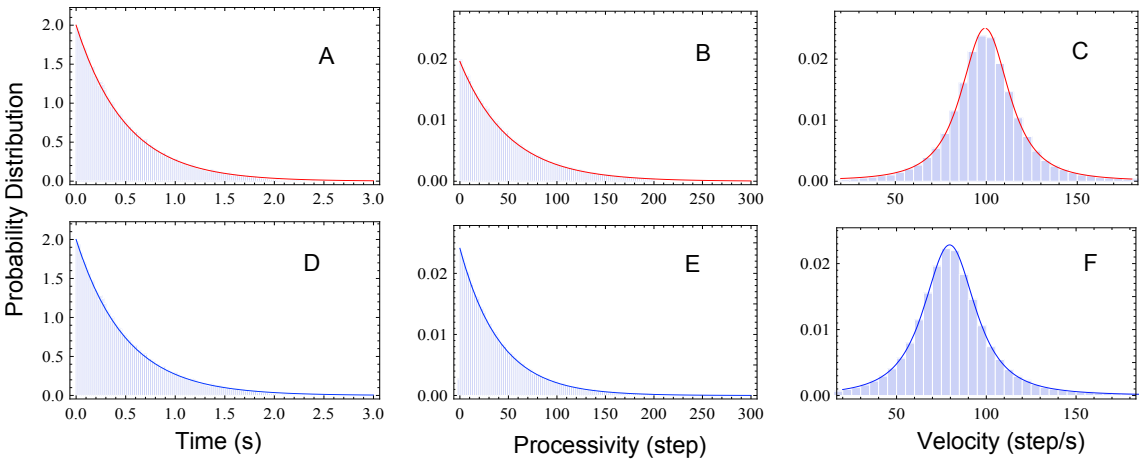


Figure A.6: Comparison of theory and Kinetic Monte Carlo simulations. **A**–**C**: No backward rate and $k^+ = 100$ /s, $\gamma = 2$ /s. **D**–**F**: $k^+ = 100$ /s, $k^- = 20$ /s, $\gamma = 2$ /s. Red and blue lines are predicted functions from the analytical model, while the blue histograms are obtained from Kinetic Monte Carlo simulations. **A**, **D** (Eq. A.21): $P(t)$ - probability distribution of time before detaching. **B** (Eq. A.22), **E** (Eq. A.44): $P(n)$ - probability distribution of number of steps that the motor takes before detaching. **C** (Eq. A.36), **F** (Eq. A.77): $P_A(v)$ - probability distribution of velocity.

forward, l step backward and then detach from the track at time t is,

$$P_i(m,l,t) = \left[\frac{t^{2(m+l)}}{(2m)!(2l)!}(k_s^+)^{2m}(k_s^-)^{2l} + \frac{t^{2(m+l)+1}}{(2m+1)!(2l)!}(k_s^+)^{2m+1}(k_s^-)^{2l} + \frac{t^{2(m+l)+1}}{(2m)!(2l+1)!}(k_s^+)^{2m}(k_s^-)^{2l+1} + \frac{t^{2(m+l+1)}}{(2m+1)!(2l+1)!}(k_s^+)^{2m+1}(k_s^-)^{2l+1}\right]\gamma\exp(-k_T t) \tag{A.79}$$

Thus, the probability for the motor to complete $n = (m-l)$ steps in time t becomes,

$$\begin{aligned} P_i(n,t) &= \sum_{m,l=0}^{\infty} \delta_{(m-l),n} P_i(m,l,t) \\ &= \gamma e^{-k_T t} t^{2n} (k_s^+)^{2n} \sum_{l=0}^{\infty}\left[\frac{t^{4l}(k_s^+k_s^-)^{2l}}{(2n+2l)!(2l)!} + tk_s^+\frac{t^{4l}(k_s^+k_s^-)^{2l}}{(2n+2l+1)!(2l)!}\right. \\ &\quad \left. + tk_s^-\frac{t^{4l}(k_s^+k_s^-)^{2l}}{(2n+2l)!(2l+1)!} + t^2k_s^+k_s^-\frac{t^{4l}(k_s^+k_s^-)^{2l}}{(2n+2l+1)!(2l+1)!}\right] \end{aligned} \tag{A.80}$$

Just as before, we partition the velocity distribution as $P_i(v) = P_i(v > 0) + P(v =$

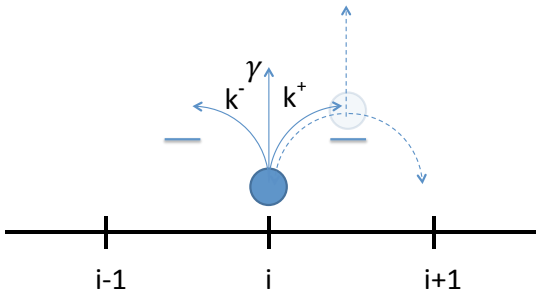


Figure A.7: Kinetic model for stepping with a chemical intermediate state at every step.

$0) + P_i(v < 0)$ where

$$P_i(v>0) = \frac{\gamma}{v}\sum_{n=0}^{\infty}\left[\left(\frac{n}{v}\right)^{2n+1}\left((k_s^+)^2 e^{-\frac{k_T}{v}}\right)^n \frac{1}{2k_s^+k_s^-n}\left(-(k_s^+ + k_s^-)v_0 F_1\left[2n, -\frac{k_s^+k_s^-n^2}{v^2}\right]\right.\right.$$
$$+ (k_s^+ + k_s^-)v_0\mathbf{F}_1\left[2n, \frac{k_s^+k_s^-n^2}{v^2}\right] + 2k_s^+ n v_0\mathbf{F}_1\left[2n+1, -\frac{k_s^+k_s^-n^2}{v^2}\right]$$
$$\left.\left. + 2k_s^+ n(k_s^- - v)_0\mathbf{F}_1\left[2n+1, \frac{k_s^+k_s^-n^2}{v^2}\right]\right)\right], \tag{A.81}$$

and

$$P_i(v<0) = \frac{\gamma}{-v}\sum_{n=0}^{\infty}\left[\left(\frac{n}{-v}\right)^{2n+1}\left((k_s^-)^2 e^{\frac{k_T}{v}}\right)^n \frac{1}{2k_s^+k_s^-n}\left((k_s^+ + k_s^-)v_0\mathbf{F}_1\left[2n, -\frac{k_s^+k_s^-n^2}{v^2}\right]\right.\right.$$
$$- (k_s^+ + k_s^-)v_0\mathbf{F}_1\left[2n, \frac{k_s^+k_s^-n^2}{v^2}\right] - 2k_s^- n v_0\mathbf{F}_1\left[2n+1, -\frac{k_s^+k_s^-n^2}{v^2}\right]$$
$$\left.\left. + 2k_s^- n(k_s^+ + v)_0\mathbf{F}_1\left[2n+1, \frac{k_s^+k_s^-n^2}{v^2}\right]\right)\right]. \tag{A.82}$$

A.5 Velocity distribution of molecular motors - model with fixed run-length and non-zero back-stepping rate

A.5.1 Velocity distribution $P_N(v)$ for homogeneous helicases taking N steps

Let $P(m,l)$ be the probability that the motor takes m steps forward and l steps backward and until reaches site $n = m - l$. For this model, $P(m,l)$ is given as

$$P(m,l) = \left(\frac{k^+}{k_T}\right)^m \left(\frac{k^-}{k_T}\right)^l \frac{(m+l)!}{m!l!}\frac{(m-l)}{(m+l)}, \tag{A.83}$$

where $k_T = k^+ + k^-$ is the total rate, $\frac{k^\pm}{k_T}$ is the probability of taking a forward (backward) step at any site, and the $\frac{(m+l)!}{m!l!}$ term is the binomial coefficient accounting for the number of ways the motor can take m forward, l backward steps, and the final term $\frac{(m-l)}{(m+l)}$ is accounted for an absorbing barrier [170].

To find an expression for $P(v = \frac{N}{t})$, we first calculate the joint distribution $P(m,l,t)$ for the motor to take m steps forward, l steps backward, and then detach at t due to the presence of an absorbing barrier:

$$\begin{aligned} P(m,l,t) &= \left(\frac{k^+}{k_T}\right)^m \left(\frac{k^-}{k_T}\right)^l \frac{(m+l)!}{m!l!}\frac{(m-l)}{(m+l)}\frac{t^{m+l-1}}{(m+l-1)!}k_T^{m+l}\exp(-k_T t) && \text{(A.84)} \\ &= \frac{t^{m+l-1}}{m!l!}(m-l)(k^+)^m(k^-)^l\exp(-k_T t). && \text{(A.85)} \end{aligned}$$

Setting $n = m - l$, which is the binding site difference with the initial position,

we obtain,

$$P(n,t) = \sum_{m,l=0}^{\infty} \delta(n-m+l)P(m,l,t) \quad \text{(A.86)}$$

$$= \frac{(k^+t)^n}{t}\exp(-k_Tt)n\sum_{l=0}^{\infty}\frac{t^{2l}(k^+k^-)^l}{(n+l)!l!} \quad \text{(A.87)}$$

$$= \frac{(k^+t)^n}{t}\exp(-k_Tt)\frac{n}{n!}{}_0\mathrm{F}_1(;n+1;t^2k^+k^-), \quad \text{(A.88)}$$

where the hypergeometric function ${}_0\mathrm{F}_1$ is defined as:

$${}_0\mathrm{F}_1(;n+1;t^2k^+k^-) = \sum_{l=0}^{\infty}\frac{n!}{(n+l)!}\frac{(t^2k^+k^-)^l}{l!} \quad \text{(A.89)}$$

In experiments, the length of DNA, N_{DNA}, is fixed. In such single moelecule experiments, time t is measured when each motor reaches binding site N^{th} ($N = \frac{N_{DNA}}{s}$). In this case, the velocity distribution $P_N(v)$ is given by,

$$P_N(v) = \sum_{n=0}^{\infty}\int_0^{\infty}\mathrm{d}tP(n=N,t)\delta\left(v-\frac{N}{t}\right) \quad \text{(A.90)}$$

$$= \frac{N}{v^2}P\left(N,\frac{N}{v}\right) \quad \text{(A.91)}$$

$$= \frac{N}{v^2}\left(\frac{N}{v}\right)^{N-1}\left(k^+\exp(-\frac{k_T}{v})\right)^N\frac{N}{N!}{}_0\mathrm{F}_1\left(;N+1;\frac{N^2}{v^2}k^+k^-\right) \quad \text{(A.92)}$$

$$= \left(\frac{N}{v}\right)^{N+1}\left(k^+\exp(-\frac{k_T}{v})\right)^N\frac{1}{N!}{}_0\mathrm{F}_1\left(;N+1;\frac{N^2}{v^2}k^+k^-\right). \quad \text{(A.93)}$$

This is an exact expression $P_N(v)$ for homogeneous motors with identical same kinetic rates taking N steps (same as Eq. 3.1 in the Chap. 3).

A.6 Velocity distribution for motors with static disorder - model with fixed runlength and Gaussian distribution of forward rates

First, let us work out an approximate equation for the velocity distribution (Eqs. 3.1, 3.2).

From Eq. 3.2, we obtain

$$P_N^A(v) \approx \left(\frac{N}{v}\right)^{n+1} \frac{1}{N!}(1+4a)^{-1/4} \exp\left[N\left(c - \frac{k_T}{v} + \mathrm{Ln}k^+\right)\right] \quad (A.94)$$

$$= \left(\frac{N}{v}\right)^{n+1} \frac{1}{N!}(1+4a)^{-1/4} \exp[N\ f(k^+)] \quad (A.95)$$

where $f(k^+) = c - \frac{k_T}{v} + \mathrm{Ln}k^+$. For N large, function $f(k^+)$ can be expand around its peak at k_0^+. After performing some algebra, it can be shown that

$$f'(k^+) = \frac{2k^-}{v^2(1+\sqrt{1+4a})} - \frac{1}{v} + \frac{1}{k^+}. \quad (A.96)$$

Set $f'(k_0^+) = 0$ we have $k_0^+ = k^- + v$. Then $f(k_0^+) = -1 + \mathrm{Ln}(v)$ and the second derivative at k_0^+ is $f"(k_0^+) = -\frac{1}{v(2k^-+v)}$. We finally obtain $P_N^A(v)$ for large N_{large},

$$\begin{aligned} P_{Nlarge}^A(v) &\approx \left(\frac{N}{v}\right)^{n+1} \frac{(1+4a)^{-1/4}}{N!} \exp\left[N\big(f(k_0^+) + \frac{1}{2}f"(k_0^+)(k^+ - k_0^+)^2\big)\right] \\ &= \left(\frac{N}{v}\right)^{n+1} \frac{(1+4a)^{-1/4}}{N!} \exp\left[N\left(-1 + \mathrm{Ln}(v) - \frac{(k^+ - k^- - v)^2}{2v(2k^- + v)}\right)\right] \\ &= \left(\frac{N}{v}\right)^{n+1} \frac{(1+4a)^{-1/4}}{N!} \exp\left[N\left(-1 + \mathrm{Ln}(v) - \frac{(k^+ - k^- - v)^2}{2v(2k^- + v)}\right)\right] \\ &= \left(\frac{N}{v}\right)^{n+1} \frac{(1+4a)^{-1/4} v^N}{N!\mathrm{e}^N} \exp\left[N\left(-\frac{(k^+ - k^- - v)^2}{2v(2k^- + v)}\right)\right] \quad (A.97) \end{aligned}$$

Using Stirling's approximation $N! \approx \sqrt{2\pi N}(\frac{N}{\mathrm{e}})^N$, we can rewrite the above equation as,

$$P_{Nlarge}^A(v) \approx \sqrt{\frac{N}{2\pi}}\left[\frac{(1+4a)^{-1/4}}{v} \exp\left(-\frac{N(k^+ - k^- - v)^2}{2v(2k^- + v)}\right)\right]. \quad (A.98)$$

If $k^+ \gg k^-$, then the major contributions to P_{Nlarge}^A comes from value at around $k_0^+ = k^- + v$, which means $v \sim k^+ \gg k^-$. Therefore, we can further approximate $(1+4a)^{-1/4} = (1 + \frac{4k^+k^-}{v^2})^{-1/4} \approx 1$, which leads to the velocity distribution,

$$P_{Nlarge}^A(v) \approx \sqrt{\frac{N}{2\pi}}\left[\frac{1}{v} \exp\left(-\frac{N(k^+ - k^- - v)^2}{2v(2k^- + v)}\right)\right]. \quad (A.99)$$

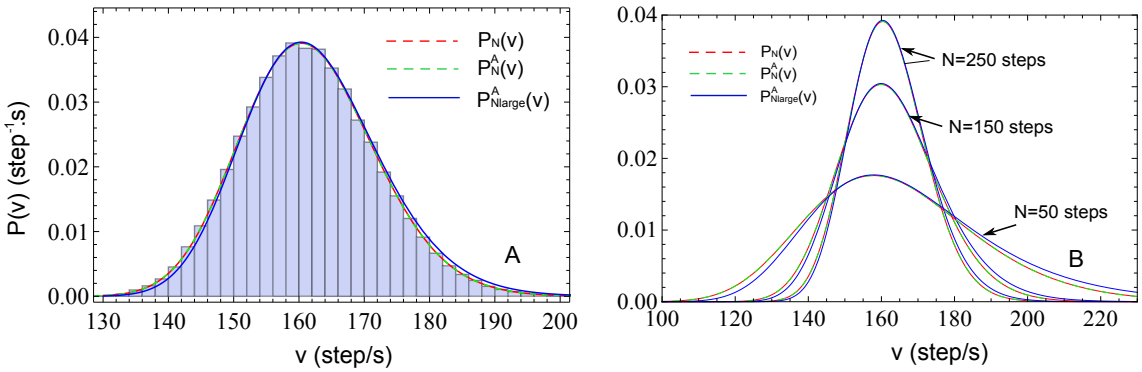


Figure A.8: Comparison of exact ($P_N(v)$ - Eq. 3.1) and approximation theory results ($P^A_N(v)$ - Eq. 3.2 and $P^A_{Nlarge}(v)$ - Eq. A.99) with Kinetic Monte Carlo simulations (blue histogram) (**A**); and with different number of steps (**B**).

Fig. A.8 shows the approximate form of velocity distribution (Eqs. 3.2 and A.99) agrees well with the exact form (Eq. 3.1) with the parameters ectracted for helicase T7 [23]. The exact velocity distribution $P_N(v)$ (dashed red, Eq. 3.1) coincides with the first approximate equation $P^A_N(v)$ (dashed green, Eq. 3.2) and the further approximate equation $P^A_{Nlarge}(v)$ (blue, Eq. A.99). All of the lines fit well with the histogram from a Kinetic Monte Carlo simulation. Fig. A.8B shows the 3 functions with different values of the number of steps as an illustration of the n needed to reach the P^A_{Nlarge} limit. Surprisingly, even for $N = 50$ steps, the approximation is very good.

With this new approximation of velocity distribution, the integral in Eq. 3.3 can be performed exactly as an integral of a product of two different Gaussian distributions for k^+. With this, we arrive at the analytical expression for velocity

distribution with a normal distributed forward rate as,

$$P_N^{H,A}(v) = \frac{\sqrt{N}}{\sqrt{2\pi v}} \frac{\mathrm{e}^{-\frac{N(k^- + v - \mu_k)^2}{2(N\sigma_k^2 + v(2k^- + v))}}}{\sqrt{\frac{N\sigma_k^2 + v(2k^- + v)}{(2k^- + v)}}}. \tag{A.100}$$

This is Eq. 3.4 in Chap. 3

Appendix B: Supporting Information for Chapter 2

B.1 Estimation of the K-1 detachment rate, γ_0

To estimate the detachment rate of kinesin, we consider the pathway in Fig B.1, where K-1 from a State 1 with both heads bound to the MT, can detach one head at a time (States 2 and 3), and before fully dissociating from the track. The detachment rate of the two-headed kinesin is $k_{total} = k_{14} = \gamma$ that can be estimated from the microscopic rates k_{ij}, defined in Fig. B.1. We assume that once the two heads dissociate, the rates for reattachment to the MT are zero ($k_{43} = k_{42} = 0$). The Laplace transform $\tilde{\pi}_i(s)$ of $\pi_i(t)$, which is probability density distribution that

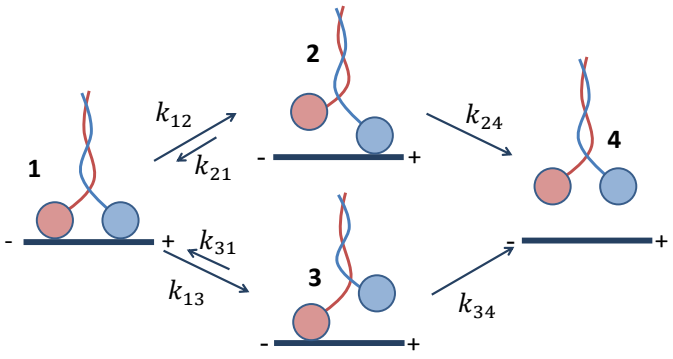


Figure B.1: Enzymatic pathway used to estimate the K-1 detachment rate, γ.

the system goes from state i to the final dissociated state 4 in time t [45], is

$$\tilde{\pi}_2(s) = \frac{k_{21}\tilde{\pi}_1(s) + k_{24}\tilde{\pi}_4(s)}{k_{21} + k_{24} + s} \tag{B.1}$$

$$\tilde{\pi}_3(s) = \frac{k_{31}\tilde{\pi}_1(s) + k_{34}\tilde{\pi}_4(s)}{k_{31} + k_{34} + s} \tag{B.2}$$

$$\tilde{\pi}_1(s) = \frac{k_{12}\tilde{\pi}_2(s) + k_{13}\tilde{\pi}_3(s)}{k_{12} + k_{13} + s} \tag{B.3}$$

$$\tilde{\pi}_4(s) = 1. \tag{B.4}$$

Solving for the distribution $\tilde{\pi}_{total}(s) = \tilde{\pi}_1(s)$ yields

$$\tilde{\pi}_{total}(s) = \frac{k_{13}k_{34}(k_{21} + k_{24} + s) + k_{12}k_{24}(k_{31} + k_{34} + s)}{k_{12}(k_{24} + s)(k_{31} + k_{34} + s) + (k_{21} + k_{24} + s)(k_{13}(k_{34} + s) + s(k_{31} + k_{34} + s))}. \tag{B.5}$$

The average detachment time $< t >= \gamma^{-1}$ is the first moment of the generating function $\tilde{\pi}_{total}(s)$,

$$< t >= \frac{1}{\gamma} = \left(-\frac{d\tilde{\pi}_{total}(s)}{ds}\right)\Big|_{s=0} = \frac{k_{13}(k_{21} + k_{24}) + (k_{12} + k_{21} + k_{24})(k_{31} + k_{34})}{k_{13}(k_{21} + k_{24})k_{34} + k_{12}k_{24}(k_{31} + k_{34})} \tag{B.6}$$

In the limit $k_{21} \to 0$ and $k_{31} \to 0$, which would result in an over estimate of k_{14}, we obtain,

$$\gamma < k_{14} = \frac{k_{13}k_{24}k_{34} + k_{12}k_{24}k_{34}}{k_{13}k_{24} + (k_{12} + k_{24})k_{34}}. \tag{B.7}$$

The microscopic rates k_{ij} for the K-1 have been reported in an experimental paper [171]. We expect the detachment rate of the kinesin trailing head (State 2) to be faster so we set $k_{12} = 48s^{-1}$ from their measurements. All other rates are set to the one-headed detachment rate $k_{13} = k_{24} = k_{34} = 3s^{-1}$. With these values, the estimated detachment rate of K-1 is $\gamma_0 = 2.8s^{-1}$. This value is in a very good agreement with $\gamma_0 = 2.3s^{-1}$ extracted by using our theory to simultaneously fit the

experimental data for $P(v)$ and $P(n)$.

Appendix C: Supporting Information for Chapter 3

C.1 Model of NA unwinding by a helicase and application for T4 gp41

In the study by Chakrabarti *et al.* [23], it was shown that only by fitting the force dependences of both the unwinding velocity and run-length, one can obtain robust parameters for a finitely processive helicase. Lionnet *et al.*'s paper [16] report that the runlength of T4 gp41 as 178 ± 48 bp at $F = 9.7 \pm 3$ pN and 2 mM ATP concentration. In the absence of additional data, we choose to fit Chakrabarti *et al.*'s model to the force dependence of the unwinding rate reported for the same ATP concentration 2 mM. Subsequently, analyze the parametes with the data point to get force-dependent processivity.

From their model [23] (CJT), the unwinding rate v_u and processivity $< d >$ depend on several parameters (See the scheme of their model in Fig. C.1 for visual picture of some of the parameters):

- g - representing the fractional position of the free energy barrier between base-pair opened and base-pair closed states ($0 \ll g \ll 1$)
- s - step-size of helicase (number of base-pairs)

- u - step-height of the interaction potential between helicase and d.s junction (in $k_B T$)

- r - range of interaction between helicase and d.s junction (number of base pair)

- r_f - forward s.s translocation rate of helicase with the driving force arising from NTP/dNTP hydrolysis. Normally, this rate can be converted from average s.s translocation rate by devided by the step-size ($r_f = \frac{<v_{ss}>}{s}$). When analyzing T4 gp41 helicase's data, we use the average s.s translocation velocity, which is equal to the mean rezipping velocity. At 2 mM ATP concentration, the measured value is $< v_{ss} >= 260 \pm 30$ bp/s [16]

- k - pure diffusion forward and backward s.s translocation rates of helicase ($k_+ = k_- = k$) Note here that we distinguished these diffusion forward and backward rates in the CJT model (k_+ and k_-) with our forward and backward rates in our model (k^+ and k^-). They related to each other as $k^+ = r_f + k_+ = r_f + k$ and $k^- = k_- = k$

- k_d - detachment rate of helicase from s.s in translocation (which is the same as γ in our model). For T4 gp41, the detachment rate is estimated to be $0.006\,s^{-1}$ (see below)

- b - ratio of the nucleic acid breathing rates under force when the helicase is far away from the junction ($b = \alpha/\beta = e^{-(\Delta G - \Delta G_F)}$)

- ΔG - stability of the particular d.s base pair. To limit one more fitting pa-

rameter, here, we take $\Delta G = 2.27\,k_BT$ as a value averaged over the sequence for the DNA substrate with 42% GC content that was used in the force dependence experiment [16] with $E_{GC} = 3.2\,k_BT$ and $E_{AT} = 1.6\,k_BT$ [172]. (We choose this value over $\Delta G = 1.95\,k_BT$ in Lionnet *et al.* [16] obtained from other parameters of E_{GC} and E_{AT} because it gives us better fits.)

- ΔG_F - force dependence free energy contribution to the free energy cost to open one base pair ($\Delta G_F = 2\frac{z_{max}}{l_K} \ln\left(\frac{1}{Fl_K}\sinh(Fl_K)\right)$). We use the same contour length per base $z_{max} = 0.58\,\text{nm/bp}$ and Kuhn length $l_K = 2.46\,\text{nm}$ obtained from the experiment for the same DNA substrate [16]

We estimate the detachment rate of T4 gp41 from its dissociation half-life of 1 min, when it translocates on ssDNA at 37 °C [173] (with a complete replisome, this dissociation half-life is reported to be 11 min [174]). Using temperature coefficient Q_{10} to be equals to 2 [175], we estimate the detachment rate of T4 gp41 at 25 °C to be $\gamma = 0.006\,s^{-1}$. This detachment rate is in the same range with the estimated detachment rates of the similar helicase, T7 [23, 176].

After estimating all possible parameters, we fit the unwinding rate $v_u(F)$ with different possible value set of step-size s and range of interaction r to the data to extract 3 parameters g, k, and u. We chose only the fitted parameter sets that give us consistent average processivities with the reported range obtained at $F = 9.7 \pm 3\,\text{pN}$ [16]. The data sets that give good fits are reported in Table C.1. Fig. C.2 shows the fitting line with the parameter set a) from the table.

From Table C.1, we can see that, with a fixed step-size value s, changing the

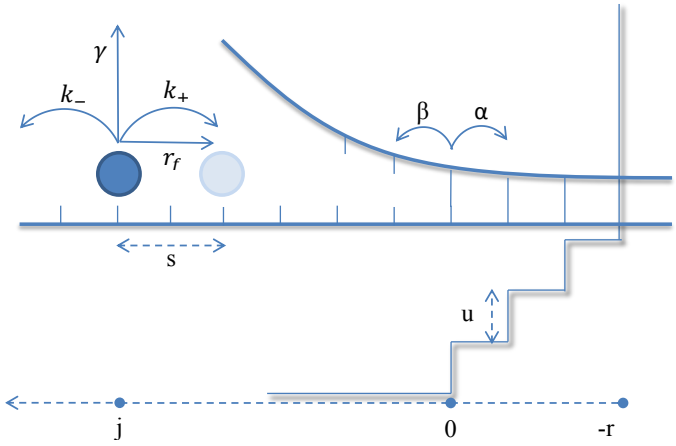


Figure C.1: Model for nucleic acid unwinding by a helicase [23]. In this model, the helicase has step-size $s = 2\,\mathrm{nt}$, and a range of interaction with the d.s junction $r = 3\,\mathrm{nt}$.

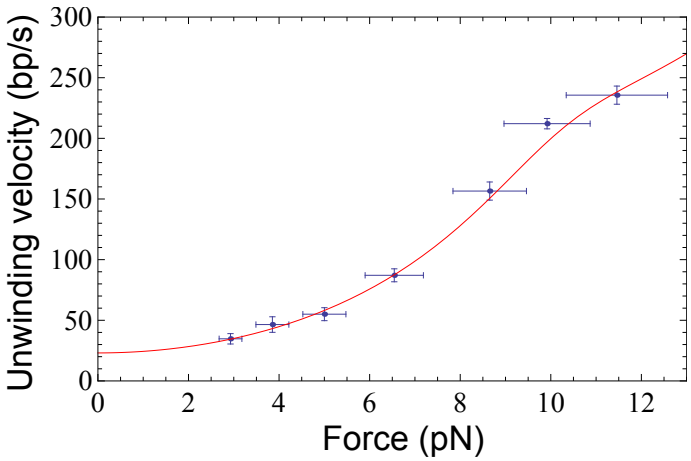


Figure C.2: Force dependence of unwinding velocity $v_u(F)$ data (in blue) at 2 mM ATP concentration [16] with the fitting line (in red) using parameters a) from the Table C.1.

Table C.1: Fitting parameters for T4 gp41

	χ^2	s (nt)	r (nt)	g	k (s^{-1})	u k_BT
a)*	0.61	1	8	0.21	14.0	0.61
b)	0.58	1	9	0.22	13.9	0.61
c)	0.56	1	10	0.23	13.9	0.62
d)	0.50	1	15	0.25	13.8	0.63
e)*	2.48	2	5	0.14	10^{-9}	1.14
f)	3.63	2	7	0.16	10^{-9}	1.15
g)	4.69	2	10	0.17	10^{-9}	1.15
h)	5.43	2	15	0.18	10^{-9}	1.15
i)*	6.76	3	5	0.10	10^{-9}	1.49
j)	8.97	3	6	0.11	10^{-10}	1.47
k)*	4.54	4	4	0.03	10^{-12}	1.82
l)	8.59	4	5	0.06	10^{-12}	1.71
* parameters used to calculate error bars						
a)*	0.61	1	8	0.2 ± 0.3	14 ± 6	0.6 ± 0.3
e)*	2.48	2	5	0.1 ± 0.8	$0 + 0.7$	1.1 ± 0.2
i)*	6.76	3	5	0.10 ± 0.04	$0.00 + 0.06$	1.5 ± 0.2
k)*	4.54	4	4	0.03 ± 0.02	$0.000 + 0.005$	1.8 ± 0.2

value of the interaction range r does not change the estimated parameters much. Indeed, all of those obtained parameters lie within the error-bars of the first fitting parameter set. For instance, for $s = 1$, all possible value of $r \geq 8$ gives us reasonable fits (Table C.1 a-d)). However, all the values of g (same to k, u) are similar and lies within the error bars of Table C.1 a)*. So for each value of s, we can not extract exact value of r, but we can be confident with the other fitting parameters, g, k, and u with the error bars as reported in the table at a), e), i), and k). We can not conclude anything further than that, due to the lack of data for processivity.

In order to apply our model to test for the heterogeneity for T4 gp41, we need the information about step-size s and the rate k of the helicase. The other parameters are not as important. Also, both of these two parameters are independent of ATP concentration. Therefore, we can still use them to analyze s.s translocation

velocity distribution at 5 mM. We perform the test for heterogeneity with all 4 possible sets Table C.1 a), e), i), and k) in the main text. The different parameters in the 2 models were converted as $k^+ = r_f + k_+ = r_f + k$ and $k^- = k_- = k$. As we expected, all of them show that T4 gp41 is heterogeneous, similar to helicase T7.

C.2 Heterogeneity in the diffusion rate does not contribute much to the width of velocity distribution

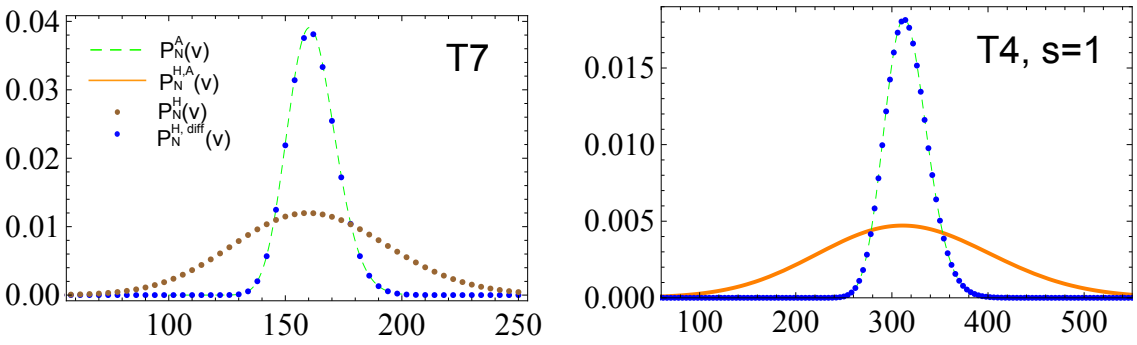


Figure C.3: Heterogeneity effect on the diffusion rate $P_N^{H,diff}$ and the forward rate (or driving rate) $P_N^H(v)$ on velocity distribution.

In Chap. 3, we explained the main model for heterogeneity with a dispersion in the forward rate k^+ alone. Conceptually, that means we have considered only the heterogeneity in the forward rate driving by NTP/dNTP hydrolysis r_f in the CJT model [23]. In this section, we consider the effect of heterogeneity in the diffusion forward and backward s.s translocation rates in the CJT model ($k^- \equiv k_- = k_+ = k$). We calculate numerically the velocity distribution, $P_N^{H,diff}(v)$ as,

$$P_N^{H,diff}(v) = \int_{-\infty}^{\infty} \mathrm{d}k^- P_N(v|k^+ \to r_f + k^-) \frac{1}{\sqrt{2\pi}\sigma_k} \mathrm{e}^{-\frac{(k^- - \mu_k)^2}{2\sigma_k^2}}. \qquad \text{(C.1)}$$

where μ_k and σ_k now are the mean and s.d. of a Gaussian distribution for k^-. Because the diffusion rates contribute to both the forward and backward rates in our model, we substitute all the forward rate k^+ with $k^+ = r_f + k^-$ in the Eq. 3.1. We consider the value of diffusion rate, k^- in this case, distributed normally with a mean μ_k and a s.d. σ_k ($r_f = 161\,\mathrm{step}^{-1}$, $\mu_k = 0.6\,\mathrm{step}^{-1}$, $\sigma_k = 1.6\,\mathrm{step}^{-1}$ for T7 and $r_f = 314\,\mathrm{step}^{-1}$, $\mu_k = 20\,\mathrm{step}^{-1}$, $\sigma_k = 20\,\mathrm{step}^{-1}$ for T4). Here, we choose only the first case of T4 with the fit for the step size $s = 1$. We see in Fig. C.3 that there is no clear effect of the heterogeneity in the diffusion rate on the velocity distribution. Basically $P_N^{H,diff}$ lies exactly on top of the velocity distribution for homogeneous case $P_N^A(v)$.

Bibliography

[1] Oren M. Becker, Alexander D. MacKerell Jr, Benoit Roux, and Masakatsu Watanabe, editors. *Computational Biochemistry and Biophysics*. CRC Press, New York, 1st edition edition, February 2001.

[2] Carolyn J. Lawrence, R. Kelly Dawe, Karen R. Christie, Don W. Cleveland, Scott C. Dawson, Sharyn A. Endow, Lawrence S. B. Goldstein, Holly V. Goodson, Nobutaka Hirokawa, Jonathon Howard, Russell L. Malmberg, J. Richard McIntosh, Harukata Miki, Timothy J. Mitchison, Yasushi Okada, Anireddy S. N. Reddy, William M. Saxton, Manfred Schliwa, Jonathan M. Scholey, Ronald D. Vale, Claire E. Walczak, and Linda Wordeman. A standardized kinesin nomenclature. *J. Cell Biol.*, 167(1):19–22, October 2004.

[3] Kristen J. Verhey, Neha Kaul, and Virupakshi Soppina. Kinesin Assembly and Movement in Cells. *Annu. Rev. Biophys.*, 40(1):267–288, 2011.

[4] Harukata Miki, Yasushi Okada, and Nobutaka Hirokawa. Analysis of the kinesin superfamily: insights into structure and function. *Trends Cell Bio.*, 15(9):467–476, September 2005.

[5] Nobutaka Hirokawa, Yasuko Noda, Yosuke Tanaka, and Shinsuke Niwa. Kinesin superfamily motor proteins and intracellular transport. *Nat. Rev. Mol. Cell Biol.*, 10(10):682–696, October 2009.

[6] Jonathon Howard. *Mechanics of Motor Proteins and the Cytoskeleton*. Sinauer Associates, Sunderland, Mass, 2001.

[7] Maria Spies. *DNA Helicases and DNA Motor Proteins*. Springer Science & Business Media, November 2012.

[8] Masayoshi Nishiyama, Hideo Higuchi, and Toshio Yanagida. Chemomechanical coupling of the forward and backward steps of single kinesin molecules. *Nat. Cell Biol.*, 4(10):790–797, October 2002.

[9] N J Carter and R a Cross. Mechanics of the kinesin step. *Nature*, 435(7040):308–12, May 2005.

[10] Bradley Efron and Robert J. Tibshirani. *An Introduction to The Bootstrap*. Chapman & Hall, London, 1993.

[11] Steven S. Rosenfeld, Geraldine M. Jefferson, and Peter H. King. ATP Reorients the Neck Linker of Kinesin in Two Sequential Steps. *J. Biol. Chem.*, 276(43):40167–40174, October 2001.

[12] Steven S. Rosenfeld, Jun Xing, Geraldine M. Jefferson, Herbert C. Cheung, and Peter H. King. Measuring kinesin's first step. *J. Biol. Chem.*, 277(39):36731–36739, September 2002.

[13] Steven S. Rosenfeld, Jun Xing, Geraldine M. Jefferson, and Peter H. King. Docking and Rolling, a Model of How the Mitotic Motor Eg5 Works. *J. Biol. Chem.*, 280(42):35684–35695, October 2005.

[14] Wilhelm J Walter, Vaclav Beranek, Elisabeth Fischermeier, and Stefan Diez. Tubulin acetylation alone does not affect kinesin-1 velocity and run length in vitro. *PLoS One*, 7(8):e42218, 2012.

[15] Daniel S Johnson, Lu Bai, Benjamin Y Smith, Smita S Patel, and Michelle D Wang. Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell*, 129(7):1299–309, June 2007.

[16] Timothe Lionnet, Michelle M. Spiering, Stephen J. Benkovic, David Bensimon, and Vincent Croquette. Real-time observation of bacteriophage t4 gp41 helicase reveals an unwinding mechanism. *Proc. Natl. Acad. Sci. U. S. A.*, 104(50):19790–19795, 2007.

[17] F. Jon Kull, Elena P. Sablin, Rebecca Lau, Robert J. Fletterick, and Ronald D. Vale. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature*, 380(6574):550–555, April 1996.

[18] Alexey Drozdetskiy, Christian Cole, James Procter, and Geoffrey J. Barton. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.*, 43(W1):W389–W394, July 2015.

[19] Benot Gigant, Weiyi Wang, Birgit Dreier, Qiyang Jiang, Ludovic Pecqueur, Andreas Plckthun, Chunguang Wang, and Marcel Knossow. Structure of a kinesin-tubulin complex and implications for kinesin motility. *Nat. Struct. Mol. Biol.*, 20(8):1001–1007, August 2013.

[20] Jennifer Turner, Robert Anderson, Jun Guo, Christophe Beraud, Robert Fletterick, and Roman Sakowicz. Crystal Structure of the Mitotic Spindle Kinesin Eg5 Reveals a Novel Conformation of the Neck-linker. *J. Biol. Chem.*, 276(27):25496–25502, July 2001.

[21] Adeline Goulet, William M. Behnke-Parks, Charles V. Sindelar, Jennifer Major, Steven S. Rosenfeld, and Carolyn A. Moores. The Structural Basis of Force Generation by the Mitotic Motor Kinesin-5. *J. Biol. Chem.*, 287(53):44654–44666, December 2012.

[22] Adeline Goulet, Jennifer Major, Yonggun Jun, Steven P. Gross, Steven S. Rosenfeld, and Carolyn A. Moores. Comprehensive structural model of the mechanochemical cycle of a mitotic motor highlights molecular adaptations in the kinesin family. *Proc. Natl. Acad. Sci. U.S.A.*, 111(5):1837–1842, February 2014.

[23] Shaon Chakrabarti, Christopher Jarzynski, and D. Thirumalai. Dynamics and universal properties of helicase unwinding of nucleic acids. *unpublished*.

[24] Harvey F. Lodish, Arnold Berk, Chris A. Kaiser, Monty Krieger, Matthew P. Scott, Anthony Bretscher, Hidde Ploegh, and Paul Matsudaira. *Molecular Cell Biology*. Freeman, New York, sixth edition, 2007.

[25] K Svoboda and S M Block. Force and velocity measured for single kinesin molecules. *Cell*, 77(5):773–84, June 1994.

[26] Manfred Schliwa and Gunther Woehlke. Molecular motors. *Nature*, 422(6933):759–765, April 2003.

[27] Daniela Stock, Andrew G. W. Leslie, and John E. Walker. Molecular Architecture of the Rotary Motor in ATP Synthase. *Science*, 286(5445):1700–1705, November 1999.

[28] S Ray, E Meyhfer, R A Milligan, and J Howard. Kinesin follows the microtubule's protofilament axis. *J. Cell Biol.*, 121(5):1083–1093, 1993.

[29] James A. Spudich and Sivaraj Sivaramakrishnan. Myosin VI: An innovative motor that challenged the swinging lever arm hypothesis. *Nat. Rev. Mol. Cell Bio.*, 11(2):128–37, February 2010.

[30] Anthony J Roberts, Takahide Kon, Peter J Knight, Kazuo Sutoh, and Stan a Burgess. Functions and mechanics of dynein motor proteins. *Nat. Rev. Mol. Cell Bio.*, 14(11):713–26, November 2013.

[31] Roop Mallik and Steven P Gross. Molecular Motors: Strategies to Get Along. *Curr. Biol.*, 14(22):R971–R982, October 2004.

[32] Timothy M Lohman, Eric J Tomko, and Colin G Wu. Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat. Rev. Mol. Cell Biol.*, 9(5):391–401, May 2008.

[33] Martin R Singleton, Mark S Dillingham, and Dale B Wigley. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, 76:23–50, January 2007.

[34] H Kojima, E Muto, H Higuchi, and T Yanagida. Mechanics of single kinesin molecules measured by optical trapping nanometry. *Biophys. J.*, 73(4):2012–2022, October 1997.

[35] Sbastien Courty, Camilla Luccardini, Yohanns Bellaiche, Giovanni Cappello, and Maxime Dahan. Tracking individual kinesin motors in living cells using single quantum-dot imaging. *Nano Lett.*, 6(7):1491–1495, 2006.

[36] Sotaro Uemura, Kenji Kawaguchi, Junichiro Yajima, Masaki Edamatsu, Yoko Yano Toyoshima, and Shin Ishiwata. Kinesin microtubule binding depends on both nucleotide state and loading direction. *Proc. Natl. Acad. Sci. U.S.A.*, 99(9):3–4, 2002.

[37] Karel Svoboda, Christoph F Schmidt, Bruce J Schnapp, and Steven M Block. Direct observation of kinesin stepping by optical trapping interferometry. *Nature*, 365(6448):721–727, October 1993.

[38] Adrian N Fehr, Charles L Asbury, and Steven M Block. Kinesin steps do not alternate in size. *Biophys. J.*, 94(3):L20–2, March 2008.

[39] Mark J Schnitzer, Koen Visscher, and Steven M Block. Force production by single kinesin motors. *Nat. Cell Biol.*, 2(October):718–723, 2000.

[40] Bojan Milic, Johan O. L. Andreasson, William O. Hancock, and Steven M. Block. Kinesin processivity is gated by phosphate release. *Proc. Natl. Acad. Sci. U.S.A.*, 111(39):14136–14140, 2014.

[41] Weiyi Wang, Luyan Cao, Chunguang Wang, Benot Gigant, and Marcel Knossow. Kinesin, 30 years later: Recent insights from structural studies. *Protein Sci.*, 24(7):1047–1056, July 2015.

[42] Anatoly B. Kolomeisky and Michael E. Fisher. Molecular motors: A theorist's perspective. *Annu. Rev. Phys. Chem.*, 58(1):675–695, 2007.

[43] Mark J Schnitzer and Steven M Block. Statistical kinetics of processive enzymes. *Cold Spring Harb. Sym.*, 60:793 – 802, 1995.

[44] Yuichi Taniguchi, Masayoshi Nishiyama, Yoshiharu Ishii, and Toshio Yanagida. Entropy rectifies the Brownian steps of kinesin. *Nat. Chem. Biol.*, 1(6):342–347, November 2005.

[45] Joshua W Shaevitz, Steven M Block, and Mark J Schnitzer. Statistical kinetics of macromolecular dynamics. *Biophys. J.*, 89(4):2277–85, October 2005.

[46] Debashish Chowdhury. Stochastic mechano-chemical kinetics of molecular motors: A multidisciplinary enterprise from a physicists perspective. *Phys. Rep.*, 529(1):1–197, August 2013.

[47] M D Betterton and Frank Jülicher. Opening of nucleic-acid double strands by helicases: active versus passive opening. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, 71(1 Pt 1):011904, January 2005.

[48] David Keller and Carlos Bustamante. The Mechanochemistry of Molecular Motors. *Biophysi. J.*, 78(2):541–556, February 2000.

[49] Frank Jülicher, Armand Ajdari, and Jacques Prost. Modeling molecular motors. *Rev. Mod. Phys.*, 69(4):1269–1282, October 1997.

[50] Jan Brugués and Jaume Casademunt. Self-Organization and Cooperativity of Weakly Coupled Molecular Motors under Unequal Loading. *Phys. Rev. Lett.*, 102(11):118104, March 2009.

[51] Ashok Garai, Debashish Chowdhury, and M. D. Betterton. Two-state model for helicase translocation and unwinding of nucleic acids. *Phys. Rev. E*, 77(6):061910, June 2008.

[52] Frank Jülicher and Jacques Prost. Spontaneous Oscillations of Collective Molecular Motors. *Phys. Rev. Lett.*, 78(23):4510–4513, June 1997.

[53] Frank Jülicher and Jacques Prost. Cooperative Molecular Motors. *Phys. Rev. Lett.*, 75(13):2618–2621, September 1995.

[54] J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, June 1977.

[55] N. Go, T. Noguti, and T. Nishikawa. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. U.S.A.*, 80(12):3696–3700, June 1983.

[56] B. Brooks and M. Karplus. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U.S.A.*, 80(21):6571–6575, November 1983.

[57] Michael Levitt, Christian Sander, and Peter S. Stern. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Bio.*, 181(3):423–447, February 1985.

[58] Steve W. Lockless and Rama Ranganathan. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science*, 286(5438):295–299, October 1999.

[59] Ruxandra I. Dima and D. Thirumalai. Determination of network of residues that regulate allostery in protein families using sequence analysis. *Protein Sci.*, 15(2):258–268, February 2006.

[60] Anatoly B. Kolomeisky and B. Widom. A Simplified "Ratchet" Model of Molecular Motors. *J. Stat. Phys.*, 93(3-4):633–645, 1998.

[61] Michael E. Fisher and Anatoly B. Kolomeisky. Simple mechanochemistry describes the dynamics of kinesin molecules. *Proc. Natl. Acad. Sci. U.S.A.*, 98(14):7748–7753, 2001.

[62] Eric C. Dykeman and Otto F. Sankey. Normal mode analysis and applications in biological physics. *J. Phys. Condens Matter*, 22(42):423202, October 2010.

[63] Qiang Cui and Ivet Bahar, editors. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*. Chapman and Hall/CRC, 1 edition edition, December 2005.

[64] Ronald D Vale, Thomas S Reese, and Michael P Sheetz. Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell*, 42(1):39–50, August 1985.

[65] Ahmet Yildiz, Michio Tomishige, Ronald D. Vale, and Paul R. Selvin. Kinesin Walks Hand-Over-Hand. *Science*, December 2003.

[66] Ahmet Yildiz and Paul R. Selvin. Fluorescence Imaging with One Nanometer Accuracy: Application to Molecular Motors. *Acc. Chem. Res.*, 38(7):574–582, July 2005.

[67] S Sack, J Müller, a Marx, M Thormählen, E M Mandelkow, S T Brady, and E Mandelkow. X-ray structure of motor and neck domains from rat brain kinesin. *Biochem.*, 36(51):16155–65, December 1997.

[68] F. Kozielski, S. Sack, A. Marx, M. Thormählen, E. Schönbrunn, V. Biou, A. Thompson, E. M. Mandelkow, and E. Mandelkow. The crystal structure of dimeric kinesin and implications for microtubule-dependent motility. *Cell*, 91(7):985–994, December 1997.

[69] Charles V Sindelar, Mary Jane Budny, Sarah Rice, Nariman Naber, Robert Fletterick, and Roger Cooke. Two conformations in the human kinesin power stroke defined by X-ray crystallography and EPR spectroscopy. *Nat. Struct. Mol. Biol*, 9(11):844–848, November 2002.

[70] Sarah Rice, Abel W. Lin, Daniel Safer, Cynthia L. Hart, Nariman Naber, Bridget O. Carragher, Shane M. Cain, Elena Pechatnikova, Elizabeth M. Wilson-Kubalek, Michael Whittaker, Edward Pate, Roger Cooke, Edwin W. Taylor, Ronald A. Milligan, and Ronald D. Vale. A structural change in the kinesin motor protein that drives motility. *Nature*, 402(6763):778–784, December 1999.

[71] Luyan Cao, Weiyi Wang, Qiyang Jiang, Chunguang Wang, Marcel Knossow, and Benoît Gigant. The structure of apo-kinesin bound to tubulin links the nucleotide cycle to movement. *Nat. Commun.*, 5:5364, 2014.

[72] Alexander Marx, Andreas Hoenger, and Eckhard Mandelkow. Structures of kinesin motor proteins. *Cell Motil. Cytoskeleton*, 66(11):958–966, November 2009.

[73] Zhiguo Shang, Kaifeng Zhou, Chen Xu, Roseann Csencsits, Jared C. Cochran, and Charles V. Sindelar. High-resolution structures of kinesin on microtubules provide a basis for nucleotide-gated force-generation. *Elife*, 3:e04686, November 2014.

[74] Joseph Atherton, Irene Farabella, I.-Mei Yu, Steven S. Rosenfeld, Anne Houdusse, Maya Topf, and Carolyn A. Moores. Conserved mechanisms of microtubule-stimulated ADP release, ATP binding, and force generation in transport kinesins. *Elife*, 3:e03680, September 2014.

[75] Charles V. Sindelar. A seesaw model for intermolecular gating in the kinesin motor protein. *Biophys. Rev.*, 3(2):85–100, June 2011.

[76] Steven M Block. Kinesin motor mechanics: binding, stepping, tracking, gating, and limping. *Biophys. J.*, 92(9):2986–95, May 2007.

[77] Jennetta W Hammond, Dawen Cai, T Lynne Blasius, Zhe Li, Yuyang Jiang, Gloria T Jih, Edgar Meyhofer, and Kristen J Verhey. Mammalian Kinesin-3 Motors Are Dimeric In Vivo and Move by Processive Motility upon Release of Autoinhibition. *PLoS Biol.*, 7(3), 2009.

[78] M. Abdel-Monem, H. Drwald, and H. Hoffmann-Berling. Enzymic unwinding of DNA. 2. Chain separation by an ATP-dependent DNA unwinding enzyme. *Eur. J. Biochem.*, 65(2):441–449, June 1976.

[79] M. Abdel-Monem and H. Hoffmann-Berling. Enzymic unwinding of DNA. 1. Purification and characterization of a DNA-dependent ATPase from Escherichia coli. *Eur. J. Biochem.*, 65(2):431–440, June 1976.

[80] Alexander E. Gorbalenya and Eugene V. Koonin. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Bio.*, 3(3):419–429, June 1993.

[81] Detlef D. Leipe, Eugene V. Koonin, and L. Aravind. Evolution and Classification of P-loop Kinases and Related Proteins. *J. Mol. Bio.*, 333(4):781–815, October 2003.

[82] Maria Manosas, Xu Guang Xi, David Bensimon, and Vincent Croquette. Active and passive mechanisms of helicases. *Nucleic Acids Res*, 38(16):5518–5526, September 2010.

[83] Jeehae Park, Sua Myong, Anita Niedziela-Majka, Kyung Suk Lee, Jin Yu, Timothy M Lohman, and Taekjip Ha. PcrA helicase dismantles RecA filaments by reeling in DNA in uniform steps. *Cell*, 142(4):544–55, August 2010.

[84] David L. Pincus, Shaon Chakrabarti, and D. Thirumalai. Helicase processivity and not the unwinding velocity exhibits universal increase with force. *Biophys. J.*, 109(2):220 – 230, 2015.

[85] Huong T. Vu, Shaon Chakrabarti, Michael Hinczewski, and D. Thirumalai. Discrete Step Sizes of Molecular Motors Lead to Bimodal Non-Gaussian Velocity Distributions under Force. *Phys. Rev. Lett.*, 117(7):078101, August 2016.

[86] M. Yusuf Ali, Hailong Lu, Carol S. Bookwalter, David M. Warshaw, and Kathleen M. Trybus. Myosin v and kinesin act as tethers to enhance each others' processivity. *Proc. Natl. Acad. Sci. U.S.A.*, 105(12):4691–4696, 2008.

[87] Virupakshi Soppina, Stephen R Norris, Aslan S Dizaji, Matt Kortus, Sarah Veatch, Michelle Peckham, and Kristen J Verhey. Dimerization of mammalian kinesin-3 motors results in superprocessive motion. *Proc. Natl. Acad. Sci. U.S.A.*, 111(15):5562–7, April 2014.

[88] Jing Xu, Stephen J King, Maryse Lapierre-landry, and Brian Nemec. Interplay between Velocity and Travel Distance of Kinesin-based Transport in the Presence of Tau. *Biophys. J.*, 105(10):L23–L25, 2013.

[89] Melanie J. I. Müller, Florian Berger, Stefan Klumpp, and Reinhard Lipowsky. *Cargo Transport by Teams of Molecular Motors: Basic Mechanisms for Intracellular Drug Delivery*, pages 289–309. John Wiley & Sons, Inc., 2010.

[90] Ahmet Yildiz, Michio Tomishige, Ronald D Vale, and Paul R Selvin. Kinesin walks hand-over-hand. *Sciences*, 303(5658):676–8, January 2004.

[91] Charles L. Asbury, Adrian N. Fehr, and Steven M. Block. Kinesin moves by an asymmetric hand-over-hand mechanism. *Science*, 302(5653):2130–2134, 2003.

[92] David L. Coy, Michael Wagenbach, and Jonathon Howard. Kinesin takes one 8-nm step for each atp that it hydrolyzes. *J. Biol. Chem.*, 274(6):3667–3671, 1999.

[93] Changbong Hyeon, Stefan Klumpp, and Jose N. Onuchic. Kinesin's backsteps under mechanical load. *Phys. Chem. Chem. Phys.*, 11:4899–4910, 2009.

[94] John Hughes, Shankar Shastry, William O Hancock, and John Fricks. Estimating Velocity for Processive Motor Proteins with Random Detachment. *J. Agr. Biol. Envir. S.*, 18(2):204–217, June 2013.

[95] Koen Visscher, Mark J Schnitzer, and Steven M Block. Single kinesin molecules studied with a molecular force clamp. *Nature*, 400(6740):184–189, July 1999.

[96] A.B. Bortz, M.H. Kalos, and J.L. Lebowitz. A new algorithm for monte carlo simulation of ising spin systems. *J. Comput. Phys.*, 17(1):10 – 18, 1975.

[97] Anatoly B. Kolomeisky and Michael E. Fisher. Periodic sequential kinetic models with jumping branching and deaths. *Physica A*, pages 1–20, 2000.

[98] H. Wang, C. S. Peskin, and T. C. Elston. A robust numerical algorithm for studying biomolecular transport processes. *J. Theor. Biol.*, 221(4):491 – 511, 2003.

[99] A. Valleriani, S. Liepelt, and R. Lipowsky. Dwell time distributions for kinesin's mechanical steps. *EPL (Europhysics Lett.*, 82(2):28011, April 2008.

[100] Denis Tsygankov, Martin Lindén, and Michael E. Fisher. Back-stepping, hidden substeps, and conditional dwell times in molecular motors. *Phys. Rev. E*, 75(2):021909, February 2007.

[101] Martin Lindén and Mats Wallin. Dwell time symmetry in random walks and molecular motors. *Biophys. J.*, 92(11):3804–16, June 2007.

[102] Jurek Kolasa and Steward T. A. Pickett, editors. *Ecological Heterogeneity.* Springer-Verlag New York, 1991.

[103] S. T. A. Pickett and M. L. Cadenasso. Landscape ecology: Spatial heterogeneity in ecological systems. *Science*, 269(5222):331–334, 1995.

[104] William C. Aird. Phenotypic heterogeneity of the endothelium: Ii. representative vascular beds. *Circ. Res.*, 100(2):174–190, 2007.

[105] Gloria H. Heppner. Tumor heterogeneity. *Cancer Res.*, 44(6):2259–2265, 1984.

[106] Rebecca A. Burrell, Nicholas McGranahan, Jiri Bartek, and Charles Swanton. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501(7467):338–345, September 2013.

[107] Edward J. Fox and Lawrence A. Loeb. Cancer: One cell at a time. *Nature*, 512(7513):143–144, August 2014.

[108] Changbong Hyeon, Michael Hinczewski, and D. Thirumalai. Evidence of Disorder in Biological Molecules from Single Molecule Pulling Experiments. *Phys. Rev. Lett.*, 112(13):138101, March 2014.

[109] Michael Hinczewski, Changbong Hyeon, and D. Thirumalai. Directly measuring single-molecule heterogeneity using force spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 113(27):E3852–E3861, July 2016.

[110] C B Anfinsen. Principles that govern the folding of protein chains. *Sci.*, 181(4096):223–30, jul 1973.

[111] Jaya G Yodh, Michael Schlierf, and Taekjip Ha. Insight into helicase mechanism and function revealed through single-molecule approaches. *Q. Rev. Biophys.*, 43(2):185–217, May 2010.

[112] H P Lu, L Xun, and X S Xie. Single-molecule enzymatic dynamics. *Sci.*, 282(5395):1877–82, December 1998.

[113] Changbong Hyeon, Jinwoo Lee, Jeseong Yoon, Sungchul Hohng, and D Thirumalai. Hidden complexity in the isomerization dynamics of Holliday junctions. *Nat. Chem.*, 4(11):907–14, November 2012.

[114] Kyung Suk Lee, Hamza Balci, Haifeng Jia, Timothy M Lohman, and Taekjip Ha. Direct imaging of single UvrD helicase dynamics on long single-stranded DNA. *Nat. Commun.*, 4:1878, May 2013.

[115] Bian Liu, Ronald J Baskin, and Stephen C Kowalczykowski. DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate. *Nature*, 500(7463):482–5, August 2013.

[116] Liisa T Chisty, Christopher P Toseland, Natalia Fili, Gregory I Mashanov, Mark S Dillingham, Justin E Molloy, and Martin R Webb. Monomeric PcrA helicase processively unwinds plasmid lengths of DNA in the presence of the initiator protein RepD. *Nucleic Acids Res.*, 41(9):5010–23, May 2013.

[117] Natali Fili, Gregory I Mashanov, Christopher P Toseland, Christopher Batters, Mark I Wallace, Joseph T P Yeeles, Mark S Dillingham, Martin R Webb, and Justin E Molloy. Visualizing helicases unwinding DNA at the single molecule level. *Nucleic Acids Res.*, 38(13):4448–57, July 2010.

[118] Salman Syed, Manjula Pandey, Smita S Patel, and Taekjip Ha. Single-Molecule Fluorescence Reveals the Unwinding Stepping Mechanism of Replicative Helicase. *Cell Rep.*, 6(6):1037–1045, March 2014.

[119] Christopher J Fischer, Lake Wooten, Eric J Tomko, and Timothy M Lohman. Kinetics of motor protein translocation on single-stranded DNA. *Methods Mol. Biol.*, 587:45–56, January 2010.

[120] Aaron L Lucius, Nasib K Maluf, Christopher J Fischer, and Timothy M Lohman. General Methods for Analysis of Sequential n -step Kinetic Mechanisms : Application to Single Turnover Kinetics of Helicase-Catalyzed DNA Unwinding. *Biophys. J.*, 84(October):2224–2239, 2003.

[121] Christopher J Fischer and Timothy M Lohman. ATP-dependent translocation of proteins along single-stranded DNA: models and methods of analysis of pre-steady state kinetics. *J. Mol. Biol.*, 344(5):1265–86, December 2004.

[122] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover Publications, New York, NY, 0009-revised edition edition, June 1965.

[123] Bo Sun, Daniel S Johnson, Gayatri Patel, Benjamin Y Smith, Manjula Pandey, Smita S Patel, and Michelle D Wang. ATP-induced helicase slippage reveals highly coordinated subunits. *Nature*, 478(7367):132–5, October 2011.

[124] Jung-Chi Liao, Yong-Joo Jeong, Dong-Eun Kim, Smita S Patel, and George Oster. Mechanochemistry of t7 DNA helicase. *J. Mol. Biol.*, 350(3):452–75, July 2005.

[125] Kenji Yoshimoto, Karunesh Arora, and Charles L Brooks. Hexameric Helicase Deconstructed: Interplay of Conformational Changes and Substrate Coupling. *Biophys. J.*, 98(8):1449–1457, April 2010.

[126] T. R. Kirkpatrick and D. Thirumalai. *Colloquium* : Random first order transition theory concepts in biology and physics. *Rev. Mod. Phys.*, 87:183–209, Mar 2015.

[127] Ludovic Berthier and Giulio Biroli. Theoretical perspective on the glass transition and amorphous materials. *Rev. Mod. Phys.*, 83(2):587–645, June 2011.

[128] Joy T. Yang, Robert A. Laymon, and Lawrence S.B. Goldstein. A three-domain structure of kinesin heavy chain revealed by dna sequence and microtubule binding analyses. *Cell*, 56(5):879 – 889, 1989.

[129] Ahmet Yildiz, Michio Tomishige, Ronald D Vale, and Paul R Selvin. Kinesin Walks Hand-Over-Hand. *Science*, 303(5658):676–678, January 2004.

[130] Zhechun Zhang and D Thirumalai. Dissecting the kinematics of the kinesin step. *Structure*, 20(4):628–40, April 2012.

[131] Hiroshi Isojima, Ryota Iino, Yamato Niitani, Hiroyuki Noji, and Michio Tomishige. Direct observation of intermediate states during the stepping motion of kinesin-1. *Nat. Chem. Biol.*, 12(4):290–297, April 2016.

[132] Susan P. Gilbert, Martin R. Webb, Martin Brune, and Kenneth A. Johnson. Pathway of processive ATP hydrolysis by kinesin. *Nature*, 373(6516):671–676, February 1995.

[133] Wonmuk Hwang, Matthew J Lang, and Martin Karplus. Force generation in kinesin hinges on cover-neck bundle formation. *Structure*, 16(1):62–71, January 2008.

[134] S Rice, Y Cui, C Sindelar, N Naber, M Matuska, R Vale, and R Cooke. Thermodynamic properties of the kinesin neck-region docking to the catalytic core. *Biophys. J.*, 84(3):1844–54, March 2003.

[135] Ahmad S Khalil, David C Appleyard, Anna K Labno, Adrien Georges, Martin Karplus, Angela M Belcher, Wonmuk Hwang, and Matthew J Lang. Kinesin's cover-neck bundle folds forward to generate force. *Proc. Natl. Acad. Sci. U.S.A.*, 105(49):19247–52, December 2008.

[136] William R. Hesse, Miriam Steiner, Matthew L. Wohlever, Roger D. Kamm, Wonmuk Hwang, and Matthew J. Lang. Modular Aspects of Kinesin Force Generation Machinery. *Biophys. J.*, 104(9):1969–1978, May 2013.

[137] Aliaksei Krukau, Volker Knecht, and Reinhard Lipowsky. Allosteric control of kinesin's motor domain by tubulin: a molecular dynamics study. *Phys. Chem. Chem. Phys.*, 16(13):6189–6198, April 2014.

[138] Yu Cheng Zhao, F. Jon Kull, and Jared C. Cochran. Modulation of the kinesin ATPase cycle by neck linker docking and microtubule binding. *J. Biol. Chem.*, 285(33):25213–25220, August 2010.

[139] Bojan Milic, Johan O. L. Andreasson, William O. Hancock, and Steven M. Block. Kinesin processivity is gated by phosphate release. *Proc. Natl. Acad. Sci. U.S.A.*, 111(39):14136–14140, September 2014.

[140] Johan O. L. Andreasson, Bojan Milic, Geng-Yuan Chen, Nicholas R. Guydosh, William O. Hancock, and Steven M. Block. Examining kinesin processivity within a general gating framework. *Elife*, 4, 2015.

[141] Changbong Hyeon and José N. Onuchic. Mechanical control of the directional stepping dynamics of the kinesin motor. *Proc. Natl. Acad. Sci. U.S.A.*, 104(44):17382–17387, October 2007.

[142] R. D. Vale and R. A. Milligan. The way things move: looking under the hood of molecular motor proteins. *Science*, 288(5463):88–95, April 2000.

[143] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26(16):1781–1802, 2005.

[144] MacKerell A. D., D Bashford, Bellott, Dunbrack R. L., J D Evanseck, M J Field, S Fischer, J Gao, H Guo, S Ha, D Joseph-McCarthy, L Kuchnir, K Kuczera, F T K Lau, C Mattos, S Michnick, T Ngo, D T Nguyen, B Prodhom, W E Reiher, B Roux, M Schlenkrich, J C Smith, R Stote, J Straub, M Watanabe, J Wiórkiewicz-Kuczera, D Yin, and M Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616, April 1998.

[145] J Lowe, H Li, K.H Downing, and E. Nogales. Refined structure of alpha beta-tubulin at 3.5 A resolution. *J. Mol. Biol.*, 313:1045–1057, 2001.

[146] Charles V. Sindelar and Kenneth H. Downing. The beginning of kinesin's force-generating cycle visualized at 9- resolution. *J. Cell Biol.*, 177(3):377–385, May 2007.

[147] Zhechun Zhang, Riina Tehver, and D. Thirumalai. Microtubule assists in kinesin's adp release and diffusional search. *unpublished*.

[148] John Eargle, Dan Wright, and Zaida Luthey-Schulten. Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics*, 22(4):504–506, Feb 2006.

[149] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

[150] John Stone. *An Efficient Library for Parallel Ray Tracing and Animation.* Master's thesis, Computer Science Department, University of Missouri-Rolla, April 1998.

[151] Arun Kumar Somavarapu and Kasper P. Kepp. The Dependence of Amyloid-β Dynamics on Protein Force Fields and Water Models. *Chemphyschem*, 16(15):3278–3289, October 2015.

[152] Nikolina Kalchishkova and Konrad J. Bhm. The Role of Kinesin Neck Linker and Neck in Velocity Regulation. *J. Mol. Biol.*, 382(1):127–135, September 2008.

[153] Ryo Nitta, Yasushi Okada, and Nobutaka Hirokawa. Structural model for strain-dependent microtubule activation of Mg-ADP release from kinesin. *Nat. Struct. Mol. Biol.*, 15(10):1067–1075, October 2008.

[154] Changbong Hyeon and José N Onuchic. Internal strain regulates the nucleotide binding site of the kinesin leading head. *Proc. Natl. Acad. Sci. U.S.A.*, 104(7):2175–2180, February 2007.

[155] Venkatesh Hariharan and William O Hancock. Insights into the Mechanical Properties of the Kinesin Neck Linker Domain from Sequence Analysis and Molecular Dynamics Simulations. *Cell. Mol. Bioeng.*, 2(2):177–189, June 2009.

[156] Ahmet Yildiz, Michio Tomishige, Arne Gennerich, and Ronald D Vale. Intramolecular strain coordinates kinesin stepping behavior along microtubules. *Cell*, 134(6):1030–41, September 2008.

[157] Robert A. Cross and Andrew McAinsh. Prime movers: the mechanochemistry of mitotic kinesins. *Nat. Rev. Mol. Cell Biol.*, 15(4):257–271, April 2014.

[158] Lukas C. Kapitein, Erwin J. G. Peterman, Benjamin H. Kwok, Jeffrey H. Kim, Tarun M. Kapoor, and Christoph F. Schmidt. The bipolar mitotic kinesin Eg5 moves on both microtubules that it crosslinks. *Nature*, 435(7038):114–118, May 2005.

[159] Youwei Yan, Vinod Sardana, Bei Xu, Carl Homnick, Wasyl Halczenko, Carolyn A. Buser, Michael Schaber, George D. Hartman, Hans E. Huber, and Lawrence C. Kuo. Inhibition of a Mitotic Motor Protein: Where, How, and Conformational Consequences. *J. Mol. Biol.*, 335(2):547–554, January 2004.

[160] Hung Yi Kristal Kaan, Venkatasubramanian Ulaganathan, David D. Hackney, and Frank Kozielski. An allosteric transition trapped in an intermediate state of a new kinesininhibitor complex. *Biochem. J.*, 425(1):55–61, January 2010.

[161] F. Jon Kull and Sharyn A. Endow. Force generation by kinesin and myosin cytoskeletal motor proteins. *J. Cell. Sci.*, 126(Pt 1):9–19, January 2013.

[162] Jun Xing, Willy Wriggers, Geraldine M. Jefferson, Richard Stein, Herbert C. Cheung, and Steven S. Rosenfeld. Kinesin Has Three Nucleotide-dependent Conformations IMPLICATIONS FOR STRAIN-DEPENDENT RELEASE. *J. Biol. Chem.*, 275(45):35413–35423, November 2000.

[163] Joost van Mameren, Karen C Vermeulen, Fred Gittes, and Christoph F Schmidt. Leveraging single protein polymers to measure flexural rigidity. *J. Phys. Chem. B*, 113(12):3837–44, March 2009.

[164] Victor A. Bloomfield, Donald M. Crothers, and Ignacio Tinoco. *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, 2000.

[165] Yariv Kafri, David K. Lubensky, and David R. Nelson. Dynamics of molecular motors with finite processivity on heterogeneous tracks. *Phys. Rev. E*, 71(4):041906, April 2005.

[166] Eric A. Toth, Ying Li, Michael R. Sawaya, Yifan Cheng, and Tom Ellenberger. The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Mol. Cell*, 12(5):1113–1123, November 2003.

[167] Eileen K. Jaffe. Impact of Quaternary Structure Dynamics on Allosteric Drug Discovery. *Curr Top Med Chem*, 13(1):55–63, January 2013.

[168] ArthurF. Voter. Introduction to the kinetic monte carlo method. In KurtE. Sickafus, EugeneA. Kotomin, and BlasP. Uberuaga, editors, *Radiation Effects in Solids*, volume 235 of *NATO Science Series*, pages 1–23. Springer Netherlands, 2007.

[169] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.

[170] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15:1–89, Jan 1943.

[171] William O. Hancock and Jonathon Howard. Kinesins processivity results from mechanical and chemical coordination between the atp hydrolysis cycles of the two motor domains. *Proc. Natl. Acad. Sci. U.S.A.*, 96(23):13147–13152, 1999.

[172] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot. Molecular stick-slip motion revealed by opening dna with piconewton forces. *Phys. Rev. Lett.*, 79:4489–4492, Dec 1997.

[173] C. C. Liu and B M Alberts. Characterization of the DNA-dependent GTPase activity of T4 gene 41 protein, an essential component of the T4 bacteriophage DNA replication apparatus. *J. Biol. Chem.*, 256(6):2813–20, March 1981.

[174] Robert D. Schrock and Bruce Alberts. Processivity of the gene 41 dna helicase at the bacteriophage t4 dna replication fork. *J. Biol. Chem.*, 271(28):16678–16682, 1996.

[175] T. W. Hegarty. Temperature Coefficient (Q10), Seed Germination and Other Biological Processes. *Nature*, 243(5405):305–306, June 1973.

[176] Dong-Eun Kim, Murli Narayan, and Smita S Patel. {T7} {DNA} helicase: A molecular motor that processively and unidirectionally translocates along single-stranded {DNA}. *J. Mol. Biol.*, 321(5):807 – 819, 2002.