Theoretical perspectives on biological machines

Mauro L. Mugnai, Changbong Hyeon, Michael Hinczewski, and D. Thirumalai

Theoretical Perspectives on Biological Machines

Mauro L. Mugnai*

*Department of Chemistry,
University of Texas, Austin,
TX 78712*

Changbong Hyeon†

*Korea Institute for Advanced Study, Seoul 02455,
Republic of Korea*

Michael Hinczewski‡

*Department of Physics,
Case Western Reserve University,
OH 44106*

D. Thirumalai§

*Department of Chemistry,
University of Texas, Austin,
TX 78712*

Many biological functions are executed by molecular machines, which like man made motors are powered by chemical energy, released by ATP hydrolysis. They have evolved to transport cargo, facilitate folding proteins and RNAs to fold, remodel chromatin, and replicate DNA. The biological machines also serve as the most natural systems for illustrating emergent phenomena in non-equilibrium active systems, thus providing a great impetus to illustrate the general principles governing their functions. Despite their architectural diversity, physics-based theories have provided unifying themes of the inner working of nanoscale biological machines. The theories address questions such as, how the trade-off between precision, energetic costs and optimal performances, are balanced. However, many complexities associated with biological machines require molecular level descriptions. Simple point mutations in the enzyme could drastically alter functions, result in unexpected diseases or dramatically restrict the capacity of molecular chaperones to help proteins fold. These examples are reminders that, while the search for principles of generality in biology is intellectually stimulating and necessary, molecular details must be accounted in order to develop a deeper understanding of their functions.

CONTENTS

* maurolm83@gmail.com
† hyeoncb@kias.re.kr
‡ mhincz@gmail.com
§ dave.thirumalai@gmail.com

I. INTRODUCTION

The opening sentence of a perspective by Bustamante (Bustamante *et al.*, 2011) on the workings of nucleic acid translocases begins with the quote, "The operative industry of Nature is so prolific that machines will be eventually found not only unknown to us but also unimaginable by our mind", attributed to Marcello Malpighi, who is considered the founder of microscopic anatomy, histology, and embryology. This statement, made over three centuries ago, is even more relevant today. It is a reminder that mechanical forces must play a fundamental role in biology. In modern times this vast subject falls under the growing field of mechanobiology. The pervasive role of mechanics in living systems controls motility on all length scales, from motion of a single cell on a substrate and collections of cells to dynamics at the molecular level. Cooperative interactions between various modules at the molecular level is thought to control functions at the mesoscale. A number of complex dynamical processes such as transcription, translation, transport of vesicles and organelles, folding of proteins and RNA, and chromosome segregation control the sustenance and growth of cells. At some level all these biologically important processes involve molecular machines, whose ability to perform their functions, often but always with high efficiency, in a noisy crowded environment is truly remarkable. It is worth remembering that the ability to execute a variety of functions distinguishes living and abiotic systems. Because of functional demands in living systems, which also includes adaptation to changing environmental conditions, it is virtually impossible to fully describe biology without evolutionary considerations. Although not the focus of our perspective, the role of evolutionary constraints must also be integrated with physical models in order to discover general principles governing the functions of biological machines.

What are the characteristics of biological machines? First, there are many varieties of machines, all of which should be thought of as enzymes, which consume energy and perform mechanical work to carry out specific tasks. A few of the machines that we consider here are kinesin, myosin, and dynein, which are collectively referred to as molecular motors (Belyy *et al.*, 2014a; Block, 2007; Reck-Peterson *et al.*, 2018; Sun and Goldman, 2011a; Vale
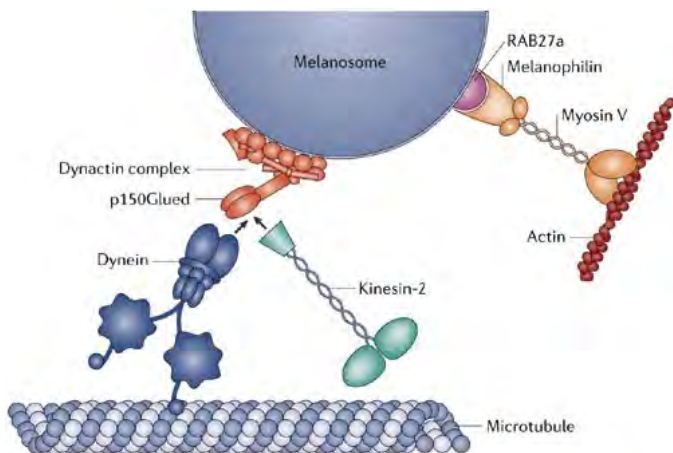

FIG. 1 Illustration of the complexity of transportation of melanosomes, which are vesicles containing melanin. Both dynein and kinesin-2 compete for the same binding site on dynactin mediated by p150Glued. Depending on the function (aggregation of melanosomes or their dispersion throughout the cell) one or the other wins. When melanosomes are dispersed in the cell they are transported by kinesin-2 and myosin V whereas when they aggregate dynein moves the cargo. Figure extracted from (Soldati and Schliwa, 2006).

and Milligan, 2000). These cytoskeletal motors transport cargo by walking, almost always unidirectionally, on filamentous actin and microtubules (MT). Under *in vivo* conditions the motors cooperate or there could be a tug-of-war in the process of transport (see Fig. 1) (Gross *et al.*, 2002; Levi *et al.*, 2006). To illustrate the diversity of motor-like functions and point out certain emerging general principles, we also provide theoretical descriptions for the functions of molecular chaperones, which assist in the folding of proteins and ribozymes (RNA enzymes) that cannot do so spontaneously, and helicases with multiple functions that includes separation of double stranded (ds) DNA strands. Both molecular chaperones and helicases bear no structural resemblance or sequence similarity to molecular motors. Nevertheless, we are convinced that by comparing the characteristics of these seemingly unrelated machines, integrating theory and experiments, unifying themes and differences between them could be elucidated. Second, all these motors and others not covered here (for example polymerases, ribosomes, and packaging motors) are all multidomain proteins, whose architectures are spectacularly different. Despite the vastly different sequences, structures, and evolutionary origins it is indeed the case that all of these machines utilize some form of chemical energy (generated by ATP or GTP hydrolysis) in order to amplify the small local conformational changes through their structural linkages to facilitate large conformational changes for functional purposes. Such conformational amplifications are examples of remarkable allostery or action at a distance, and could also be couched in terms of in-

formation transfer between structural subunits that are spatially well separated, which in some cases (dynein for example (Bhabha *et al.*, 2014)) can be as large as 25nm.

A large number of experiments have unveiled many of the details of how these machines move by converting chemical energy into mechanical work (see for example (Hartman *et al.*, 2011; Spudich and Sivaramakrishnan, 2010) (De La Cruz and Ostap, 2004; Holzbaur and Goldman, 2010; Sweeney and Houdusse, 2010). A remarkable number of experimental methods have been developed to address various aspects of biological machines. These include, but are not restricted to, ensemble experiments (stopped flow and fluorescent labeling) that provide the much needed data on ATP hydrolysis and ADP release rates, single molecule experiments that yield dwell time distributions in molecular motors, processivity and velocity as a function of external loads in motors and helicases. In addition, a combination of ensemble experiments and determination of structures using X-ray crystallography and cryo-EM experiments has produced a vivid picture of the molecular basis of chaperone function. Of particular note are optical trap experiments, which have been used to obtain mean motor velocity as a function of a resistive force and ATP concentration in kinesin, myosin, and dynein, and the dependence of velocity and processivity in a number of helicases. No point would be served in reviewing the detailed results from these experiments as there are many articles that the interested reader might consult.

Our focus here is to describe theoretical approaches that are rooted in a number of areas in physics in order to understand principally outcomes of *in vitro* experiments. The theoretical approaches in these studies were developed, and continue to be the focus of current research, in order to quantitatively explain the experimental observations, and shed light on new puzzles that seem to arise with ever improving advances in experimental techniques. Perhaps, the most detailed view of how biological machines operate might be obtained from molecular dynamics simulations. Although atomically detailed molecular dynamics simulations have been performed to get a molecular picture of certain aspects of the functions of motors (Hwang *et al.*, 2008, 2017) and other machines (see for example (Elber and West, 2010; Ma *et al.*, 2000; Stan *et al.*, 2005), the current limitations of such approaches prevent them from making direct contact with experiments. Improvements in the development of accurate energy functions (also known as force fields) and enhancement in computer capacity to enable simulations for long times will in the future bridge the gap between what is currently possible and what is needed for realistic description of molecular machines.

The actions of molecular machines, like all biological processes, are complicated, involving cooperative dynamics on multiple time scales. This is illustrated using the typical time scales involved in a single step of myosin V

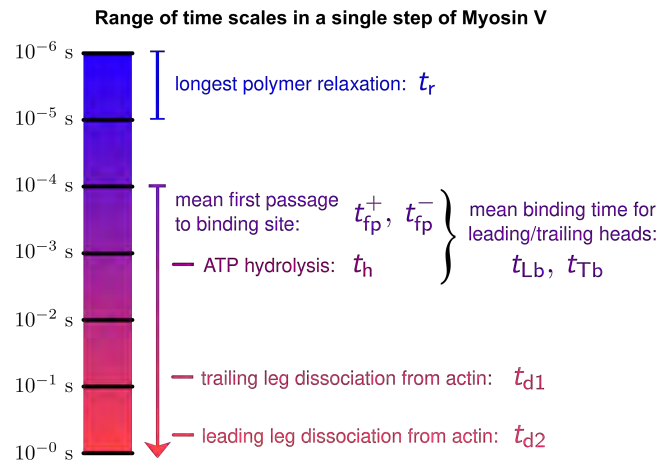**Range of time scales in a single step of Myosin V**

FIG. 2 Significant physical timescales associated with myosin V dynamics, from the coarse-grained polymer theory model of (Hinczewski *et al.*, 2013).

(Fig. 2). In many problems in physics it is the hoped that simple models, that capture the essence of the problem can be created, which be can be solved, in order to obtain insights into highly non-trivial systems. This strategy is hard to implement for biological machines because the interaction energies are highly heterogeneous (like in spin glasses), thus making it hard to construct a reasonable coarse-graining strategy. Nevertheless, in order to make progress one has to devise tractable coarse-grained models, which can be either simulated or solved analytically (at least approximately). The efficacy of such approaches can then be assessed by direct comparisons with experimental results, and their abilities to provide insights into the mechanisms of their functions. Remarkably, inspired by advances in experiments in the last decade, several theoretical models have been proposed, which have greatly contributed to our understanding of molecular machines. These developments have occurred in different contexts, which belie the underlying unifying principles. To bring these issues to the fore, we provide our perspectives on the application of these models, which should be thought of as coarse-grained network models or their generalizations. Such models have been created for explaining not only the motility of motors, but also the functions of molecular chaperones, and helicases. We focus on the applications of these theoretical ideas in order to account for the functions of these intrinsically non-equilibrium systems. Collectively, these approaches show that, by examining in detail the workings of many machines, universal principles, both at the conceptual and practical levels, might emerge.

The literature on the functions of biological machines is vast. Therefore, we restricted ourselves to only a few topics that focus on theoretical approaches, which are sufficiently general that they can be applied to an array of problems in the field. Rather than describe many

results in detail, we walk the the reader through a selection of theoretical methods, which is necessarily biased, so that she or he can access the literature readily. Before getting to the details of this review we will be remiss if we did not point out one prescient monograph (Howard, 2001) and two forward looking reviews (Jülicher *et al.*, 1997; Kolomeisky and Fisher, 2007). The monograph by Howard, written nearly twenty years ago, covers all aspects of cytoskeletal motors and is a landmark in this field. It provides conceptual and practical guidelines needed to understand the fundamentals of motor mechanics. Questions of generality, such as describing movement of a generic motor either in isolation or as a collection, were addressed using principally the Brownian ratchet model by Julicher, Adjari, and Prost (Jülicher *et al.*, 1997), whereas Kolomeisky and Fisher (Kolomeisky and Fisher, 2007) summarized the development and practical applications of stochastic kinetic models. Here, our focus is on the more recent developments and applications of physical principles that have started to provide a unified perspective for not only molecular motors but a large of class of machines with vastly different functions. These new insights have become possible by expanding the scope of traditional stochastic chemical kinetics models for motors, helicases, and molecular chaperones, incorporation of polymer physics concepts to account for the architecture of motors, and use of coarse-grained models in simulating the stepping kinetics of motors and the dynamics of large scale allosteric transitions.
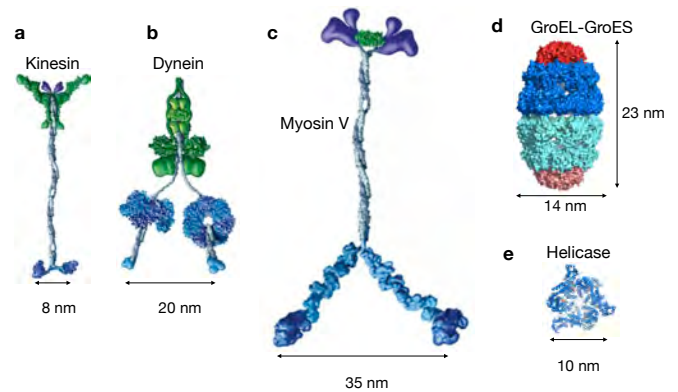

FIG. 3 Schematic representations of the structures of six biological machines. Molecular motors, like conventional kinesin (a), dynein (b), and myosin V (c), show great variations in their structures and sizes. The motor heads in kinesin and myosin V, shown in dark blue, bind directly to the microtubule and actin, respectively (not shown) but in dynein the microtubule binding domain (light blue) is separated from the motor domain (hexameric ring structures) by nearly twenty five nanometers. (d) Structure of the GroEL in the symmetric state (Protein Data Bank (PDB) code 4PKO (Fei *et al.*, 2014)), which is the functional state in the presence of misfolded proteins. The blue and aqua colors represent the seven fold symmetric GroEL rings. The red and the pink correspond to the co-chaperonin GroES. (e) Structure of the DnaB helicase with six subunits assembled as a ring. The figure was created using the coordinates in the PDB code 4ESV (Itsathitphaisarn *et al.*, 2012). Structures (a)-(c) are adapted from (Vale, 2003a).

II. STRUCTURES

The structures of the biological machines that we consider in our perspective (kinesin, myosin, dynein, the *E. Coli.* chaperonin machinery GroEL/GroES, and helicases) are shown in Fig. 3. Inspection of the figures shows that there are considerable variations in the architectures, although kinesin-1 (or conventional kinesin) and myosin V share some structural similarities.

A. Molecular motors

Let us first consider the three motors whose structures are schematically shown in Fig. 3 (a-c). Although there are differences between them, the parts list is roughly the same. These motors are dimeric. The nucleotide binding sites are in the two motor heads, which are connected through a mobile linker (the lever arm in myosin) to a tail domain involved in dimerization and cargo-binding. Changes due to nucleotide binding and hydrolysis in the motor heads result in conformational changes in the linker that propel the motor on the cytoskeletal filaments. For the motors to be processive, which

means they take multiple steps before disengaging from the track, one head has to be bound till the detached head rebinds to a site on the track. This involves communication between the heads and is referred to as gating, the origin of which is still not fully understood at the molecular level.

Kinesin: There are at least forty five members belonging to the kinesin superfamily in mouse and human genomes (Hirokawa *et al.*, 2009). They are all microtubule (MT) bound motors, which walk unidirectionally towards the plus end of the MT (for example kinesin-1) or to the minus end (Ncd motor). Kinesin-1 takes precisely 8.2 nm steps, which corresponds to the distance between two adjacent α/β tubulin dimers, which are the building blocks of the MT. The size of kinesin-1 is a few nanometers whereas the length of the stalk is in the range of 30-40 nm. The kinesin-1 velocity depends on the concentration of ATP, saturating at high values. The motor moves towards the plus end of the MT with maximal velocity of approximately 800 nm/s (Visscher *et al.*, 1999), and is capable of resisting forces on the order of 7 pN (Carter and Cross, 2005; Visscher *et al.*, 1999) ($k_B T = 4.1$ pN·nm where k_B is the Boltzmann constant and T is the temperature).

Myosin: The number of genes encoding for myosin motors in roughly thirty five (Sellers, 2000). The superfamily of the actin bound myosins are divided into fifteen classes (Hartman and Spudich, 2012; Sellers, 2000). With the exception of myosin VI all other members of this family walk towards the plus end of actin. The structure of myosin V (Fig. 3) shows that the motor heads are connected to the lever arms, which are made up of six IQ motifs. The lever arm is $\approx$ 36 nm-long stiff unit (the persistence length exceeds 100 nm), whose size is commensurate with half the helical repeat length of F-actin. The maximal velocity of myosin V, whose motor head is larger than kinesin-1, is roughly 500 nm/s (Baker *et al.*, 2004) with a stall force between 2-3 pN (Kad *et al.*, 2008; Mehta *et al.*, 1999; Uemura *et al.*, 2004a; Veigel *et al.*, 2002).

Dynein: Cytoplasmic dynein, discovered over fifty years ago (Gibbons and Rowe, 1965) and pictured in Fig. 3b, walks somewhat erratically with a broad step-size distribution (DeWitt *et al.*, 2012; Reck-Peterson *et al.*, 2006a) on the MT towards the minus end. The structural features of dynein are different when compared with other cytoskeletal motors (see Fig. 3). First, the motor head belongs to the class of AAA+ family, which means dynein must have evolved from a different lineage compared with myosin and kinesin. Second, other AAA+ enzymes, such as bacterial chaperonin GroEL (Fig.3d) and protein degradation machines, are oligomeric assemblies. In contrast, the hexameric ring that constitutes the motor domain assembles from a single polypeptide chain! Third, the size of the dynein motor head is significantly larger than those of kinesin and myosin. The length of the motor head of dynein along its longest axis is about 25 nm, which is in contrast to the diameter of kinesin, which is only about 5 nm. Finally, although there are six nucleotide binding sites in dynein, hydrolysis in only two (perhaps three) are relevant for its motility. The velocity of dynein at 1 mM ATP is roughly ∼100 nm/s (Reck-Peterson *et al.*, 2006b) with a stall force that is $\approx$ 7pN (Gennerich *et al.*, 2007; Toba *et al.*, 2006).

Chaperonins: The beautiful and unusual structure with seven fold symmetry of the *E. Coli.* chaperonin machinery, consisting of a complex between GroEL and GroES, resembles an American football (Fig. 3d). The GroEL/GroES machine helps in the folding of recalcitrant proteins that do not fold spontaneously. The structure in Fig. 3d (reported in (Fei *et al.*, 2014)) is the functional state of this stochastic machine that is one of the populated states during the catalytic cycle of GroEL in the presence of misfolded substrate proteins (SPs). There are two chambers in which the SPs could be sequestered. The volume of each of the chambers in the structure in Fig. 3e is about 185,000 Å^3, which is about twice the volume in the relaxed structure in the absence of nucleotides and GroES, the co-chaperonin. The spectacular change in the chamber volume that occurs during the catalytic cycle, with accompanying alterations in the chemical nature of the cavity interior, changing from hydrophobic to polar, is the mechanism of annealing action of this machine (Todd *et al.*, 1996).

Helicases: Helicases, which are involved in all aspects of nucleic acid metabolism (Lohman, 1992; Lohman and Bjornson, 1996), function by coupling nucleoside triphosphate (NTP) hydrolysis, to either translocate on single strand nucleic acids (ssNAs) or unwind double-stranded (ds) DNA. Depending on their sequences they are classified into six super families (SFs). The structural diversity of helicases can be appreciated by noticing that sequences classified under the SF1 and SF2 families are non-ring forming whereas those in SF3-SF6 form ring structures. The hexametric structure (Itsathitphaisarn *et al.*, 2012) of the replicative helices (DnaB) from bacteria, which does belong to the AAA+ family, is shown in Fig. 3d. However, unlike GroEL (see Fig. 3d) DnaB has the expected crystallographically allowed six fold symmetry. How the NTP chemistry is coupled to translocation and unwinding remains an outstanding unsolved theoretical problem.

B. Catalytic Cycle

All machines undergo a catalytic cycle, not unlike man-made motors, in which fuel, usually in the form of ATP, bound to a nucleotide binding site (or sites) is hydrolyzed. These events trigger conformational changes that produce motion, which in motors and helicases results in stepping on the polar tracks or translocation on single stranded nucleic acids. In chaperones the nucleotide chemistry is linked to conformational changes, which in turn perform work on the protein or ribozyme to be folded. The link between ATPase cycle and function is somewhat different in GroEL, which we describe below. The catalytic cycle for myosin V in the simplest form, which suffices for our purposes, is reproduced in Fig. 4. There are four crucial steps (see Fig. 4). In the initial state, ATP binds to the trailing head (TH), with ADP in the (LH). The premature release of ADP from the LH is slowed by rearward tension, which is an example of gating. Upon ATP binding to the TH the interaction with F-actin is weakened, resulting in its detachment from F-actin. During the diffusive search by the TH for the forward binding site, ATP is hydrolyzed, producing ADP and the inorganic phosphate, P_i. In this state, the TH binds to F-actin after which P_i is released from the new LH followed by ADP release from the TH, and the cycle continues till the processive run ends. Of course, this simple description is incomplete because the rates for nucleotide binding vary depending on the nucleotide concentration. Nevertheless, this simple reaction cycle, whose main features hold for all myosin motors, is sufficient to nearly quantitatively characterize many ex-
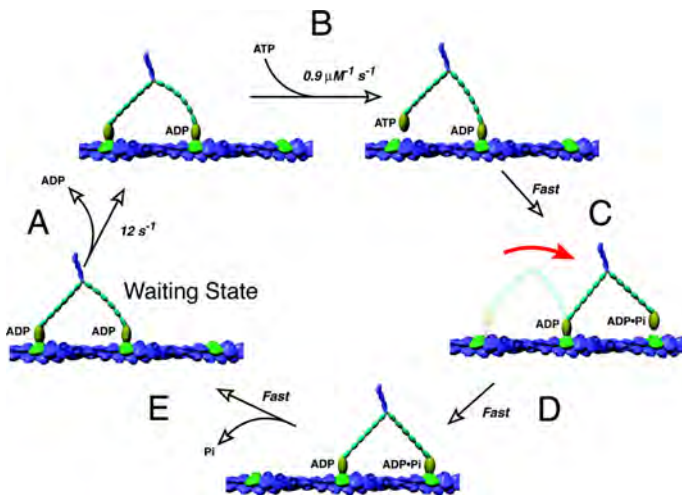
FIG. 4 A simple representation of the catalytic cycle of myosin V describing the stepping of the trailing head towards the plus end of actin, which is the dominant pathway in the absence of resisting force. The main text describes the details. As described later in this perspective there are four other pathways that have to be accounted for in order to produce a quantitative theory of the stepping kinetics of this motor. The figure is reproduced from (Vale, 2003b)

perimental observables (see below). The catalytic cycle for kinesin is similar except that the interaction between the MT and kinesin is weakest if the motor head contains ADP. All other nucleotide states (the no nucleotide *apo*, ATP bound state, the state with ADP and P_i) bind strongly to the MT.

III. BIOLOGICAL MACHINES ARE ACTIVE SYSTEMS.

Biological machines are non-equilibrium systems that are driven by non-conservative forces require constant supply of energy. Therefore, it is not surprising that detailed balance (DB) relation and the fluctuation-dissipation theorem (FDT) are violated (Battle *et al.*, 2016; Gladrow *et al.*, 2016). These are the key features characterizing the out-of-equilibrium nature of biological machines driven by non-conservative forces. Consequently, molecular machines can be thought of as active systems, and hence a note on diffusive motion is warranted. Chemical free energy released upon hydrolysis of ATP or GTP is the driving force for directed motion of molecular motors. However, if the energy source is not explicitly modeled in the description of the associated dynamics, molecular motors could be regarded as self-propelled active particles. To our knowledge such an approach has not been pursued to calculate experimentally measurable quantities, such as the force-velocity relation or the distribution of run lengths of motors.

The dynamics of active particles is fundamentally different from passive particles at equilibrium or moving under the influence of a conservative external field. For example, the mean drift velocity of a spherical colloidal particle with a drag coefficient γ, and charge q subject to a constant electric field E is $V_D = qE/\gamma$. This relation is readily obtained by balancing the Stokes force (γV_D) and the force exerted by the field (qE). The diffusion constant is $D = k_B T/\gamma$, which is the Stokes-Einstein relation. Hence, the distribution of the particle position is given by the probability density,

$$P(x,t) \sim \exp\left[-\frac{(x - x_0 - V_D t)^2}{4Dt}\right]. \qquad (1)$$

Starting from an initial position x_0, the particle moves on average at velocity V_D and the distribution of positions spreads with time as $\langle(\delta x)^2\rangle = 2(k_B T/\gamma)t$. Of particular note is that D is defined independently of the particle velocity V_D, so that the magnitude of D is not altered by increasing the field strength. Conversely, the velocity of the particle does not depend on the ambient temperature T. Thus, V_D and D are mutually independent of one another.

In stark contrast, for an active particle, whose motion is powered by the *internal* fuel, the diffusivity is no longer independent from the driving velocity. The *effective diffusion constant*, D_{eff}, defined as $\lim_{t\to\infty}\langle(\delta x)^2\rangle/2t$, depends on a set of non-thermal parameters and violates the FDT (Liu *et al.*, 2011; Tailleur and Cates, 2008). For transport motors exhibiting one-dimensional movement along the cytoskeletal filament, both $V = \lim_{t\to\infty} d\langle x(t)\rangle/dt$ and $D = (1/2)\lim_{t\to\infty} d\langle\delta x(t)^2\rangle/dt$ can be measured directly using a given time trace of a motor, which can be measured using single molecule optical tweezer experiments. Indeed, such measurements on kinesin-1 (Visscher *et al.*, 1999) were used to obtain the velocity (V) and diffusion coefficient (D) from the global analysis of the stepping trajectories. These values were used (Visscher *et al.*, 1999) to estimate the randomness parameter $r = 2D/d_0 V$ where $d_0 (\approx 8.2$ nm) is the step size of kinesin-1. When the diffusion constant D is calculated from the randomness parameter with the knowledge of V and d_0, and D is compared with V measured under the same conditions, it can be shown that D increases monotonically with V (Hwang and Hyeon, 2017) in contrast to passive diffusion. This result is a consequence of the enzyme catalytic turnover (Hwang and Hyeon, 2017). As already discussed the effective diffusivity D for kinesin-1, which is an active particle whose dynamics is fueled by ATP hydrolysis free energy, does not obey the FDT. This illustration shows that these energy consuming enzymes operate out of equilibrium.

IV. STOCHASTIC KINETIC MODELS

In this section we discuss the development of stochastic kinetic models (SKMs) as a means of describing the

function of different types of molecular machines. As stated above and shown explicitly in Fig. 4 for myosin V, all molecular machines go through cycles during which they hydrolyze ATP (or GTP) and perform a some function. Bulk-kinetics, single-molecule experiments, and structural studies have shown that the intermediates explored by the molecular machines during the cycle have distinguishable structural or kinetic features and have provided the rate for going from one state to the next. The temporal resolution in these studies is on the order of milliseconds [although new techniques enable the exploration of time-scales as fast as $\approx 55\,\mu s$ (Isojima *et al.*, 2016)]. It follows that the dynamics of the molecular machine can be described as a discrete-state (the distinguishable intermediates should be experimentally observed), continuous-time Markov model – a framework that falls under the rubric of the "master equation." In this model, the cycle of a molecular machine is reduced to a network of states connected by edges that identify the available transitions between the intermediates. Thus, by solving the dynamics associated with an appropriate network it is possible, in principle, to account for experimental observables in terms of the underlying network parameters.

SKMs have a number of appealing features: (i) they have a direct connection with the biochemistry associated with the molecular machines; (ii) the rates that constitute the model are measurable; (iii) it is sometimes possible to find an analytical solution, and (iv) more complex networks may be solved numerically; (v) finally, SKMs can be directly related with thermodynamics, which makes the model instructive and appeals to the interests, and intuition of a vast scientific community.

Nevertheless, SKMs also have some shortcomings: first and foremost, they often have a large number of parameters. For instance, a simple model for a molecular machine performing mechanical work against a fixed load and described by a single-cycle network with N intermediate states has $4N - 2$ independent parameters, which have to be determined by fitting to appropriate experiments. However, typically the number of observables are very few, making matching the predictions of the SKM to experiments is difficult. Furthermore, it is often difficult to interpret the physical meaning of the extracted parameters in terms of the underlying motor architecture and the underlying biochemical cycle. As a consequence, it is necessary to strike a balance between the "minimality" of a model, which reduces the risk of over-fitting and increases the generality of SKMs and the predictive power, and the "comprehensiveness" of the network considered. This comes at some risk; for instance, one may be tempted to neglect certain transitions that are "fast" compared to others, and therefore are not expected to contribute to the overall phenomenology of the machine. In addition, it may seem reasonable to ignore off-pathway, slow and rare transitions. On the other hand, in

order to understand the function of molecular machines, experimentalists often probe their response to changes in the environment (for instance modifying the ATP concentration), or by applying mechanical perturbations and studying the response of the motor. As the environment or perturbations change, the nature of the "rate-limiting" step as well as the likelihood of alternative stepping pathways might change. Therefore, a simplified description that ignores certain intermediate states might only work for a limited set of experimental conditions, thereby reducing the number of measurements that can be used to train the parameters or to falsify the predictions. Finally, we note *en passant* that the determination of an appropriate catalytic cycle is predicated on a few experimental observables only, and it may not be complete enough for all the states that a machine might sample during its function.

Another issue with SKMs is that structural information is incorporated in the model only in a very approximate way, normally by introducing a parameter that allows the rate to depend on the load applied according to the Bell model (Bell, 1978). This reduces the capability of the model to incorporate the wealth of experimental structural information, and decreases the predictive power of the proposed paradigm. However, in some cases a tractable analytical theory that incorporates lever arm structural information into a kinetic model is possible (Hinczewski *et al.*, 2013), as discussed in more detail below. Despite these limitations, the framework underlying SKMs not only provides a convenient way to analyze experiments, but also has been used to address conceptual questions related to the efficiency of biological machines. For these reasons, the theory underlying SKMs is an integral part of describing the function of molecular motors.

A. The Master Equation

We assume that the system is described by N distinguishable intermediate states. The probability of being in state i at time t is given by $p_i(t)$, and the time evolution of this probability is governed by the master equation,

$$\frac{dp_i(t)}{dt} = \sum_j p_j(t)w_{ji} - \sum_j p_i(t)w_{ij}, \qquad (2)$$

where $w_{ij} \geq 0$ is the rate for the transition $i \rightarrow j$, subject to the constraint $\sum_i p_i(t) = 1$. Because of microscopic reversibility, if $w_{ij} \neq 0$ then $w_{ji} \neq 0$. As $t \rightarrow \infty$, the probabilities $p_i(t)$ become independent of time. This stationary solution of the master equation describes an equilibrium system (p_i^{eq}) if detailed balance holds, that is if across all the edges of the network the net flux is zero: $p_i^{eq}w_{ij} - p_j^{eq}w_{ji} = \Delta J_{ij} = 0$. However, $dp_i/dt = 0$ is also satisfied by the less stringent relation
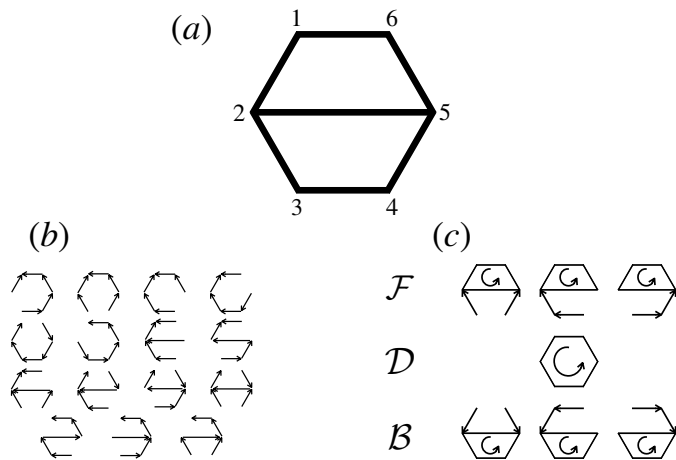
FIG. 5 Example of kinetic network. The network has 6 states and 7 edges [see panel (a)]. Panel (b): the partial graphs oriented in order to extract the stationary probability of state 1. Panel (c): There are three cycle fluxes, $\mathcal{F}$ (upper three figures), $\mathcal{B}$ (bottom three), and $\mathcal{D}$ (central figure). From the figure we extract the values of Σ_ν in Eq. 4 for the three cycles. These are given by the product of the rates flowing towards the cycle [see T. L. Hill (Hill, 2005a) for details]: $\Sigma_{\mathcal{F}} = w_{32}w_{45} + w_{43}w_{32} + w_{34}w_{45}$; $\Sigma_{\mathcal{B}} = w_{12}w_{65} + w_{61}w_{12} + w_{16}w_{65}$; $\Sigma_{\mathcal{D}} = 1$.

$\sum_j(p_i^{ss}w_{ij} - p_j^{ss}w_{ji}) = \sum_j \Delta J_{ij} = 0$. Under these conditions the system could be in a non-equilibrium steady state (NESS). An isolated system is expected to reach an equilibrium state, whereas coupling with an external energy source enables the creation and persistence of a NESS.

From a mathematical standpoint, given a kinetic network described by the master equation, the stationary probabilities p_i^{ss} (or p_i^{eq}) can be obtained using graph-theoretical arguments. The details may be found in the works of Hill (Hill, 1966, 2005a; Hill and Chen, 1975). Briefly, one may construct the set of partial diagrams such that all the states are visited but no cycles are formed (see Fig. 5b). The stationary probability of being in state i is proportional to the sum of all the partial graphs oriented in such a way that the fluxes converge towards state i, or σ_i. The normalization factor is $\Sigma = \sum_i \sigma_i$, so that the stationary probability is,

$$p_i^{ss} = \frac{\sigma_i}{\Sigma}. \tag{3}$$

The stationary flux along one edge may be written as a sum of contributions from all the cycle fluxes of the system. Each cycle can be completed in two directions: one, counterclockwise, labeled as "+"; the other, clockwise, termed "-". The cycle fluxes in the "+" and "-" directions are given by the following relationships (Hill,

2005a),

$$J_{\nu\pm} = \frac{\Sigma_\nu \Pi_{\nu\pm}}{\Sigma}, \tag{4}$$

where Σ_ν is a combination of rates that are specific for cycle ν, $\Pi_{\nu\pm}$ is the products of the rates of the cycle performed in the "+" ($\Pi_{\nu+}$) or "-" ($\Pi_{\nu-}$) direction, and the denominator Σ is defined above (see Fig. 5c). It follows that,

$$\frac{J_{\nu+}}{J_{\nu-}} = \frac{\Pi_{\nu+}}{\Pi_{\nu-}} = e^{\beta A_\nu}, \tag{5}$$

where A_ν (or βA_ν) has been termed the action functional (Lebowitz and Spohn, 1999) or affinity (Schnakenberg, 1976) of the cycle ν. Note that the net direction of completion of cycle ν is given by the sign of $\Delta J_\nu = J_{\nu+} - J_{\nu-}$. It is easy to show that (Hill and Simmons, 1976),

$$\Delta J_\nu A_\nu \geq 0. \tag{6}$$

The above equality holds only when $J_{\nu,+} = J_{\nu,-}$. In other words, ΔJ_ν and A_ν have the same sign, which means that the value of the affinity dictates the direction of the cycle (Hill and Simmons, 1976).

In order to identify the physical interpretation of these mathematical identities we need to connect the rates with thermodynamic quantities, such as energy, entropy, and the chemical potential. Following three different approaches we show that the affinity (Eq. 5) is related to the entropy produced during a cycle. The first two strategies involve the Shannon entropy (Liepelt and Lipowsky, 2007b; Lipowsky and Liepelt, 2008; Schnakenberg, 1976) and fluctuation theorems (Crooks, 1998; Seifert, 2005a,b, 2012), and they will be discussed without specifically referring to the experimental hallmarks for motor velocity described in the previous sections. The last method, which is based on establishing a connection between pseudo-first-order rate constants and the chemical potential (Hill, 2005a), will be developed in the context of molecular motors.

B. Thermodynamics

Molecular machines are enzymes (E) that catalyze the chemical transformation of substrate molecules ($S \rightarrow P$), which in general can be described by the Michaelis-Menten kinetics ($E + S \rightleftharpoons ES \rightarrow E + P$). Of course, the process is reversible, and the enzyme also catalyzes the reverse reaction – the transformation of P into S. In the case of molecular machines, the substrate is ATP, and the products of ATP hydrolysis are ADP and orthophosphate (P_i). (Some molecular machines catalyze the hydrolysis of GTP; mutatis mutandis our considerations do not change.)

We imagine the following experimental setup: a molecular machine operates in a solution containing ATP, ADP, and P_i. The chemical potentials of these three species are given by (Hill, 2005b),

$$
\begin{aligned}
\mu_{\mathrm{ATP}} &= \mu_{\mathrm{ATP}}^0 + k_{\mathrm{B}}T \ln [\mathrm{ATP}] \\
\mu_{\mathrm{ADP}} &= \mu_{\mathrm{ADP}}^0 + k_{\mathrm{B}}T \ln [\mathrm{ADP}] \\
\mu_{\mathrm{P_i}} &= \mu_{\mathrm{P_i}}^0 + k_{\mathrm{B}}T \ln [\mathrm{P_i}] \\
X = -\Delta\mu_{\mathrm{hyd}} &= \mu_{\mathrm{ATP}} - \mu_{\mathrm{ADP}} - \mu_{\mathrm{P_i}} = \\
&= k_{\mathrm{B}}T \ln \frac{\mathrm{K_{eq}}[\mathrm{ATP}]}{[\mathrm{ADP}][\mathrm{P_i}]},
\end{aligned}
\tag{7}
$$

where μ_c (μ_c^0) is the (standard) chemical potential of species c, $[c]$ is the concentration in solution, and $\mathrm{K_{eq}} \approx 4.9 \cdot 10^5$ M (Howard, 2001) is the equilibrium constant for ATP hydrolysis. (Note that we should use activities a_c in Eq. 7 instead of concentrations $[c]$; throughout the review we will assume that $a_c \approx [c]$.) At equilibrium, $([\mathrm{ADP}][\mathrm{P_i}]/[\mathrm{ATP}])_{\mathrm{Eq}} = \mathrm{K_{eq}}$, and $\Delta\mu_{\mathrm{hyd}} = 0$. On the other hand, if $[\mathrm{ATP}]\mathrm{K_{eq}} \gg [\mathrm{ADP}][\mathrm{P_i}]$, then $\Delta\mu_{\mathrm{hyd}} < 0$, and the hydrolysis of ATP is a spontaneous reaction. We exclude un-catalyzed hydrolysis/synthesis of ATP, as it occurs over time-scales beyond our interest (Hulett, 1970); as a consequence, in the absence of molecular machines the concentrations [ATP], [ADP], and $[P_i]$ are constant. The presence of the molecular machine does not alter the equilibrium features of ATP hydrolysis (namely, $\mathrm{K_{eq}}$), however it accelerates the rate of ATP synthesis/hydrolysis. Let the initial concentrations of ATP, ADP, and P_i make ATP hydrolysis a spontaneous reaction. After each catalytic cycle the enzyme attains the same conformation that it had at the start. However, the solution conditions have changed as a substrate (ATP) has been consumed and products (ADP and P_i) have been created. Thus, in the presence of the molecular machine the solution approaches the equilibrium ratio of concentrations of ATP, ADP, and P_i, and monitoring the dynamics under these conditions corresponds to studying the time-dependent relaxation towards equilibrium. At equilibrium, the rate of hydrolyzing and synthesizing ATP is the same.

However, this is not what happens in a cell, where the concentration of ATP is maintained under homeostatic control (Wang et al., 2017) far away from equilibrium; typical concentrations are $[\mathrm{ATP}] \approx 1\mathrm{mM}$, $[\mathrm{ADP}] \approx 10\mu\mathrm{M}$, $[\mathrm{P_i}] \approx 1\mathrm{mM}$ (Howard, 2001), resulting in a chemical potential $X = -\Delta\mu_{\mathrm{hyd}}$ of ≈ 25 (Howard, 2001), which makes ATP hydrolysis a spontaneous reaction ($\Delta\mu_{\mathrm{hyd}} < 0$). This means that the enzymatic cycle of the molecular machine will be driven in the direction that consumes ATP, and ATP hydrolysis provides the driving force that enables the molecular machine to perform work. Without accounting explicitly for the whole cellular apparatus involved in dictating and maintaining the set-point ATP level, in the simplest theoretical model the system is assumed to be in contact with some devices referred to as "chemostats" capable of maintaining the initial concentrations of ATP, ADP, and P_i (Lipowsky and Liepelt, 2008; Lipowsky et al., 2009; Qian and Beard, 2005; Seifert, 2011b). After each cycle the chemostat removes the products from solution and replenishes the substrates, thereby ensuring that after each enzymatic cycle the initial condition is reset, which enables the creation of a NESS for $t \to \infty$.

The system is also in contact with a thermal reservoir with which it exchanges heat (Lipowsky and Liepelt, 2008; Lipowsky et al., 2009; Seifert, 2011b). We assume that the thermal reservoir operates "quasi-statically" on the time-scale of the heat exchange, and the combination of the system and the thermal reservoir is energetically isolated. Because after a cycle the motor has not changed, the total entropy produced is equal to the ratio between the heat absorbed by the thermal reservoir, Q, and the constant temperature T. We exclude pV work from our formulation, but we include the possibility that the molecular machine performs work against a fixed load, so that for a forward (backward) step $W_{\mathrm{mech}} = f d_0$ ($W_{\mathrm{mech}} = -f d_0$), where the choice of the sign implies that a positive force opposes forward movement. Note that the load f is assumed to be clamped, so at every step, regardless of the position of the motor, the cycle is repeated under identical conditions.

The energy $X = -\Delta\mu_{\mathrm{hyd}} > 0$ consumed during a cycle is partitioned into work performed by the motor (W_{mech}) and heat released (Q), that is,

$$
\Delta\mu_{\mathrm{hyd}} + Q + W_{\mathrm{mech}} = 0,
\tag{8}
$$

implying that the total change of entropy of the system plus environment over one cycle is given by,

$$
\Delta S = \frac{Q}{T} = \frac{-\Delta\mu_{\mathrm{hyd}} - f d_0}{T} \geq 0.
\tag{9}
$$

(For backward steps, the sign in front of $f d_0$ is the opposite). The inequality is due to the second law of thermodynamics, which states that we should to find an increase in entropy during the cycle because the combination of system and environment is isolated. Eq. (9) leads to the observation that the maximum amount of force that the motor can resist is given by,

$$
f_{\max} = \frac{-\Delta\mu_{\mathrm{hyd}}}{d_0}.
\tag{10}
$$

A number of authors have discussed variations of this thermodynamic framework. In particular, Seifert elucidated the contribution of the thermodynamic contribution of chemostats (Seifert, 2011b); Qian and Beard discussed a network of reactions in which some metabolites are clamped, while other are introduced at a constant rate (Qian and Beard, 2005).

C. Rate of Entropy Production

Let the entropy of the motor be (Lipowsky and Liepelt, 2008),

$$S_m = -k_{\mathrm{B}} \sum_i p_i \ln p_i, \qquad (11)$$

where the sum extends over all the available conformations of the system. Let us take a time derivative of S_m, and after imposing the conservation of probability ($\sum_i dp_i/dt = 0$) condition, and plugging in the master equation we get,

$$\frac{dS_m}{dt} = k_{\mathrm{B}} \sum_i \sum_j (p_i w_{ij} - p_j w_{ji}) \ln p_i. \qquad (12)$$

Note that in a NESS, $\sum_j (p_i w_{ij} - p_j w_{ji}) = \sum_j J_{ij} = 0$, and as a consequence $dS_m/dt = 0$, as expected. We first symmetrize the result with respect to the indexes i and j, and then add and subtract $1/2 k_{\mathrm{B}} \sum_i \sum_j (p_i w_{ij} - p_j w_{ji}) \ln(w_{ij}/w_{ji})$, which leads to (Lipowsky and Liepelt, 2008; Schnakenberg, 1976),

$$\frac{dS_m}{dt} = k_{\mathrm{B}} \frac{1}{2} \sum_i \sum_j J_{ij} \ln \frac{p_i w_{ij}}{p_j w_{ji}} - k_{\mathrm{B}} \frac{1}{2} \sum_i \sum_j J_{ij} \ln \frac{w_{ij}}{w_{ji}}. \qquad (13)$$

The first of the two terms on the r.h.s. of Eq. 13 is always non-negative, and it has been identified as the rate of entropy production, dS/dt (Lipowsky and Liepelt, 2008; Schnakenberg, 1976). We note parenthetically that some authors (Hill and Simmons, 1976) do not include the factor $1/2$ as they interpret the summation to be carried out over the edges of the network, and not over the states. Because in a NESS $dS_m/dt = 0$, it follows that the second term of the r.h.s of Eq. 13 is the rate of entropy out-flux, and,

$$\begin{aligned}
\frac{dS}{dt} &= k_{\mathrm{B}} \frac{1}{2} \sum_i \sum_j J_{ij} \ln \frac{p_i w_{ij}}{p_j w_{ji}} = \\
&= k_{\mathrm{B}} \frac{1}{2} \sum_i \sum_j J_{ij} \ln \frac{w_{ij}}{w_{ji}} \geq 0.
\end{aligned} \qquad (14)$$

Consider now a cyclic network of N states; the only transitions allowed from state i are $i \rightleftharpoons i+1$ and $i \rightleftharpoons i-1$. (The cyclic nature of the network implies periodic boundary conditions, and so state $i = 0$ and $i = N$ are the same states). In a NESS, the condition $\sum_j J_{ij} = 0$ becomes $\Delta J_{i,i+1} = \Delta J_{i-1,i}$, so that the net flux across all the edges is the same and equal to the net cycle flux, $\Delta J = J_+ - J_-$. It follows that,

$$\frac{dS}{dt} = k_B \Delta J \ln \prod_{i=0}^{N-1} \frac{w_{i,i+1}}{w_{i+1,i}} = \frac{\Delta J A}{T} \qquad (15)$$

where $w_{i,i+1}$ and $w_{i+1,i}$ are the forward and backward rates, respectively. The term $A = k_{\mathrm{B}} T \ln \prod_{i=0}^{N-1} \frac{w_{i,i+1}}{w_{i+1,i}}$ is

the affinity (see Eq. 5), which Hill refers to as a thermodynamic force driving the system out of equilibrium and imposing a NESS (Hill, 2005a). Liepelt and Lipowky identified the thermodynamic force with the entropy production (Liepelt and Lipowsky, 2007b); if the rate for completing the cycle in the direction d is J_d, one can integrate over a time $1/J_d$ the rate of entropy production and identify the entropy ΔS^d produced during one cycle in the d direction as,

$$T\Delta S^+ = k_{\mathrm{B}} T \ln \prod_{i=0}^{N-1} \frac{w_{i,i+1}}{w_{i+1,i}} = -T\Delta S^-, \qquad (16)$$

and, therefore, from Eq. 17 we get,

$$\frac{dS}{dt} = \Delta J \Delta S^+. \qquad (17)$$

The argument can be extended to the case in which the kinetic network is characterized by multiple cycles (Liepelt and Lipowsky, 2007b; Lipowsky and Liepelt, 2008). If ΔJ_ν^{-1} is the average time for completing a cycle ν, and ΔS_ν is the entropy produced with that cycle, we can write,

$$\frac{dS}{dt} = \sum_\nu J_\nu \Delta S_\nu \geq 0. \qquad (18)$$

Equation 6 indicates that each term in the summation is non-negative (Hill and Simmons, 1976).

To summarize, in this section we showed how the action functional (or affinity) can be identified with the entropy produced. We now present a different argument based upon fluctuation theorems leading to the same conclusions.

D. Fluctuation Theorems

Let the probability density of taking a forward path of n steps be $\mathcal{P}_F(i_0, t_0; i_1, t_1; \cdots; i_n, \tau)$, in which the system is in state i_0 at t_0, transitions to state i_1 at t_1 and so on until it reaches state i_n at time τ (see Fig. 6a). We use the subscript "F" to indicate that the transitions are completed forward in time. Using the Markov property, we can rewrite this joint probability as (Crooks, 1998),

$$\begin{aligned}
&\mathcal{P}_F(i_0, t_0; i_1, t_1; \cdots; i_n, \tau) = \\
&= p_{i_0}(t_0) P(i_1, t_1 | i_0, t_0) \cdots P(i_n, \tau | i_{n-1}, t_{n-1}).
\end{aligned} \qquad (19)$$

Assuming that the system is in a steady state at $t = t_0$, we replace $p_{i_0} = p_{i_0}^{ss}$. The conditional probabilities are given by $P(j, t_j | i, t_i) = w_{ij} e^{-W_i(t_j - t_i)}$, where $W_i = \sum_j w_{ij}$. The conditional probability $P(j, t_j | i, t_i)$ is the product of two probabilities. The first is w_{ij}/W_i, the probability of making a transition to state j among
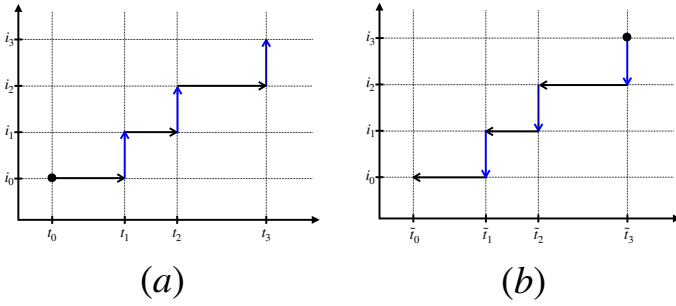
FIG. 6 Path in state space executed forward (a) and backward (b) in time. The starting point is the dot, waiting times at a fixed state are in black, transitions are in blue.

all the other possibilities when starting in state i. The second is $W_i e^{-W_i(t_j - t_i)}$, the probability that this transition happens after time $(t_j - t_i)$. The reason for the W_i in the exponent is that the mean waiting time to transition from state i to any other state is W_i^{-1}, independent of state j. This independence is a well-known property of escape times in Markov networks, which may be computed through well-known methods (Van Kampen, 2007). Therefore, the joint probability density becomes (Seifert, 2012),

$$
\mathcal{P}_F(i_0, t_0; i_1, t_1; \cdots; i_n, \tau) =
$$
$$
= \mathcal{N}^{-1} p_{i_0}^{ss} \prod_{\alpha=0}^{n-1} w_{i_\alpha, i_{\alpha+1}} e^{-W_{i_\alpha}(t_{\alpha+1} - t_\alpha)}, \quad (20)
$$

where $t_n = \tau$ and $\mathcal{N}(t_0, \tau)$ is a normalization factor to ensure that $\sum_{i_0, i_1, \cdot, i_n} \int_{t_0}^\tau dt_1 \int_{t_1}^\tau dt_2 \cdots \int_{t_{n-1}}^\tau dt_n \mathcal{P}_F = 1$. The probability of completing the time-reversed path (Fig. 6b) is,

$$
\mathcal{P}_B(i_n, \tilde{\tau}, i_{n-1}, \tilde{t}_{n-1}; \cdots; i_0, \tilde{t}_0) =
$$
$$
= \tilde{\mathcal{N}}^{-1} p_{i_n}^{ss} \prod_{\alpha=0}^{N-1} w_{i_{\alpha+1}, i_\alpha} e^{-W_{i_\alpha}(\tilde{t}_\alpha - \tilde{t}_{\alpha+1})}, \quad (21)
$$

in which we started in state i_n at time $\tilde{t}_n = \tau - t_n$ and retrace the forward path until we reach state i_0 at time $\tilde{t}_0 = \tau - t_0$. Note that the initial probability is now $p_{i_n}^{ss}$ and because each path could be performed forward or backward in time, the normalization factor $\mathcal{N} = \tilde{\mathcal{N}}$. Following Seifert (Seifert, 2005b), we define R to be the logarithm of the ratio between the probability for forward and backward paths, and for the sake of simplicity we consider only cyclical pathways, in which $i_0 = i_n$. Therefore,

$$
R = \ln \frac{\mathcal{P}_F}{\mathcal{P}_B} = \ln \prod_{\alpha=0}^{n-1} \frac{w_{i_\alpha, i_{\alpha+1}}}{w_{i_{\alpha+1}, i_\alpha}}. \quad (22)
$$

Note that the time-dependent terms in the joint probabilities cancel out, so we neglect them in the following.

(In alternative, we may consider the time at which these jumps occur to be fixed.) Here, R has a structure analogous to that of the affinities introduced in the previous sections. More precisely, given a set of states visited we can rewrite,

$$
R = \sum_\nu n_\nu \beta A_\nu, \quad (23)
$$

where n_ν is the net number of times the cycle ν has been completed during the n-step path associated with R. In order to justify this expression, note that every time that during the path a cycle is not completed, the path re-traces it self and those branch-like excursions do not contribute to R. If we average over all paths that complete a cycle of length n, we find the following sequence of identities (Seifert, 2005b),

$$
\langle e^{-R} \rangle = \sum_{\text{paths}} P_F e^{-R} = \sum_{\text{paths}} P_B = 1. \quad (24)
$$

Using Jensen's inequality, $\langle e^{-R} \rangle \geq e^{-\langle R \rangle}$, leads to,

$$
\langle R \rangle \geq 0. \quad (25)
$$

From Eq. 23 we obtain,

$$
\langle R \rangle = \sum_\nu \langle n_\nu \rangle \beta A_\nu \geq 0. \quad (26)
$$

We now introduce an Arrhenius-type relationship between the rates, $w_{ij}/w_{ji} = e^{-\beta \Delta F_{ij}}$, where $\Delta F_{ij} = F_j - F_i$ is the free energy difference between state i and state j. The free energy difference accounts for three contributions: (i) the intrinsic free energy of a state may change; (ii) the motor may bind/release ATP, ADP, and P_i; (iii) the motor may perform work against an external load. After a cycle, the motor returns to the initial state, and as a consequence the intrinsic free energy does not carry any contribution to the cycle. The hydrolysis of ATP contributed with the release of energy equal to $-\Delta \mu_{\text{hyd}}$, and the work performed per each displacement of size d_0 against a load f is equal to $-f d_0$. It follows that (Lipowsky and Liepelt, 2008),

$$
\prod_{|i,j\rangle \in \nu} \frac{k_{ij}}{k_{ji}} = e^{-n_{\nu,\text{ATP}} \beta \Delta \mu_h - \beta f l_\nu} = e^{\frac{\Delta S_\nu}{k_B}}, \quad (27)
$$

where, following the notation of Lipowsky and Liepelt (Lipowsky and Liepelt, 2008), $|i,j\rangle$ corresponds to a directed edge from state i to j, and the sum is extended over all the directed edges belonging to cycle ν. Here, $n_{\nu,\text{ATP}}$ is the net number of ATP hydrolyzed during cycle ν, and l_ν is the net displacement along the track (a multiple of d_0). The last identity follows from Eq. 9, and from Eq. 26 gives,

$$
k_B \langle R \rangle = \sum_\nu \langle n_\nu \rangle \Delta S_\nu \geq 0. \quad (28)
$$

V. MOLECULAR MOTORS – MODELS WITHOUT DETACHMENT

In this section we develop models of molecular motors with increasing complexity, from one-state models, to uni-cyclic models, ending with multi-cycle kinetic networks. We show that experimental evidence and thermodynamic insights suggest that the introduction of multiple cycles creates models that are physically more sensible. We discuss more in detail the efficiency of the motors, and we conclude the section by describing models that account for the detachment of the molecular motor from the track.

A. Molecular Motors are Processive Enzymes

The catalytic cycle of kinesin, dynein, and myosin proceeds via multiple steps, in which ATP binding, hydrolysis, and release of ADP and P_i lead to substantial changes in the motor conformation. The presence of the cytoskeletal filament, CF (F-actin in the case of myosin, microtubules for dynein and kinesin), increases the rate of ATP hydrolysis (De la Cruz *et al.*, 2001, 1999; Hackney, 1988; Homma and Ikebe, 2005; Ori-McKenney *et al.*, 2010). When the structural changes are properly rectified under the specially designed interactions with the appropriate CFs, the conformational fluctuations in M are transduced to a predominantly uni-directional motion along the track, and generates mechanical forces against an external load. The free energy source necessary to perform this movement (or work) is provided by the hydrolysis of ATP.

Enzymes that work in conjunction with substrates (as is the case for Ms and CFs) are said to be "processive" if they perform multiple catalytic cycles without fully disengaging from the CFs (Schnitzer and Block, 1995). Processive motors move long distances along the CFs by hydrolyzing one ATP molecule per step without dissociating from the CF. In general (although with some fascinating exceptions (Inoue *et al.*, 2002; Post *et al.*, 2002)) processive movement requires the cooperation of multiple enzymes: some members of the ensemble proceed forward while the others hold tight onto the CF. From here on, we focus our discussion exclusively on dimeric processive motors made of two identical enzymes of the same family, which we refer to as "heads." With this we may identify the leading head (LH) as the one that is in front of the dimeric complex, while the other is the trailing head (TH). We reserve the word "motor" for the motile construct, which is a dimer in the case of myosin V (Mehta *et al.*, 1999), VI (Rock *et al.*, 2001), and X (Sun *et al.*, 2010), conventional kinesin (kinesin-1) (Howard *et al.*, 1989), and cytoplasmic dynein (Reck-Peterson *et al.*, 2006b).

In the well-accepted hand-over-hand stepping mechanism (Ökten *et al.*, 2004; Sun *et al.*, 2010; Toba *et al.*, 2006; Yildiz *et al.*, 2003, 2004a,b), the TH of the motor detaches from the filament, overcomes the bound head, and advances the motor by reaching a forward target binding site (TBS). After the step is completed, the role of the two heads is reversed, and the process is repeated identically, until the motor detaches from the CF. A few key ingredients are necessary for processive, hand-over-hand movement to occur. First, in order to prevent premature end of the processive run, the head that remains bound to the track during a step should not dissociate from the CF at least until the detached head reaches the TBS. Second, the two heads should go through their catalytic cycles out-of-phase (Block, 2007), so that the TH is more likely to detach from CF than the LH. This requirement necessarily implies that the two heads ought to communicate with one another through action at a distance, which may be viewed as a complex form of allostery (Thirumalai *et al.*, 2019). Third, during a step a combination of conformational transitions in the two heads biases the diffusive motion of the free head towards the TBS unless a large resistive force is applied.

Meeting the first two conditions requires specific features of the hydrolytic cycles that the two heads must undergo. For instance, if the rate-limiting step occurs in a CF-bound conformation, the chances of prematurely ending the processive run are diminished. In addition, the interaction between the two heads is believed to be key in ensuring that the TH steps first, and that the LH is unlikely to detach during a step. This is possible if the interplay between the two heads slows down (or "gates") some of the steps of the LH catalytic cycle, or accelerates them in the trailing head (also referred to as "gating") (Block, 2007; Hancock, 2016; Sweeney and Houdusse, 2010). The structural bias towards the TBS is provided by a conformational transition of the CF-bound motor that projects forward the stepping head. For the three classes of motors mentioned before, the lever arm swing in myosin, the neck-linker docking in kinesin, and the bent→straight transition of the linker domain in dynein provide the requisite forward bias. Recently, there has been both theoretical (Mugnai *et al.*, 2019) and computational studies on dynein (Goldtzvik *et al.*, 2018a; Wang *et al.*, 2018) that have focused on the gating mechanism, which could depend on both ATP and the hydrolysis product, ADP. Despite considerable work a full understanding of how gating occurs in motors has not been quantitatively elucidated. It is remarkable that these structural transitions, induced by ATP binding and hydrolysis, which occur on relatively small length scales are amplified by responses in other structural elements. The terms "power stroke" (Dominguez *et al.*, 1998) and "Brownian ratchet" (Rice *et al.*, 1999) have been used to schematize this forward bias, depending on whether the emphasis is placed on the mechanistic (order→order) or

stochastic (disorder→order) nature of the conformational transitions.

In the simplest cases, the step-size of a motor is commensurate with the filament repeat – for motors moving along the MT, $d_0 \approx 8.2$ nm, whereas $d_0 \approx 36$ nm for actin-based motors (myosins). However, not all motors walk precisely by following this rule: myosin VI and dynein, for instance, display a broad step-size distribution, in which hand-over-hand steps are intertwined with inchworm-like movements and frequent backward steps (Nishikawa *et al.*, 2010; Reck-Peterson *et al.*, 2006a). Of course, the variability in the step sizes is also a consequence of the architecture of the motor. Thus, both the structural design of the motor and the coupling to catalytic cycles determine not only the stepping kinetics but also the precision of the step sizes.

B. Processive Motor Velocity

A variety of single-molecule techniques are able to inform on the motile properties of molecular motors. The motor or the track may be labeled with a fluorescent (Nishikawa *et al.*, 2010; Yildiz *et al.*, 2003) or refractive (Isojima *et al.*, 2016; Mickolajczyk *et al.*, 2015) probe. Monitoring the time-dependent changes in the location of the such probes enables the determination of the velocity of the motor as a function of nucleotide concentration. Alternatively, optical trapping techniques may be used to follow the displacement of the motor or the filament, and to investigate the motor response to the external load (Block *et al.*, 1990; Finer *et al.*, 1994; Howard *et al.*, 1989). Much of our current knowledge about kinesin comes from single molecule experiments (Block *et al.*, 1990; Howard *et al.*, 1989), which were developed approximately at the time of the discovery of kinesin (Brady, 1985; Vale *et al.*, 1985). It has been shown that each step of kinesin motors along MTs is tightly coupled with hydrolysis of one ATP molecule, that is, a single event of ATP hydrolysis by kinesin leads to ∼8 nm step along MTs (Schnitzer and Block, 1997). Thus, within the Michaelis Menten (MM) scheme, the velocity of kinesin is associated with enzyme turnover rate is,

$$V = d_0 \frac{k_{\text{cat}}[\text{ATP}]}{K_M + [\text{ATP}]}, \tag{29}$$

where $d_0 = 8$ nm is the step size of the kinesin, k_{cat} and K_M are the rate of catalysis and the Michaelis constant, respectively. For kinesin-1, the typical experimental value for the Michaelis constant is $K_M \gtrsim 50$ μM, and for $[\text{ATP}] \gtrsim 1$ mM the velocity of kinesin is saturated to its maximum value $V_{\text{max}} = d_0 k_{\text{cat}} \approx 8$ nm/10 ms = 0.8 μm/s (Kolomeisky and Fisher, 2007; Visscher *et al.*, 1999). Similar single-molecule studies have been conducted on other motors, showing that the Michealis-Menten scheme provides a good paradigm to rationalize the ATP concentration dependence of velocity. However, the parameters of the fit depend on the motor: for myosin V it was found that $V_{\text{max}} \approx 0.55$ μm/s, with $K_M \approx 163$ μM (Baker *et al.*, 2004), whereas the slower myosin VI has $V_{\text{max}} \approx 0.16$ μm/s and $K_M \approx 274$ μM [fit of Eq. 29 to data from (Elting *et al.*, 2011)].

Experiments employing optical tweezers have shown that external loads affect the ATP chemistry in the catalytic site as well as the motility of the motor (Block *et al.*, 2003; Oguchi *et al.*, 2008; Veigel *et al.*, 2005; Visscher *et al.*, 1999). Thus, the resulting force-velocity-ATP relationship has been a landmark measurement, which is used as a constraint to decipher the mechanism underlying kinesin motility, and to construct the appropriate SKMs. Incorporation of the effect of load into MM model was suggested by making the parameters k_{cat} and K_M force-dependent (Schnitzer *et al.*, 2000),

$$V(F, [\text{ATP}]) = \frac{d_0 k_{\text{cat}}(F)[\text{ATP}]}{k_{\text{cat}}(F)/k_b(F) + [\text{ATP}]} \tag{30}$$

where $k_{\text{cat}}(F) = k_{\text{cat}}^o/(p_{\text{cat}} + (1 - p_{\text{cat}})e^{F\delta_{\text{cat}}/k_B T})$ and $k_b(F) = k_b^o/(p_b + (1 - p_b)e^{F\delta_b/k_B T})$. The fit of the F-dependent velocity data of motor using Eq.30 is reasonable as long as the magnitude of load (F) is small (Schnitzer *et al.*, 2000). However, the conventional MM model has an intrinsic drawback when the external load approaches the stall force value ($\approx$ 7 pN for kinesin) and becomes greater than the stall force. An increase of load ought to induce backward stepping ($V < 0$); yet no modification of Eq.29 produces a negative velocity! One possible fix for the failure of the MM model at large load is to permit in the model a reverse reaction current, which is realized by rendering every elementary reaction step within the kinesin cycle "reversible." We discuss later how such models have been constructed in the context of Markov jump processes.

C. Periodic Lattice Model

In a molecular motor (i) binding, release, and chemical transformations of ATP, ADP and P_i from the motor head domain, and (ii) the advancement along the track occur over much slower times (ms - s) compared to the fluctuations of molecular conformations (ns - μs). The time-scale separation enables the construction of a Markov jump processes in which the state of the system is defined by two coordinates: (i) one of the N intermediates defining the chemical state of the substrate, and (ii) the location l along the track. The probability density of the i-th state ($i = 1, 2, \ldots, N$) at filament site l obeys

FIG. 7 Example of simple kinetic networks. Panel (a): the network has only one state, $N = 1$. The forward rate is given by $u = u^*[\text{ATP}]$, the backward rate is $w = w^*[\text{ADP}][\text{P}_i]$. Both forward and backward rates are pseudo-first-order, with units $\text{mM}^{-1}\text{s}^{-1}$ and $\text{mM}^{-2}\text{s}^{-1}$, respectively. Panel (b): multi-state model. Here, the rates u_1 and w_1 are arbitrarily chosen to depend on [ATP] and [ADP][P_i], respectively. Panel (c): a cycle corresponding to the multi-state model.

the master equation,

$$\frac{dP_i(l,t)}{dt} = \sum_{l'=l-1}^{l+1} \sum_j [w_{j,l\to i,l'} P_j(l',t) - w_{i,l\to j,l'} P_i(l,t)] \tag{31}$$

where $\sum_l \sum_i P_i(l,t) = 1$, and $w_{j,l\to i,l'}$ is the rate for going from state j in location l to state i on F site $l' = l, l\pm 1$. Under the assumption that the track is infinite and periodic, that is that the rates do not depend on l, one may sum the r.h.s and the l.h.s of the master equation and eliminate l. As a result, we obtain the master equation in Eq. 2, and we can use the theoretical framework presented before.

Periodic one-dimensional models, which were originally suggested by Derrida (Derrida, 1983) in order to study the mean velocity and diffusion constant of random systems, have been widely adopted in describing the dynamics of molecular motors. In particular, Fisher and Kolomeisky (Fisher and Kolomeisky, 1999a, 2001; Kolomeisky and Fisher, 2007) have pioneered this approach and specifically used the reversible kinetic model to describe the motility of kinesin-1 and myosin V. Numerous other studies have used variants of these models to discuss the thermodynamic features of stochastic motors. We review the main successes and shortcomings of these models in the context of molecular motors as their complexity is increased.

1. One-state Models

In the simplest possible model, the motor is defined by a single ($N = 1$) chemical state at each site along the track (Fig. 7a). The rate of forward stepping, u, depends on the concentration of ATP. Microscopic reversibility demands that the backward transition is possible as well, which is then nominally associated with ATP synthesis. The rates of forward and backward stepping are modeled as pseudo-first order processes,

$$u = u^*[\text{ATP}], \quad w = w^*[\text{ADP}][\text{P}_i], \tag{32}$$

leading to $u/w = u^*/w^* \cdot [\text{ATP}]/([\text{ADP}][\text{P}_i])$. In equilibrium, the probability of going forward is identical to the probability of backward stepping, which implies that $(u/w)_{\text{Eq}} = 1$. From Eq. 7, at equilibrium $\Delta\mu_{\text{hyd}} = 0$ and $([\text{ADP}][\text{P}_i]/[\text{ATP}])_{\text{Eq}} = \text{K}_{\text{eq}}$. As a consequence, using Eq. 7 we obtain that the ratio between forward and backward rates is given by,

$$\begin{aligned} (a) \quad & \frac{u}{w} = \frac{u^*}{w^*}\frac{[\text{ATP}]}{[\text{ADP}][\text{P}_i]} = e^{\beta X} \\ (b) \quad & \frac{u^*}{w^*} = \text{K}_{\text{eq}}, \end{aligned} \tag{33}$$

where we recall that $X = -\Delta\mu_{\text{hyd}}$. The average, stationary velocity is given by the step-size times the difference between the forward and backward rates,

$$v = d_0(u - w) = d_0 w\left(e^{\beta X} - 1\right). \tag{34}$$

The crucial insight from this expression pertains to the interplay between the non-equilibrium steady-state (NESS) homeostatically maintained by the cell and molecular motor function: in the cytosol $X > 0$, thus $v > 0$; at equilibrium, $X = \Delta\mu_{\text{hyd}} = 0$, implying that $v = 0$.

In the absence of applied load, the motor moves without performing any work. In this case, the entire free energy extracted from the hydrolysis of ATP is dissipated into heat, $Q = -\Delta\mu_{\text{hyd}}$. From Eq. 17, the rate of heat dissipation per step is given by,

$$T\frac{d_iS}{dt} = \dot{Q} = k_{\text{B}}T(u - v)\ln\frac{u}{v} = (u - v)(-\Delta\mu_{\text{hyd}}) \geq 0. \tag{35}$$

Clearly, $\dot{Q} \geq 0$, and the equality holds in equilibrium, where $u = v$. If instead the motor is subjected to a force, f, opposing its movement, at each step the motor performs mechanical work $W = fd_0$. In this case, the free energy released by ATP hydrolysis is partitioned between work and heat, $Q = -\Delta\mu_{\text{hyd}} - W$. In order to account for the presence of applied load, the rates may be modified using the Bell model (Fisher and Kolomeisky, 1999b),

$$\begin{aligned} (a) \quad & u(F) = u^*[\text{ATP}]e^{-\beta\theta fd_0}, \\ (b) \quad & w(F) = w^*[\text{ADP}][\text{P}_i]e^{\beta(1-\theta)fd_0}. \\ (c) \quad & \frac{u}{w} = e^{\beta(X - fd_0)} \end{aligned} \tag{36}$$

Here, the coefficient θ indicates how the load is partitioned between the forward and backward steps, and it identifies the position of the transition state along the direction of motion (Fisher and Kolomeisky, 2001). For $\theta = 0$, the forward step is unaffected by force, whereas backward steps are insensitive to load if $\theta = 1$. These two extreme cases have been referred to as "power stroke" and "ratchet", respectively (Howard, 2006). In a "power stroke", the chemical step (crossing the transition state) occurs before the movement forward of the motor. In contrast, in the case of a "ratchet" the motor fluctuates between the sites l and $l + 1$ until it is captured in the forward site by ATP hydrolysis. In the presence of backward load the velocity becomes,

$$v = d_0 w(0) e^{\beta(1-\theta)fd_0} \left(e^{\beta X - \beta fd_0} - 1 \right). \qquad (37)$$

It is clear that the motor stalls under a resistive load of,

$$f_{\text{stall}} = \frac{X}{d_0} = -\frac{\Delta\mu_{\text{hyd}}}{d_0}, \qquad (38)$$

at which the mean motor velocity becomes $v = 0$. It is clear that $f_{\text{stall}} = f_{\text{max}}$ for this model. Using Eq. 38 together with the value of $\Delta\mu_{\text{hyd}}$ in the cell and the average step-size of a motor, one can predict the value of f_{stall}: for kinesin $f_{\text{stall}} \approx 12.5$pN, for myosin V $f_{\text{stall}} \approx 2.8$nm. In the case of myosin V, f_{stall} approximates the value of the stall force measured experimentally; for kinesin, the f_{max} is about twice as large as the stall force. Note also that according to Eq. 38, F_s depends on [ATP]; a few experiments have suggested that this is not the case for a variety of motors (Carter and Cross, 2005; Gennerich *et al.*, 2007; Nishiyama *et al.*, 2002; Toba *et al.*, 2006; Uemura *et al.*, 2004a), although other studies have observed such a dependence (Visscher *et al.*, 1999). The independence of F_s on [ATP] might be reasonable because each head has only one binding pocket for ATP, and hence [ATP] should not determine F_s. This assumption would be valid only if the experiments are performed by continuously regenerating ATP and removing the products. Otherwise, it would be difficult to prove that F_s would be independent of [ATP]. The relationship $v(f, \Delta\mu)$ allows one also to compute the power output of the motor, defined as $P = vF$, and the motor efficiency,

$$\eta(f, X) = \frac{fd_0}{X}, \qquad (39)$$

that is the ratio of the work produced over the energy released by the hydrolysis of ATP. Note that the power is zero when $f = 0$ and at the stall force (where $v = 0$), thus for some force $0 < f^* < F_s$ it reaches a maximum P^*. The force-velocity curve, power output, efficiency, and efficiency at maximum power (η^*) have been used to compare ratchet-like motors ($\theta = 1$) with power-stroke-driven ($\theta = 0$). Given a value of the driving force X, power strokes result in larger velocity at fixed F, and

display higher efficiencies and exert more power at a given velocity (Wagoner and Dill, 2016). Furthermore, both η^* and P^* at fixed X increase as $\theta \to 0$ (Schmiedl and Seifert, 2008). Finally, from Eq. 17 we write,

$$T\frac{d_iS}{dt} = \dot{Q} = (u - w)\ln\frac{u}{w} = (u - w)(-\Delta\mu_{\text{hyd}} - fd_0), \qquad (40)$$

where we used Eq. 36c and Eq. 37. Note that, again, $\dot{Q} \geq 0$, and the equality holds at equilibrium ($u = w$).

One-state models are appealing because of their simplicity; for instance, they have only two fitting parameters with straightforward physical interpretations: w^* sets the time-scale (see Eq. 34), and θ establishes the location of the transition state during a step. However, these models fail in reproducing the dependence of the velocity on ATP concentration observed in experiments (Eq. 29). According to Eq. 34 v grows without saturating as [ATP] increases, which is not physical. In order to solve this problem, one must use kinetic models with $N > 1$.

2. Multi-state, Uni-cycle Models

Figure 7b shows a kinetic model for molecular motors with $N > 1$ intermediates. Assuming that the filament is periodic, we can drop the label l and construct an equivalent uni-cycle in Fig. 7c, where the rate u_N is associated with a forward step, and w_1 with a backward displacement. From Eq. 4 we obtain the following relationships,

$$
\begin{aligned}
(a) \quad & J^+ = \frac{\prod_{i=1}^{N} u_i}{\Sigma(\{u, w\})}, \quad J^- = \frac{\prod_{i=1}^{N} w_i}{\Sigma(\{u, w\})}, \\
(b) \quad & \frac{\prod_{i=1}^{N} u_i}{\prod_{j=1}^{N} w_j} = \frac{J^+}{J^-} = e^{\beta X}, \\
(c) \quad & \frac{u_1^* \prod_{i>1}^{N} u_i}{w_1^* \prod_{j>1}^{N} w_j} = \text{K}_{\text{eq}},
\end{aligned}
\qquad (41)
$$

where J^+ and J^- are the fluxes to complete a cycle in the clockwise and counterclockwise direction, respectively, and $\Sigma(\{u, w\})$ is a function of all the rates (see Eq. 3). The connection with thermodynamics is provided by Eq. 41b, which is equivalent to Eq. 33a; Eq. 41c holds because in equilibrium $X = 0$ (see Eq. 33b for the $N = 1$ case). For the case of uni-cyclic network models, the net steady state flux ($\Delta J = J^+ - J^-$) along the cycle is obtained by calculating $\Delta J = w_{i,i+1}p_i^{ss} - w_{i+1,i}p_{i+1}^{ss}$ at any edge between the two neighboring chemical states along the cycle. Therefore, the velocity, which is obtained as

the flux across edge $N \underset{w_1}{\overset{u_N}{\rightleftarrows}} 1$, is given by,

$$v = d_0 \Delta J = d_0 J^- (e^{\beta X} - 1) = d_0 \frac{\prod_{i=1}^N w_i}{\Sigma(\{u, w\})} (e^{\beta X} - 1),$$
(42)

The similarity between Eq. 34 and Eq. 42 is clear: both of the expressions relate the velocity to the exponential of the free energy released upon the hydrolysis of ATP, and in both cases at equilibrium the motor does not move. However, the expression in Eq. 42 saturates at large [ATP], in agreement with experiments (see Eq. 29).

Compared to the case of a single chemical state, new load distribution factors are necessary for each transition rate, i.e.

$$u_i(f) = u_i(0) e^{-\beta \theta_i^+ f d}, \quad w_i(f) = w_i(0) e^{\beta \theta_i^- f d}. \quad (43)$$

The distribution factors obey the relationship $\sum_{i=1}^N (\theta_i^+ + \theta_i^-) = 1$ (Fisher and Kolomeisky, 1999b). It follows that,

$$\frac{\prod_{i=0}^{N-1} u_i(f)}{\prod_{i=0}^{N-1} w_i(f)} = \frac{J^+(f)}{J^-(f)} = e^{\beta X - \beta f d_0}, \quad (44)$$

and given $\Delta J(f) = J^+(f) - J^-(f)$, the force-velocity relation is,

$$v(f) = d_0 \Delta J(f) = d_0 \frac{\prod_{i=1}^N w_i(f)}{\Sigma(f)} \left(e^{\beta X - \beta f d_0} - 1 \right). \quad (45)$$

Although $v(F)$ is a complicated function of all the parameters (Fisher and Kolomeisky, 1999b; Wagoner and Dill, 2016), the expression for a theoretical estimation of the stall force is the same as the one found for $N = 1$, given by Eq. 38, and it is again equal to $f_{\max}$ in Eq. 10.

In principle, one could design kinetic networks with an arbitrary number of intermediates. However, because the number of free parameters becomes $4N - 2$, it is desirable to establish the minimal N that enables an accurate description of the experimental data as accurately as possible. Let $D = \lim_{t \to \infty} \frac{d}{dt} (\langle x(t)^2 \rangle - \langle x(t) \rangle^2)$ be the dispersion, and $v = \lim_{t \to \infty} \frac{d}{dt} \langle x \rangle$ is the velocity; the randomness parameter for a quantity x is defined as,

$$r = \frac{2D}{d_0 v}. \quad (46)$$

It can be shown that $N \geq 1/r$ (Kolomeisky and Fisher, 2007; Koza, 2002), and therefore r^{-1} sets a lower bound to N that is needed to account for measurements. Note that r is a function of ATP concentration and external load.

Following Eq. 17, the rate of heat released during a cycle (Hill, 2005a; Qian, 2007; Qian and Beard, 2005; Toyabe et al., 2010; Zimmermann and Seifert, 2012) is,

$$\dot{Q} = \Delta J k_B T \ln \frac{J^+}{J^-} = \Delta J (-\Delta \mu_{\text{hydr}} - f d_0) = \dot{E} + \dot{W} \geq 0,$$
(47)

where $\dot{E}$ is the rate of free energy expended per cycle. Note that in uni-cyclic network models, $\dot{W} = -\Delta J(f) F d_0 = 0$ either at $f = 0$ or if the motor is subject to an opposing stall force ($f = F_s$), where $\Delta J(F_s) = 0$; thus the work production ($\dot{W}$) is a non-monotonic function of f, whereas $\dot{E}$ and $\dot{Q}$ decrease monotonically with F.

Although the conventional N-state unicyclic models (Fisher and Kolomeisky, 1999a, 2001; Kolomeisky and Fisher, 2003) appear successful in describing the motility and thermodynamics of molecular motors, unicyclic models encounter two serious problems, especially when the molecular motor is stalled or starts taking backward steps at large hindering loads (Astumian and Bier, 1996; Hyeon et al., 2009; Liepelt and Lipowsky, 2007a). First, the backward step in the unicyclic network, by construction, is produced by a reversal of the forward cycle, which implies that the backward step is always realized via the synthesis of ATP from ADP and P_i. This is the case for rotary ATP-synthases, which function as transducers of electrochemical potential into the synthesis of ATP in the mitochondria (Alberts et al., 2008), and depending on the sign of $\Delta \mu_{\text{hyd}}$ and on environmental conditions (e.g. the presence and strength of a proton gradient and applied load), these spectacular motors can reverse their function (Itoh et al., 2004; Turina et al., 2003). Although ATP hydrolysis is reversible for linear molecular motors (Bagshaw and Trentham, 1973; Hackney, 2005), back-stepping has not been associated with ATP synthesis. Rather they are associated with ATP-independent "slippage" or coupled to fuel consumption just as forward stepping (Carter and Cross, 2005; Clancy et al., 2011; Gebhardt et al., 2006; Ikezaki et al., 2013). This requires an update of the kinetic network.

In addition, the unicyclic network models lead to $\dot{Q} = 0$ under stall conditions (Eq. 47), which contradicts physical reality. For example, an idling car still burns fuel and dissipates heat ($\dot{Q} \neq 0$)! There certainly exist fundamental differences between molecular world and macroscopic counterpart in that the former is subject to a large degree of fluctuations, which permit the reverse process of ATP hydrolysis (i.e., synthesis) or negative heat dissipation for a sub-ensemble of the entire realizations. Yet, the mean of entropy production $\dot{Q}/T$ is still bound to be positive as demanded by thermodynamics, and the probability of a local violation of this principle becomes vanishingly small as the system size grows. To ameliorate the aforementioned physically problematic interpretation, associated with the backward step mechanism, it has been proposed to extend the unicyclic network into a network with multiple cycles (Clancy et al., 2011; Hyeon et al., 2009; Liepelt and Lipowsky, 2007a; Yildiz et al., 2008), so that the kinetic pathway of ATP-induced (fuel-burning) backward step or ATP-consuming stall can naturally be considered in the model.
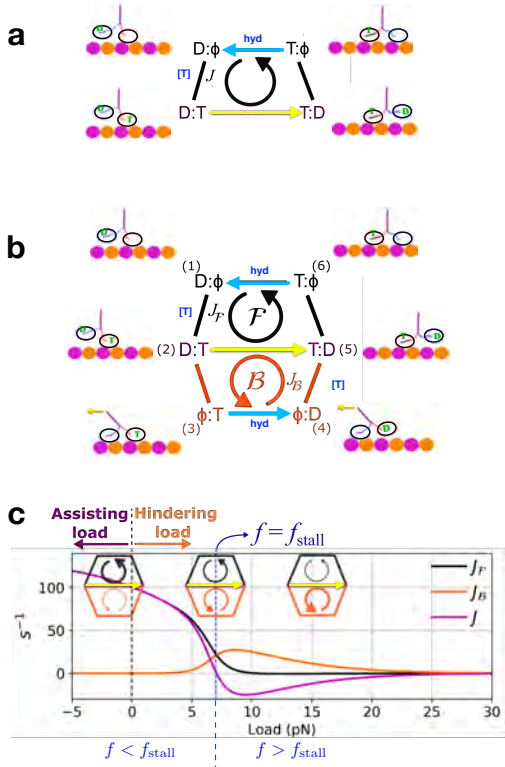
FIG. 8 **b.** Schematics of the network model representing the dynamics of double-headed kinesin-1. T, D, and ϕ denote ATP-, ADP-bound, and apo state, respectively. Through ATP binding $[(1)\rightarrow(2)]$, mechanical step $[(2)\rightarrow(5)]$, release of ADP $[(5)\rightarrow(6)]$, and hydrolysis of ATP $[(6)\rightarrow(1)]$, kinesin moves forward in the $\mathcal{F}$-cycle $[(1) \rightarrow (2) \rightarrow (5) \rightarrow (6) \rightarrow (1)]$, backward in the $\mathcal{B}$-cycle $[(4) \rightarrow (5) \rightarrow (2) \rightarrow (3) \rightarrow (4)]$. The arrows in the figure depict the direction of major counterclockwise reaction current in each cycle. In both cycles, each chemical step is reversible and k_{ij} defines the transition rate from the i-th to j-th state. **c.** Reaction current $J_{\mathcal{F}}$, $J_{\mathcal{B}}$, and J as a function of load. The three cartoons illustrate the amount of current flowing along the $\mathcal{F}$ and $\mathcal{B}$ cycles as a function of f. $f < 0$ and $f > 0$ are the assisting and hindering load, respectively.

3. Multi-cycle Models

The $(N = 6)$-state double cycle network in Fig. 8b is discussed here as a minimal kinetic model to account for subtleties in the physics of kinesin under external load. In the model, the rate constants at the edges between the adjacent chemical states are given as

$$k_{25}(f) = k_{25}^o e^{-\theta f d_0/k_B T}$$

$$k_{52}(f) = k_{52}^o e^{(1-\theta)f d_0/k_B T}$$

$$k_{ij}(f) = 2k_{ij}^o(1 + e^{\chi_{ij} f d_0/k_B T})^{-1} \text{ for } ij \neq 25 \text{ or } 52 \quad (48)$$

For the sake of simplicity, it is assumed that $(2) \rightleftharpoons (5)$, the step associated with the switching of the trailing and leading head positions (the yellow arrow in Fig.8b), obeys

the Bell-like force dependence. When $f > 0$, the force exerted on the motor hinders the forward step and enhances the backward step. Other steps $(ij \neq 25 \text{ or } 52)$ are abolished $(k_{ij} \rightarrow 0)$ at large $f(> 0)$, which corresponds to the physical situation where the large external force impedes the chemical steps such as ATP binding, hydrolysis, and ADP release by deforming the conformation of molecular motor. The model consists of two sub-cycles $\mathcal{F}$ $((1) \rightleftharpoons (2) \rightleftharpoons (5) \rightleftharpoons (6) \rightleftharpoons (1))$ and $\mathcal{B}$ $((2) \rightleftharpoons (3) \rightleftharpoons (4) \rightleftharpoons (5) \rightleftharpoons (2))$. At small $(f \approx 0)$ or assisting force $(f < 0)$, which renders $k_{25} \gg k_{52}$, the reaction current is mainly formed along the counterclockwise direction of $\mathcal{F}$ cycle $(\mathcal{F}^+)$; in contrast, at large hindering force $(f > 0)$ above the stall condition $(f = f_s)$, the current flow along the counterclockwise direction of $\mathcal{B}$ cycle $(\mathcal{B}^+)$, which corresponds to the ATP hydrolysis-induced backstep.

Owing to the microscopic reversibility, each cycle can in principle be performed in both counterclockwise (+ sign, $\mathcal{F}^+,\mathcal{B}^+$) and clockwise (− sign, $\mathcal{F}^-,\mathcal{B}^-$) directions. Forward (backward) steps occur via the $(2) \rightarrow (5)$ $((5) \rightarrow (2))$ transition; the edges $(1) \rightarrow (2)$ $((2) \rightarrow (1))$ and $(4) \rightarrow (5)$ $((5) \rightarrow (4))$ are associated with ATP binding (dissociation); the edges $(6) \rightarrow (1)$ $((1) \rightarrow (6))$ and $(3) \rightarrow (4)$ $((4) \rightarrow (3))$ are associated with ATP hydrolysis (synthesis) (The cyan arrows in $\mathcal{F}$ and $\mathcal{B}$ cycles highlight ATP hydrolysis). The model in principle accommodates 4 different pathways: (i) ATP-hydrolysis-induced forward step $(\mathcal{F}^+)$; (ii) ATP-hydrolysis-induced backward step $(\mathcal{B}^+)$; (iii) forward step that synthesizes ATP $(\mathcal{B}^-)$; (iv) backward step that synthesizes ATP $(\mathcal{F}^-)$. The reaction currents $J_{\mathcal{F}}$ flowing through $(6) \rightleftharpoons (1)$ and $J_{\mathcal{B}}$ through $(3) \rightleftharpoons (4)$ can be calculated by the generating function technique by Koza (Hwang and Hyeon, 2017; Koza, 1999), or utilizing the approach based on the large deviation theory (Lebowitz and Spohn, 1999). The expressions of $J_{\mathcal{F}}$ and $J_{\mathcal{B}}$ in terms of $\{k_{ij}\}$ for the double-cycle network are generally lengthy and too complicated to be shown here (see Eq. S25 in (Hwang and Hyeon, 2017)).

In the double-cyclic network model the total heat generated from the kinetic cycle depicted in Fig. 8b is decomposed into two contributions from the subcycles, $\dot{\mathcal{Q}}_{\mathcal{F}}$ and $\dot{\mathcal{Q}}_{\mathcal{B}}$, each of which is the product of reaction current and affinity (Barato and Seifert, 2015; Ge and Qian, 2010; Liepelt and Lipowsky, 2007a; Qian, 2004; Qian and Beard, 2005; Seifert, 2012)

$$\dot{Q} = J_{\mathcal{F}}\mathcal{A}_{\mathcal{F}} + J_{\mathcal{B}}\mathcal{A}_{\mathcal{B}}, \quad (49)$$

where the two driving forces (affinities) $\mathcal{A}$ are given by,

$$(a) \quad \mathcal{A}_{\mathcal{F}} = k_B T \log\left(\frac{k_{12}k_{25}k_{56}k_{61}}{k_{21}k_{52}k_{65}k_{16}}\right) = -\Delta\mu_{\text{hyd}} - f d_0$$

$$(b) \quad \mathcal{A}_{\mathcal{B}} = k_B T \log\left(\frac{k_{23}k_{34}k_{45}k_{52}}{k_{32}k_{43}k_{54}k_{25}}\right) = -\Delta\mu_{\text{hyd}} + f d_0$$

$$(50)$$

$$\beta\dot{Q} = \beta\left(J_{\mathcal{F}}\mathcal{A}_{\mathcal{F}} + J_{\mathcal{B}}\mathcal{A}_{\mathcal{B}}\right)$$

$$= J_{\mathcal{F}}\log\left(\frac{k_{12}k_{25}k_{56}k_{61}}{k_{21}k_{52}k_{65}k_{16}}\right) + J_{\mathcal{B}}\log\left(\frac{k_{23}k_{34}k_{45}k_{52}}{k_{32}k_{43}k_{54}k_{25}}\right)$$

$$= \underbrace{J_{\mathcal{F}}\log\left(\frac{k_{12}k_{56}k_{61}}{k_{21}k_{65}k_{16}}\right)J_{\mathcal{B}}\log\left(\frac{k_{23}k_{34}k_{45}}{k_{32}k_{43}k_{54}}\right)}_{\text{chemical}} + \tag{51}$$

$$+ \underbrace{(J_{\mathcal{F}} - J_{\mathcal{B}})\log\left(\frac{k_{25}}{k_{52}}\right)}_{\text{mechanical}} \tag{52}$$

Note that at $f = 0$, the chemical driving forces for $\mathcal{F}$ and $\mathcal{B}$ cycles are identical to be $-\Delta\mu_{\text{hyd}}$. The above decomposition of affinity associated with each cycle into the chemical driving force and the work done by the motor follows naturally from the f-dependent expression of $\{k_{ij}\}$ given in Eq.48 (Fisher and Kolomeisky, 1999a, 2001; Liepelt and Lipowsky, 2007a). The entropy production from the double cycle model is expressed as the sum of entropy produced from the two sub-cycles $\mathcal{F}$ and $\mathcal{B}$. Or simply from Eq. 49 and Eq. 50, $\dot{Q}$ can be recast into the difference between the total free energy input $[\dot{E} = (J_{\mathcal{F}} + J_{\mathcal{B}})(-\Delta\mu_{\text{hyd}})]$ and work production $[\dot{W} = (J_{\mathcal{F}} - J_{\mathcal{B}})fd_0]$:

$$\dot{Q} = J_{\mathcal{F}}\mathcal{A}_{\mathcal{F}} + J_{\mathcal{B}}\mathcal{A}_{\mathcal{B}}$$

$$= (J_{\mathcal{F}} + J_{\mathcal{B}})(-\Delta\mu_{\text{hyd}}) - (J_{\mathcal{F}} - J_{\mathcal{B}})fd_0$$

$$= \dot{E} - \dot{W}. \tag{53}$$

Notice that in order to reiterate our earlier remark that the mechanical equilibrium at the stall condition does not correspond to the thermodynamic equilibrium, the last line of $\beta\dot{Q}$ (Eq.52) is decomposed into the contributions by chemical and mechanical processes. In the proposed double-cycle kinetic model, the average velocity of the motor is given by $V = d_0(J_{\mathcal{F}} - J_{\mathcal{B}})$. Thus, if the numbers of forward and backward steps taken over time are balanced to each other, satisfying $J_{\mathcal{F}} = J_{\mathcal{B}}$ (Carter and Cross, 2005), we expect no net directional movement of the motor ($\langle x(t)\rangle = 0$), which is equivalent to say $V = 0$. It is important to recognize that displacement (or travel distance) $x(t)$ is the most direct observable to an external observer if one were to use optical tweezers or fluorescence dyes. In the stall condition ($J_{\mathcal{F}} = J_{\mathcal{B}} \equiv J_s \neq 0$), however, because the chemical processes associated with "two" ATP hydrolysis events, one along the forward and the other along the backward step, are still at work, the heat production is finite ($\dot{Q} > 0$)

$$\beta\dot{Q} = J_s\log\left(\frac{k_{12}k_{23}k_{34}k_{45}k_{56}k_{61}}{k_{21}k_{32}k_{43}k_{54}k_{65}k_{16}}\right) = 2J_s(-\Delta\mu_{\text{hyd}}). \tag{54}$$

This is a point of great importance, which cannot be capture by the uni-cyclic network model (Fig.8a).

Figure 8c depicts the currents flowing through the two sub-cycles $J_{\mathcal{F}}$ and $J_{\mathcal{B}}$ calculated from the set of rate constants $(\{k_{ij}(f)\})$ determined against the motility data of kinesin-1. Under small hindering ($f \gtrsim 0$) or assisting load ($f < 0$), kinesin-1 predominantly moves forward through the $\mathcal{F}$-cycle, whereas it starts taking more number of backsteps through the pathway $\mathcal{B}^+$ as the load increases further. At stall conditions, the net current associated with the mechanical stepping defined between the states (2) and (5) vanishes ($J = J_{\mathcal{F}} - J_{\mathcal{B}} = 0$); however, non-vanishing current along the futile cycle hydrolyzing ATP still persists along the reaction path of $(1) \rightleftharpoons (2) \rightleftharpoons (3) \rightleftharpoons (4) \rightleftharpoons (5) \rightleftharpoons (6) \rightleftharpoons (1)$ (see Figure 8b). A further increase of f beyond f_s leads to $J_{\mathcal{F}} < J_{\mathcal{B}}$, increasing the chance of backsteps via ATP hydrolysis. Although a backward step satisfying $J_{\mathcal{F}} < 0$ could be realized through an ATP synthesis, a theoretical analysis (Hyeon et al., 2009) of the experimental data (Carter and Cross, 2005; Nishiyama et al., 2002) suggest that such event ($J_{\mathcal{F}} < 0$, ATP synthesis induced backstep) is practically negligible compared with the one associated with ATP hydrolysis-induced backstep ($J_{\mathcal{B}} > 0$), so that $|J_{\mathcal{B}}| \gg |J_{\mathcal{F}}|$. This is in agreement with energetic analysis suggesting that backward and forward steps are not the reverse of each other (Hackney, 2005).

D. Additional remark on non-equilibrium nature of motors

It is worth mentioning a couple of recent studies that further highlight the non-equilibrium aspect of molecular motors. First, the Harada-Sasa equality (Harada and Sasa, 2006) quantifies the heat generated from Langevin processes out of equilibrium as follows,

$$\dot{Q} = \sum_{i=1}^{N}\gamma_i\left\{\bar{v}_i^2 + \gamma\int_{-\infty}^{\infty}[\tilde{C}_{ii}(\omega) - 2k_BT\tilde{R}'_{ii}(\omega)]\frac{d\omega}{2\pi}\right\}. \tag{55}$$

It equates the total heat dissipation rate from the system $\dot{Q}$ with the expression in terms of friction coefficient γ_i of an i-th variable x_i, the Fourier component of autocorrelation of the velocity fluctuation $C_{ii}(t) \equiv \langle(\dot{x}_i(t) - \bar{v}_i)(\dot{x}(0) - \bar{v}_i)\rangle_0$, and the real part of the response function $\tilde{R}'_{ii}(\omega)$. At equilibrium, when the detailed balance condition is satisfied, $\bar{v}_i = 0$, which leads to the standard fluctuation dissipation theorem, $\tilde{C}_{ii}(\omega) = 2k_BT\tilde{R}'_{ii}(\omega)$. Thus, the heat dissipation is zero. Therefore, the equality relates the extent of violation of the fluctuation-response relation in NESS with the rate of energy dissipation from the system into the bath. At least two studies have so far used this equality to experimentally assess the heat dissipation from molecular motors, one for kinesin and the other for F1-ATPase. Ariga (Ariga et al., 2018) have

adapted this equality in the form,

$$\dot{Q}_x = \gamma_x \left\{ \overline{v}_x^2 + \int_{-\infty}^{\infty} [\tilde{C}_{vv}(\omega) - 2k_B T \tilde{R}'_{vv}(\omega)] \frac{d\omega}{2\pi} \right\} \quad (56)$$

where the displacement of kinesin motor $x(t)$ was experimentally monitored, and the Fourier component of autocorrelation function and response function of velocity fluctuation were directly calculated using the time traces obtained from single molecule measurements. They showed that the total heat dissipation assessed using Eq.56 and the work done by the kinesin do not add up to the total chemical free energy input to the motor. The authors suggested that there are other elements of heat dissipation that do not involve the dynamics of the motor along the direction of motion $x(t)$ that can be monitored by their experiment. Although the proposed double cycle model in Fig.8b has already accommodated other possibilities, such as ATP-hydrolysis involved futile cycle without stepping, other scenarios such as slippage induced by mechanical force without chemical process (Yildiz *et al.*, 2008) are not included. Further elements of heat dissipation could be included in kinetic models in order to describe the function of the motors.

The finding by Ariga *et al.* should be contrasted with another experimental study using the Harada-Sasa equality on a single F_1-ATPase (Toyabe *et al.*, 2010). These authors showed that the total free energy input to the enzyme was partitioned to the heat and work production with little loss. This indicates that the cycle of ATP synthesis corresponds to the reversed cycle of the hydrolysis-driven motor rotation. However, a recent careful theoretical analysis by Sumi and Klumpp on F_1-ATPase (Sumi and Klumpp, 2019) suggested that the reversibility (100% efficiency) of the rotary motor is only attained under certain conditions. Mechanical slip can occur in the presence of high external torque without chemomechanical coupling, the effect of which is amplified at low ATP and ADP concentrations. This reduces the previously estimated 100 % free-energy transduction efficiency. It was argued that in addition to the viscous dissipation of the probe, heat dissipation could occur from the rotary motor itself as the torque applied to the biological nanomachine induced mechanical slip. This occurs as a result of deformation of molecular conformation, which is best optimized to function in the absence of torque.

E. Applications

1. Myosin V and Kinesin-1

SKMs have been successfully used to model the function of molecular motors. Sometime ago Leibler and Huse (Leibler and Huse, 1993) devised a model to investigate the differences between motors that act as "porters" and those that function as "rowers." The former, like the processive molecular motors described so far, work in small groups in order to transport cargoes; myosin II and axonemal dynein are examples of the latter, and work in large teams in order to slide or bend filaments.

In two landmark papers Fisher and Kolomeisky analyzed the data for kinesin-1 (Fisher and Kolomeisky, 2001) and myosin V (Kolomeisky and Fisher, 2003) in order to train the parameters in their SKMs. A 2-state and a 4-state model for kinesin-1 described accurately the motor velocity as a function of ATP and resistive load (see Fig. 9a-b); although the 2-state model was also used to describe the run-length of kinesin (see the section VI for details about the run length), the analysis of the randomness parameter as a function of [ATP] and f highlighted the need for a $N = 4$ model. In the case of myosin V, Kolomeisky and Fisher considered the data for ATP and load-dependent dwell time before a forward step (Mehta *et al.*, 1999); they fit the parameters of their 2-state model (Fig. 9c) and predicted the value of the randomness parameter as a function of [ATP] and f. The analysis of the parameters trained with the experimental data revealed in both circumstances the existence of a substep, and that most of the load dependence on the rates was carried by the reverse processes, indicating that the transition states are closer to the initial state than to the final state.

Liepelt and Lipowsky (Liepelt and Lipowsky, 2007a) devised a multi-cycle model (see Fig. 8b), which allowed them to incorporate backward steps fueled by ATP hydrolysis. The experimental results of (Carter and Cross, 2005) and (Visscher *et al.*, 1999) were recovered with a di-cyclic, $N = 6$ model (see Fig. 10), although Liepelt and Lipowsky showed that another state (and a new cycle) was necessary in order to account for the velocity as a function of ADP concentration (Schief *et al.*, 2004).

A further development was put forth by (Hyeon *et al.*, 2009), where a more sophisticated network was proposed in order to consider an alternative pathway for backward stepping based upon known structural features of the kinesin-1 dimer. (Clancy *et al.*, 2011) proposed yet another cycle for kinesin, in this case to fit the data for a mutant with an extended neck linker; the model successfully recovered the velocity, randomness parameter, and ratio between number of forward and backward steps as a function of ATP and resistive load, even in the super-stall regime.

2. Dynein – an erratic motor

In part due to its complex architecture and the paucity of details of the nucleotide chemistry much less is known about the kinematics of dynein stepping kinetics. We
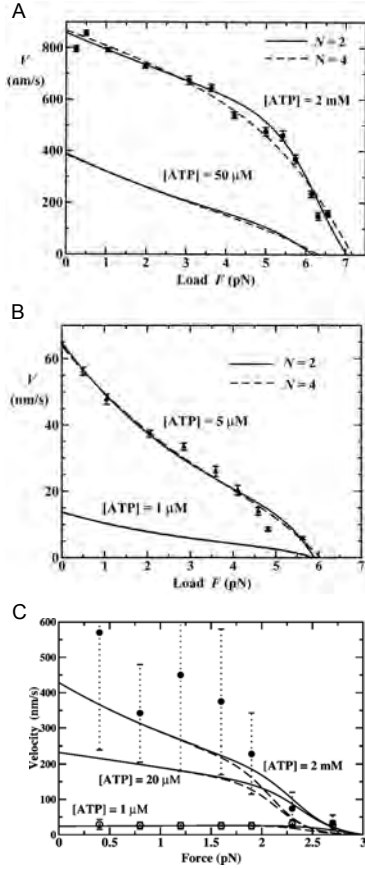
FIG. 9 Unicycle models for kinesin-1 and myosin V. Panel (a-b): comparison for kinesin-1 of the results of a uni-cycle model with experiments from (Visscher *et al.*, 1999). Panel (c): myosin V, comparison between a $N = 2$ model for myosin V and the velocity from experiments, obtained from the mean dwell times τ (Mehta *et al.*, 1999) via the relationship $v = 36\text{nm}/\tau$. Panels (a-b) and panel (c) are adapted from (Fisher and Kolomeisky, 2001) and from (Kolomeisky and Fisher, 2003), respectively.

FIG. 10 Fit of experimental data for kinesin-1 with a di-cycle, $N = 6$ model. The data in panels (a) and (b) is from (Carter and Cross, 2005), whereas the experiments considered in panels (c) and (d) are from (Visscher *et al.*, 1999). The figure is reproduced from (Liepelt and Lipowsky, 2007a).

should note that both from the perspective of structure determination (Schmidt *et al.*, 2015) and single molecule experiments (Belyy *et al.*, 2014b; DeWitt *et al.*, 2012; Reck-Peterson *et al.*, 2006a) there have been spectacular advances in recent years, which could be used to create new theories. A theoretical model different from those described in this perspective, which by necessity treated the structural and transition kinetics between pre- and post-power stroke states approximately, was introduced (Tsygankov *et al.*, 2011) to calculate the distribution of step size. Despite several untested approximation, experiments (Reck-Peterson *et al.*, 2006a) and kinetic Monte Carlo simulations of the model were in fair agreement (see Fig. 3A in (Tsygankov *et al.*, 2011)). A much more elaborate model that couples structural aspects of dynein with a model for the catalytic cycle, which is similar in spirit to the theory for myosin V (Hinczewski *et al.*, 2013), was proposed more recently (Sarlah and Vilfan, 2014).

The predictions of the model, which has a large number of parameters, were successful in obtaining the step-size distribution as well as an estimate of the stall force. However, simple theories that can account for the unusually broad step size distributions reflecting the erratic nature of this motor, force-velocity curves as a function of ATP concentration, as has been done for kinesin and myosin V, are lacking.

VI. MOLECULAR MOTORS – MODELS WITH DETACHMENT

All molecular motors take only a finite number of steps along their tracks before detaching. Therefore, prominent features of motor motility are the run-length, L, and the run-time, T, which correspond to the distance covered during a processive run, and the amount of time spent bound to the filament before detaching, respectively. In addition, the ratio between L and T constitutes an alternative and intuitive definition of velocity, $v = L/T$, which is simply the ratio between the spatial displacement of a motor and the amount of time taken to complete the movement. Processive motors take many steps along the filament before dissociating from the track. The number of steps depends on the motor, and on a number of parameters such as the concentration of nucleotides, and the value of the applied load. This underscores the importance of accounting for the end of the processive run in a way that is consistent with the enzymatic cycle of a specific motor and the experimentally

determined run-length, run-time etc.

The models described in the previous sections assume that the motor is permanently attached to the track, and ignore the end of the processive run. We discuss here a number of strategies that have been used in order to account for the detachment from the filament. (i) The most direct way is to increase the number of states by N_d, the number of detached states, and to introduce the rates of releasing and attaching to the filament. This method, which was used to model the function of kinesin-1 with $N_d = 1$ (Liepelt and Lipowsky, 2007a), enables a steady-state treatment of the corresponding Markov jump process, but requires an increase in the number of states, the number of edges, and the number of cycles. (ii) Alternatively, it is possible to enforce a new stationary state in the Markov jump process by diverting the flux into the detached state towards the filament-bound conformations. Although this approach does not require any other state ($N_d = 0$), it is necessary to increase the number of edges, which results in the creation of new cycles. A theoretical description of this steady-state approach can be found in the remarkable work by Hill (Hill, 2005a). It is unlikely that the methods described by Hill could be used to calculate velocity and run-length distribution. However, one could calculate their averages as a function of nucleotide concentration or external load. (iii) It is also possible to account for the detached trajectories by re-normalizing quantities such as the average velocity over the decreasing number of bound motors. This strategy, put forth by Kolomeisky and Fisher, is described in (Kolomeisky and Fisher, 2000), and has been used to test an approximate equation for kinesin run length (Fisher and Kolomeisky, 2001). (iv) One could define the average run length as the ratio between the average velocity v and the rate of detachment γ ($L \approx v/\gamma$). In order to do so, γ could be estimated as the product between the stationary probabilities of occupying the "vulnerable" states and newly introduced detachment rates ($\gamma = \sum_{i \in \text{vulnerable}} p_i^S k_{i,\text{det}}$) (Fisher and Kolomeisky, 2001; Maes and Van Wieren, 2003). (v) Finally, the detached state could be treated as an absorbing state. A stationary solution is not possible: after a sufficient time ($t >> \gamma^{-1}$) all the motors will be absorbed to the detached state. However, one may account for all of the possible trajectories leading to absorption, and perform averages over the ensemble of these paths. In the remainder of this section we focus on this last method.

A. Velocity Distribution

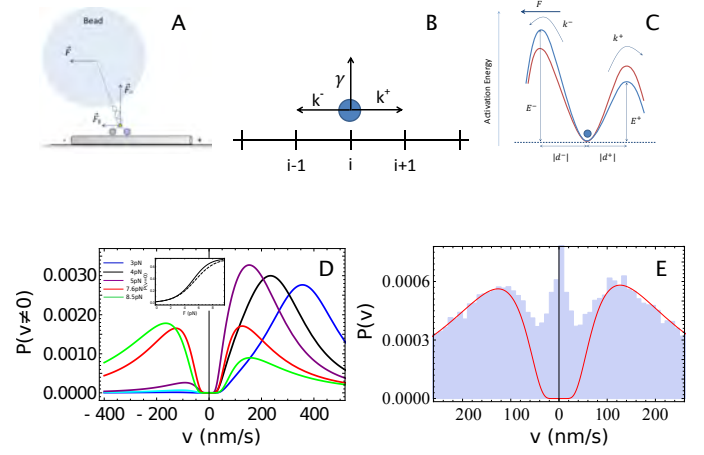The simplest model of molecular motor with detachment is shown in Fig. 11, in which forward (backward)

FIG. 11 One-state model for kinesin motility with detachment. (A) Kinesin walks along the MT, subject to an external load F. (B) The MT is represented as a linear discrete set of sites separated by $d_0 = 8.2$nm; the motor advances towards the + end with rate k^+, backsteps with rate k^-, and at every site γ establishes the detachment rate. (C) Load affects the rates according to Bell model, which in turn corresponds to a modification of the energy profile: k^+ diminishes as load increases, whereas the resistive force favors back-stepping (increases k^-). The transition state location with respect to the starting state, $\theta = \frac{|d^-|}{(|d^-|+|d^+|)}$. (D) The velocity distribution displays a markedly bimodal shape at a variety of resistive forces, with the negative peak that increases as the load approaches stall. (E) Simulations performed assuming an experimentally motivated error on the determination of d_0 maintain the bimodal structure of $P(v)$. Figure adapted from (Vu et al., 2016).

steps occur with rate k^+ (k^-), and the detachment rate is γ. For this model it is possible to determine analytically the probability distribution $p(\bar{v})$, where the velocity $\bar{v}$ is the ratio between the net number of steps taken ($n = m - l$, where m is the number of forward steps and l is the number of backward steps) divided by the run-time ($\bar{v} = n/T = v/d_0$). Vu et al. showed that (Vu et al., 2016),

$$p(\bar{v} \gtrless 0) =$$
$$= \frac{\gamma}{|\bar{v}|} \sum_{n=0}^{\infty} \left(\frac{n}{|\bar{v}|}\right)^{n+1} \frac{1}{n!} (k^{\pm} e^{-k_{\text{T}}/|\bar{v}|})^n {}_0F_1(; n+1; \frac{n^2 k^+ k^-}{|\bar{v}|^2}),$$
(57)

where ${}_0F_1(; n+1; \frac{n^2}{|\bar{v}|^2} k^+ k^-)$ is a hypergeometric function, and $k_{\text{T}} = k^+ + k^- + \gamma$. For an alternative derivation, see a more recent study (Zhang and Kolomeisky, 2018). This model was adopted to describe the stepping mechanism of kinesin-1; the parameters k^+ and k^- were established from experiments and the run length and velocity distribution (Walter et al., 2012) were fitted using only one parameter, γ. The result in Eq. 57 hold also in the presence

of force, which increases the probability of back-stepping and detaching from the microtubule. Using this model a number of noteworthy results were obtained (Vu *et al.*, 2016): (i) even at zero load, the velocity distribution is not Gaussian, although it approaches a normal distribution if the run length is large (that is, $k^+/\gamma >> 1$ and $k^+/k^- >> 1$); (ii) the probability distribution of the instantaneous velocity (the size of the step divided by the dwell time) differs from Eq. 57, and does not match the experimental distribution (Walter *et al.*, 2012), suggesting that it is not the correct way to compute velocity; (iii) most surprisingly, the velocity distribution is bimodal, with distinct peaks for $\bar{v} > 0$ and $\bar{v} < 0$. The prediction that $p(v)$ is bimodal, becoming most prominent at stall force, was unexpected. At $f = f_{\text{stall}}$, the naïve expectation would be that $p(v)$ would have a peak at $v = 0$, and the area under $p(v)$ with $v > 0$ and $v < 0$ would be the same. The bimodal distribution is a consequence of the discrete nature of kinesin step size, and it is exaggerated at large f, when the probability of taking backward steps and the probability of detaching are larger.

The model in Fig. 11 was recently extended in order to include an intermediate step (Fig. 12), which was used to describe the kinetics of a molecular motor as a combination of an ATP-dependent and an ATP-independent transition (Takaki *et al.*, 2019). The resulting velocity distribution is,

$$p(\bar{v} \gtrless 0) =$$
$$\frac{\gamma}{\bar{v}} \sum_{\substack{m,l \\ m \gtrless l}} \frac{m-l}{\bar{v}} \frac{\sqrt{\pi}}{m!l!} \frac{k^{n+1}(k^+)^m(k^-)l}{|k-(k^++k^-+\gamma)|^{n+1/2}} \left(\frac{m-l}{\bar{v}}\right)^{n+1/2}$$
$$e^{-\frac{k+k^++k^-+\gamma}{2}\frac{m-l}{\bar{v}}} I_{n+1/2}\left(\frac{|k-(k^++k^-+\gamma)|}{2}\frac{m-l}{\bar{v}}\right),$$
$$(58)$$

where I is a modified Bessel function of the first kind. Interestingly, the bimodal structure of the velocity distribution is robust to changes of the nucleotide concentration, and it is independent of which step ($1 \to 2$ or $2 \to 1$) is ATP-dependent. Therefore, an experiment aimed at testing the predicted multi-modality of $p(v)$ may be conducted under arbitrary ATP concentration and resisting load.

The model in Fig. 12 was used to tackle a vexing conundrum concerning the stepping mechanism of kinesin (Takaki *et al.*, 2019). Recently, two groups have performed similar experiments in which they monitored the movement of one kinesin head conjugated with a gold-nanoparticle (AuNP) by tracking the position of AuNP (Isojima *et al.*, 2016; Mickolajczyk *et al.*, 2015). From the analysis of the trajectories, one group proposed that ATP binds to kinesin when the trailing head is still bound to the MT (or at least in the vicinity of the pre-stepping site) (Mickolajczyk *et al.*, 2015), the other group suggested that the dissociation from the MT and forward movement of the trailing head
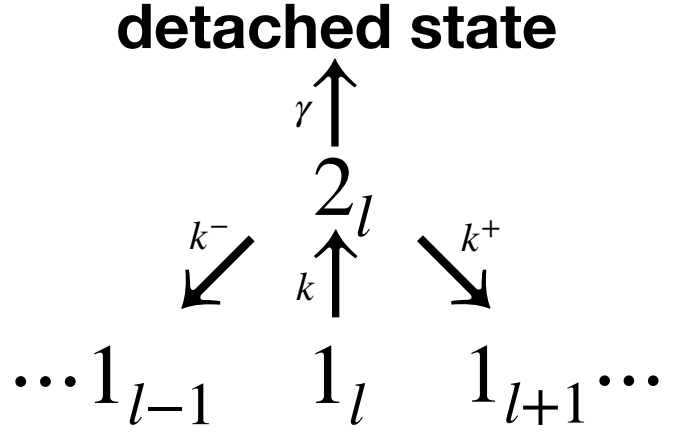

FIG. 12 Two-state model with detachment. The forward and backward stepping rates are k_+ and k_-, respectively, and they connect state 2 to state 1. The $1 \to 2$ transition occurs with rate k. Detachment occurs only from state 2, with rate γ.

detachment precede ATP binding (Isojima *et al.*, 2016). Using the model in Fig. 12, Takaki et al. (Takaki *et al.*, 2019) predicted that the ATP-dependent profile of the randomness parameter may be used in order to determine whether ATP binds to the two-head-bound or one-head-bound conformation of the kinesin-1 dimer.

B. Alternate Models with Detachment

More complicated kinetic schemes that include motor detachment have been used in order to compute the processivity and velocity of molecular motors. Elting et al. (Elting *et al.*, 2011) found an analytical solution for a 8-state model of myosin VI which they proposed in order to investigate the nature of the gating mechanism. This model did not account for the possibility of backward stepping, which was instead included by Caporizzo et al. (Caporizzo *et al.*, 2018) in a kinetic model that was devised in order to investigate the motility of myosin X as a function of the structure of the filament (individual versus bundled actin) and the geometry of the tail of the dimer (parallel versus anti-parallel). More recently, a new theoretical framework capable of extracting motility characteristics such as average velocity, average and distribution of number of steps, and probability of backward stepping for a kinetic network of arbitrary geometry has been proposed (Mugnai *et al.*, 2019). The dynamics of the motors is described in terms of the following Markov chain,

$$\vec{P}_{x+1} = (\hat{S} + \hat{F} + \hat{B})\vec{P}_x = \hat{M}\vec{P}_x, \quad (59)$$

where the N-dimensional vector $\vec{P}_x$ contains the probability of occupying any of the N filament-bound states of the motor after x transitions have occurred, and $\hat{S}$,
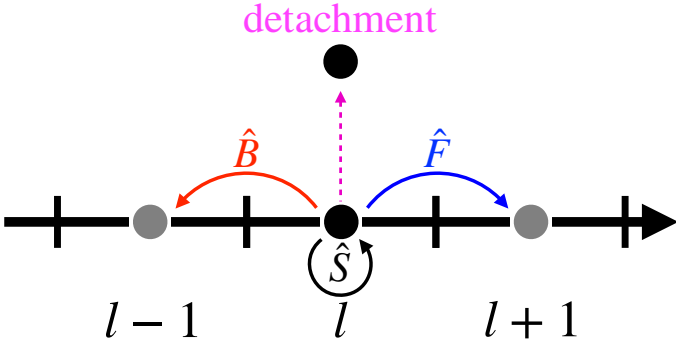
FIG. 13 Model for processive motor moving on a periodic track. From each site l of a filament the motor can change its conformation (e.g. the chemical state) within the same location (transitions included in the matrix $\hat{S}$, in black) move forward (the matrix $\hat{F}$, in red, accounts for these steps) or backward (the matrix $\hat{B}$, in blue, refers to these displacements) along the track, or detach (magenta).

$\hat{F}$, and $\hat{B}$ are the $N \times N$ transition probability matrices for undergoing a transition at a fixed location, stepping forward, or backward, respectively (see Fig. 13). Eq. 59 holds under the assumption that the track is periodic, which makes $\hat{S}$, $\hat{F}$, and $\hat{B}$ independent on the location of the motor along the filament. By accounting for all of the possible stepping pathways, it was found that the probability of taking n steps (forward or backward) before detaching is,

$$\mathbb{P}(n) = \vec{1}^{\intercal}(\hat{I} - \hat{P}_{\text{step}})(\hat{P}_{\text{step}})^n \vec{P}_0, \qquad (60)$$

where $\vec{1}^{\intercal}$ is the transpose of an N-dimensional vector of ones, $\vec{P}_0$ is the initial probability (appropriately normalized, $\vec{1}^{\intercal} \cdot \vec{P}_0 = 1$), and $\hat{P}_{\text{step}} = (\hat{F} + \hat{B})(\hat{I} - \hat{S})^{-1}$. This expression is a generalization of a well-known result for $N = 1$, in which $\mathbb{P}(n) = (1 - \pi)\pi^n$, where $\pi < 1$ is the probability of stepping and $1 - \pi$ is the probability of detaching. The average number of steps $\langle n \rangle$ may be computed using simple linear algebra, as well as the distribution and average number of forward and backward steps, $\langle m \rangle$ and $\langle l \rangle$, respectively, with $\langle m \rangle + \langle l \rangle = \langle n \rangle$. The average run length is then $\langle L \rangle = d_0(\langle m \rangle - \langle l \rangle)$, and by using a classical result for the mean-first-passage-time to absorption it is possible to compute $\langle \tau \rangle$ (Oppenheim *et al.*, 1977), and define an average velocity $v = \langle L \rangle / \langle \tau \rangle$. With Kinetic Monte Carlo simulations it is possible to show that $\langle L \rangle / \langle \tau \rangle$ differs from $\langle L/\tau \rangle$, which was adopted in (Takaki *et al.*, 2019; Vu *et al.*, 2016). On the other hand, it was shown empirically that the differences become smaller as the average number of steps taken by the motor increases (Mugnai *et al.*, 2019). The advantages of this theoretical framework are its flexibility (it works with any network) and ease of implementation (it only requires matrix algebra). Therefore, it can be used to fit complicated models against experimental data. The model was used to solve a compli-

cated model for myosin VI motility inspired by model of Yanagida and coworkers (Ikezaki *et al.*, 2012, 2013; Nishikawa *et al.*, 2010), which accounted for a motor taking both hand-over-hand and inchworm-like steps (Mugnai *et al.*, 2019), and predicted the gating mechanism [in agreement with (Dunn *et al.*, 2010; Elting *et al.*, 2011)], size of the backward steps [matching experimental observations (Altman *et al.*, 2004; Nishikawa *et al.*, 2010) within discrepancies likely related to fluctuations and the site of probe attachment], pathway for backward stepping [as suggested by (Ikezaki *et al.*, 2013)], and importance of foot-stomping in breaking the tight-coupling of myosin VI stepping.

VII. POLYMER PHYSICS-BASED APPROACHES INCORPORATING STRUCTURAL FEATURES INTO KINETIC THEORIES

Even with coarse-grained numerical models like the one proposed by (Craig and Linke, 2009), which is described below, collecting statistics from multiple simulations covering entire motor trajectories (i.e. for myosin V each run averaging tens of steps until detachment) can be computationally expensive. To more comprehensively explore how motor architecture affects stepping dynamics, particularly in light of experiments that perturb structural features like lever arm length (Oke *et al.*, 2010; Sakamoto *et al.*, 2005), alternative approaches are needed. We focus on one particular example, an analytical theory for myosin V dynamics (Hinczewski *et al.*, 2013), that also highlights several aspects discussed earlier: the starting point of the theory is a six-state stochastic kinetic network model consisting of multiple cycles that explicitly includes detachment of the myosin V dimer from the actin filament. Incorporating detachment allows us to model finite run lengths of the motor on actin, while the multi-cycle network topology captures the dominant pathways of myosin V dynamics under load forces below or near stall ($\lesssim 2 - 3$ pN): these include both forward and backward steps as well as so-called "stomps", where either the trailing or leading head detaches and then reattaches near its original binding location. Stomps are challenging to detect with single-molecule fluorescence techniques, but have been observed experimentally using high-speed atomic force microscopy (Kodera *et al.*, 2010).

However unlike the kinetic models treated so far, the force-sensitive transition rates will not be described through a phenomenological Bell-like exponential dependence as in Eq. (36). Instead, the goal of the approach by (Hinczewski *et al.*, 2013) is to model the structural mechanics underlying this force dependence through a coarse-grained polymer theory for the myosin V dimer. This replaces the discrete semiflexible polymer description of Craig and Linke (Craig and Linke, 2009) (inter-

acting monomers representing the motor head and lever arm IQ domains) with an even simpler construct: a single continuous semiflexible polymer that represents the combined motor head and lever arm domains. Thus, the dimer becomes two polymer "legs" attached at a flexible junction. The load force transmitted through the cargo domain is modeled by an effective force F applied at the junction. Because the load force changes the ensemble of conformations for the two polymer legs, and hence the three-dimensional distribution of positions explored by the unbound motor head (see Fig. 14), it affects the rates at which the unbound head reaches potential binding sites.

The most direct advantage of this simpler coarse-grained description is that the effects of force on stepping dynamics become (to an excellent approximation) analytically tractable. It is also readily generalized to more complex contexts. For example, while the model we focus on here considers only a flexible junction, backwards load forces parallel to the actin axis, and restricts steps to actin binding sites separated at actin helical half lengths (the most probable step separation for myosin V), all of these assumptions can be relaxed. A recent extension of the polymer theory approach (Hathcock et al., 2019) explores the effects of a possible structural constraint at the junction (inspired by experimental evidence (Andrecka et al., 2015)), considers off-axis forces, and incorporates the full distribution of steps at all possible actin binding sites. This allows us to understand previously observed step distributions of mutant myosins with various lever arm lengths (Oke et al., 2010; Sakamoto et al., 2005), as well as the robustness of myosin V dynamics in experiments where various off-axis load forces are applied to the motor via optically trapped cargo (Oguchi et al., 2010). Moreover the polymer approach is not limited to myosin V: an analogous treatment of dynein, with the two motor domains approximated as rigid rods (the large stiffness limit of semiflexible polymers), can successfully reproduce the complex details of dynein step distributions on microtubules (Goldtzvik et al., 2019). For both myosin V and dynein each half of the dimer is modeled as a single polymer / rod, and this description is likely to be applicable to other dimeric processive motors with fairly stiff lever arms, like myosin XI (Tominaga and Nakano, 2012). However there are cases, like myosin VI or X, where the lever arm structure is more heterogeneous, mixing stiff and flexible domains (Sun and Goldman, 2011b). Here any future attempt to apply coarse-grained polymer modeling would need to represent each lever arm as a series of polymer chains with different bending rigidities.

To understand the basic details of the polymer theory for myosin V in its simplest form, we first summarize the underlying six-state kinetic model, illustrated in Fig. 15A. State 1 is the waiting state, where both motor heads have bound ADP and are strongly attached to actin. The lever arms for both heads are in the post-powerstroke conformation: in other words in the absence of other constraints the lever arms would orient in the forward direction (toward the plus end of actin), but because the junction pulls the lever arm of the leading head backwards the entire structure is in a strained state known as the telemark or reverse arrowhead stance (Kodera et al., 2010; Walker et al., 2000). State $1'$ is also a waiting state, but with the entire motor displaced forward by one half helical length of actin ($\Delta = 36$ nm). The possible kinetic pathways are as follows:

1. *Forward step* $(1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1')$: ADP release from the trailing head is followed by ATP binding, which leads to detachment of the head from actin $(1 \rightarrow 2)$. We assume saturating ATP conditions, so ATP binding is fast compared to ADP release, and hence the entire process is subsumed into a single transition described by a rate t_{d1}^{-1}. Here t_{d1} is the mean detachment time scale, dominated by the waiting time for ADP release, $t_{d1} \sim 12$ s^{-1} (De la Cruz et al., 1999). The next transition $(2 \rightarrow 3)$ involves ATP hydrolysis into ADP+P$_i$, along with a recovery stroke where the head / lever-arm orientation changes from post- to pre-powerstroke. This transition occurs at a rate $t_h^{-1} = 750$ s^{-1} (De la Cruz et al., 1999). While there are scenarios where the reverse hydrolysis rate is significant (for example in motors with modified light chain composition (Dunn and Spudich, 2007)), for this discussion we make the typical assumption that the forward hydrolysis rate dominates. In general, while every transition arrow in Fig. 15 has an associated reverse transition in principle, here we make the simplifying assumption that the reverse rates are negligible relative to the forward rates. This assumption captures the dominant kinetic pathways of the motor (our focus here), but we would need to explicitly consider reverse rates to look at thermodynamic features of the system (like entropy production). Once ATP is hydrolyzed, the motor head has the ability to associate with actin again. The three-dimensional diffusive search can result in either binding to the forward site (36 nm ahead of the bound leg), or the original binding site. The rate at which it binds to the forward site is $(t_{fp}^+)^{-1}$, which depends on the mean first passage time t_{fp}^+ to the forward site. The dependence of t_{fp}^+ on the load force is related to how the force biases the diffusive search, and the underlying physics is discussed in more detail below. A binding to the forward site $(3 \rightarrow 4)$ is quickly followed by P$_i$ release and a powerstroke $(4 \rightarrow 1')$, which occurs with a rate t_{ps}^{-1}. We assume this step is rapid relative to the others in the network, such that $t_{ps} \ll t_{d1}, t_{fp}^+, t_h$, and hence
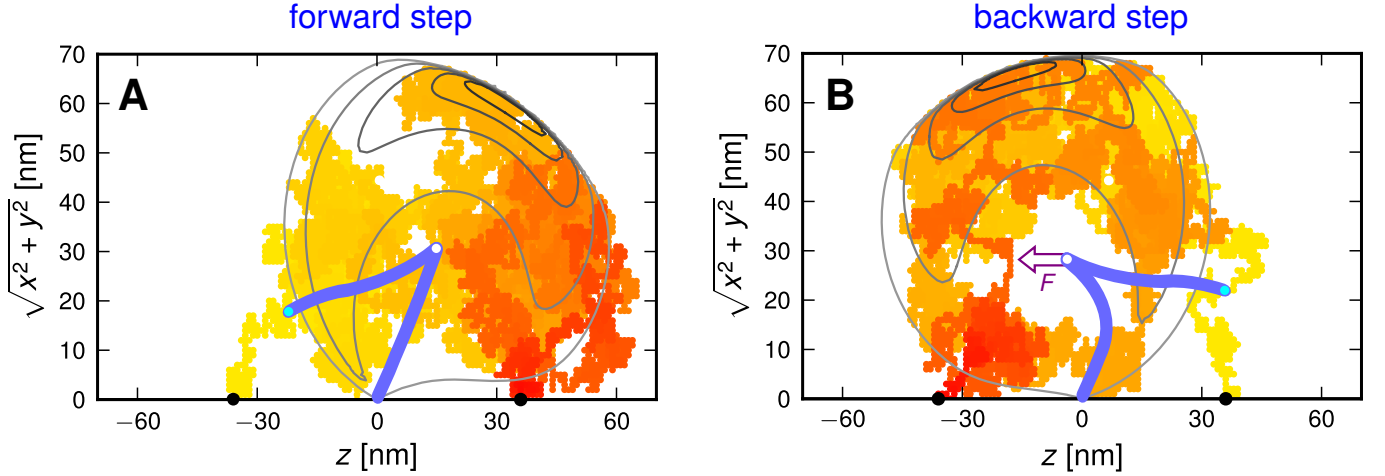
FIG. 14 Examples of the diffusive search trajectory of the detached myosin V head for two different values of backwards load force: **A.** $F = 0$; **B** $F = 2$ pN. The actin filament lies along the z axis, and the trajectories are projected onto the radial distance $\sqrt{x^2 + y^2}$ away from the actin axis. The colors represent time, with lighter colors (yellow) occurring earlier than darker colors (red). In panel A the trajectory ends in a forward step (the detached head going from the backward binding site at $z = -36$ nm to the forward binding site at $z = 36$ nm. In contrast, panel B shows a backwards step, which becomes increasingly favored as a kinetic pathway under larger load forces. Superimposed on the trajectories are cartoon snapshots of the two polymer legs (each leg representing a lever arm plus head) of the coarse-grained model for myosin V, at a time shortly after detachment. The white dot is the junction of the two legs, and the cyan dot is the location of the detached head at that time step. The contours correspond to the equilibrium probability distribution $\mathcal{P}(\mathbf{r})$ of the detached head, calculated from the analytical polymer theory (darker contours correspond to higher probabilities). As load force is increased, the attached leg and junction are pulled backwards, biasing the entire distribution away from the forward binding site, and thus increasing the chance of backwards versus forward steps (see also the corresponding experimental data of Fig. 16A). Adapted from (Hinczewski *et al.*, 2013).
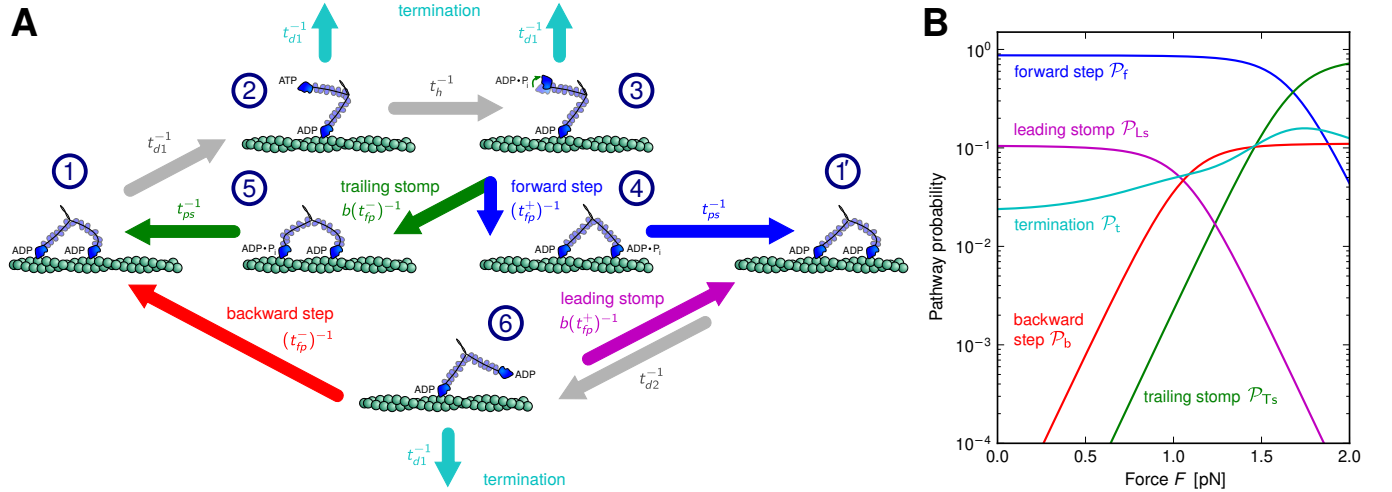
FIG. 15 **A.** Six-state kinetic model for myosin V that underlies the coarse-grained polymer analytical theory of (Hinczewski *et al.*, 2013). States 1 and 1′ are identical waiting states, with both motor heads bound to ADP and actin, except that 1′ is displaced 36 nm (a half helical length of actin) toward the plus end of the filament. Arrows are marked by inverse timescales that denote the transition rates, the details of which are described in the text. Colored arrows are transitions that belong to specific kinetic pathways: forward step (dark blue); trailing stomp (green); leading stomp (purple); and backward step (red); termination (light blue). Gray arrows are transitions that are shared between multiple pathways. **B.** Predicted pathway probabilities from the model, using the kinetic network together with mean first passage-times for the diffusive search step $t_{fp}^{\pm}$ from polymer theory. Adapted from (Hinczewski *et al.*, 2013).

t_{ps} will not enter explicitly into our estimates for motor properties below.

2. *Trailing stomp* $(1 \to 2 \to 3 \to 5 \to 1)$: This pathway starts in the same way as the forward step, with unbinding of the trailing head $(1 \to 2)$ followed by hydrolysis and the recovery stroke $(2 \to 3)$, but the diffusive search now ends at the original binding site from where the head detached $(3 \to 5)$. The mean first passage time to this site is t_{fp}^{-} (which also depends on force as described below). However the head no longer binds in the same configuration as before, because the lever arm has undergone a recovery stroke and is now in the pre-powerstroke orientation. Thus binding requires not only first passage to the site but overcoming the energetic barrier of an unfavorable geometric orientation. We effectively model this through a factor $0 < b < 1$ that reduces the rate, so that the overall rate from $3 \to 5$ is given by $b(t_{fp}^{-})^{-1}$. From fitting to experimental data, described below, the value of $b \approx 0.065$. As in the forward step, binding is followed by P_i release and a power stroke $(5 \to 1)$ occurring at a fast rate t_{ps}^{-1}.

3. *Leading stomp* $(1' \to 6 \to 1')$: This pathway is initiated when the leading head detaches with ADP still bound $(1' \to 6)$. The reason for this alternative head detachment mechanism is an asymmetry that arises from the strain in the myosin V dimer in the waiting state. The backward tension on the leading lever arm in the telemark stance (estimated to be on around 2.7 pN (Hinczewski *et al.*, 2013)) inhibits ADP release by a factor of 50-70 (Kodera *et al.*, 2010; Rosenfeld and Lee Sweeney, 2004). Rather than releasing ADP and binding ATP to detach from actin, the dominant pathway for a head under backward load is direct detachment retaining the ADP. This has been observed experimentally in single-headed myosin V, where backwards loads of around 2 pN lead to a detachment rate $t_{d2}^{-1} = 1.5$ s^{-1} (Purcell *et al.*, 2005) independent of ambient ATP and ADP concentrations. We thus take t_{d2}^{-1} as the transition rate from $1' \to 6$. The difference between the overall detachment rates for the trailing and leading heads, with t_{d1}^{-1} eight-fold larger than t_{d2}^{-1}, underlies the gating (Purcell *et al.*, 2005; Veigel *et al.*, 2005, 2002) mechanism: the trailing head is much more likely to detach first. Since detachment of the leading head occurs with the ADP still bound, the head can directly reattach back to the forward binding site, with one caveat: the post-powerstroke configuration creates an energy barrier to reattachment due to geometry (since the lever arm has to adopt a strained stance) and so as before we introduce a factor b to scale the rate to the forward site. Thus the overall reattachment rate

from $6 \to 1'$ is $b(t_{fp}^{+})^{-1}$.

4. *Backward step* $(1' \to 6 \to 1)$: This pathway starts with leading head detachment $(1' \to 6)$ like the leading stomp, but the detached head after diffusion finds the backward binding site. The post-powerstroke conformation with ADP bound is geometrically favorable for binding to this site (because the lever arm does not have to be bent) and hence the rate from $6 \to 1$ is $(t_{fp}^{-})^{-1}$. Note that this description of backward stepping is approximately valid for load forces near or below stall, but does not include additional features that may become important for larger superstall forces $(F > 3$ pN), such as power stroke reversal (Sellers and Veigel, 2010).

5. *Termination*: For the kinetic pathways described above, if the motor is in one of the three states where only one head is bound to actin (2, 3, or 6), then detachment of the second head (with rate t_{d1}^{-1}) will lead to dissociation of the entire motor from the actin filament, terminating the run.

In the kinetic network described above, mechanical forces enter in two ways: the first is through the internal strain that leads to the gating mechanism, and the second is through the external load on the junction that affects the diffusive search and hence the first passage times $t_{fp}^{\pm}$. We have already taken into account the asymmetry due to internal strain by using experimentally estimated values for t_{d1} and t_{d2}, but the force dependence of $t_{fp}^{\pm}$ has not been specified. This is precisely what the coarse-grained polymer description allows us to do. We take advantage of a separation of time scales: when either head detaches, the equilibration time t_r over which the polymer legs relax to an approximately equilibrium distribution of conformations is fast compared to the first passage times, $t_r \ll t_{fp}^{\pm}$. Brownian dynamics simulations and analytical arguments (Hinczewski *et al.*, 2013) show that $t_r \lesssim 5$ μs, while the fastest first passage times are at least $t_{fp}^{\pm} \sim \mathcal{O}(0.1$ ms). A summary of all the time scales in the problem is shown in Fig. 2. Thus to an excellent approximation the detached head fully explores some equilibrium distribution of positions, $\mathcal{P}(\mathbf{r})$, before reaching either of the binding sites. In this case the first passage time to reach the forward (+) or backward (-) binding site at position $\mathbf{r}_{\pm}$ is inversely proportional to $\mathcal{P}(\mathbf{r}_{\pm})$, the probability density of finding the detached head at that position (Hinczewski *et al.*, 2013):

$$t_{fp}^{\pm} \approx \frac{1}{4\pi D_h a \mathcal{P}(\mathbf{r}_{\pm})}, \tag{61}$$

Here D_h is the diffusion coefficient of the detached head (which can be estimated as $D_h \approx 5.7 \times 10^{-7}$ cm^2/s from the crystal structure using HYDROPRO (Ortega *et al.*,

2011)). The capture radius $a \approx 1$ nm is the distance between the head and binding site for which the interactions become strong enough that binding occurs, comparable to the Debye screening length under physiological conditions. Eq. (61), which can be derived from standard first passage time analytical approaches like the renewal method (Van Kampen, 2007), is extremely useful: it converts the dynamical problem of finding mean first passage times of a complex diffusion process into the more tractable problem of calculating the equilibrium distribution $\mathcal{P}(\mathbf{r})$ of the end-point of a two-legged semiflexible polymer system. The latter can be found analytically by extending an earlier mean-field theory for semiflexible polymers (Thirumalai and Ha, 1998), incorporating the orientational constraint of the bound leg due to the post-powerstroke conformation of the lever arm with respect to the motor head. Only a handful of structural parameters enter into the theory, determining $\mathcal{P}(\mathbf{r})$: the contour length $L = 35$ nm and persistence length $l_p \approx 310$ nm of the polymer leg; the angle of the orientational constraint with respect to the actin filament, $\theta_c \approx 60°$; and a parameter ν_c describing the strength of the orientational constraint. The first three are all known from earlier experimental estimates (Craig and Linke, 2009; Dunn and Spudich, 2007; Moore *et al.*, 2004). ν_c is along with b one of the two free parameters in the entire theoretical description, and can be estimated based on fitting to experimental data.

The full expression for $\mathcal{P}(\mathbf{r})$ and its derivation can be found in (Hinczewski *et al.*, 2013). The contours in Fig. 14 illustrate $\mathcal{P}(\mathbf{r})$ for two different values of backwards load force: A) $F = 0$ pN; B) $F = 2$ pN. Superimposed are sample trajectories of the detached endpoint, corresponding to a forward and backward step respectively, along with snapshots of the polymer conformation at a time shortly after detachment. The actin filament lies along the z-axis, with $z = 0$ corresponding to the location of attached head, and the forward/backward binding sites for the detached head at $z_{\pm} = \pm\Delta$. At zero force the peak of the distribution is at $z > 0$ due to the post-powerstroke orientational constraint on the bound leg. Thus the endpoint probabilities are biased toward the forward binding site and $\mathcal{P}(\mathbf{r}_+) \gg \mathcal{P}(\mathbf{r}_-)$. When $F = 2$ pN the situation is reversed: the load force pulls the junction backwards, counteracting the post-powerstroke constraint, and the distribution is shifted such that $\mathcal{P}(\mathbf{r}_-) \gg \mathcal{P}(\mathbf{r}_+)$. The dependence of $\mathcal{P}(\mathbf{r})$ on load force translates into corresponding changes in $t_{fp}^{\pm}$ through Eq. (61).

Once $t_{fp}^{\pm}$ as a function of F is known, a variety of physical quantities can be calculated directly from the kinetic network model (Hinczewski *et al.*, 2013). For example both forward steps and trailing stomps have the same mean duration: the average time t_{Tb} from the detachment of the trailing head to its subsequent reattachment at either the forward or backward site. Similarly both

backward steps and leading stomps have a mean duration t_{Lb} associated with how long the leading head takes to reattach. These two timescales are given by:

$$t_{Tb} = t_h + \frac{t_{fp}^+}{1 + b\alpha}, \qquad t_{Lb} = \frac{t_{fp}^+}{b + \alpha} \qquad (62)$$

where $\alpha \equiv t_{fp}^+/t_{fp}^-$. Note that t_{Tb} is bounded from below by t_h, because hydrolysis is a necessary intermediate step after trailing head detachment, whereas it is not involved after leading head detachment. The probabilities that the motor takes a forward step ($\mathcal{P}_f$), trailing stomp ($\mathcal{P}_{Ts}$), leading stomp ($\mathcal{P}_{Ls}$) and backward step ($\mathcal{P}_b$) are:

$$\mathcal{P}_f = \frac{g}{1 + g} \frac{t_{d1}^2}{(1 + b\alpha)(t_{d1} + t_h)(t_{d1} + t_{Tb} - t_h)},$$
$$\mathcal{P}_{Ts} = b\alpha\mathcal{P}_f, \quad \mathcal{P}_{Ls} = \frac{1}{1 + g} \frac{bt_{d1}}{(b + \alpha)(t_{d1} + t_{Lb})}, \qquad (63)$$
$$\mathcal{P}_b = b^{-1}\alpha\mathcal{P}_{Ls},$$

where $g \equiv t_{d2}/t_{d1} = 8$ quantifies the strength of gating. These are plotted as a function of load force for the substall regime in Fig. 15B, along with the termination probability $\mathcal{P}_t = 1 - \mathcal{P}_f - \mathcal{P}_{Ts} - \mathcal{P}_{Ls} - \mathcal{P}_b$. Forward steps predominate at small forces, but are overtaken by trailing stomps as F approaches the stall value $F_{\text{stall}} \approx 1.9$ pN, defined as when $\mathcal{P}_f = \mathcal{P}_b$. The force dependence of stomp probabilities has not yet been measured, and so constitutes a prediction of the theory, but there is experimental data on the backward-to-forward ratio $\mathcal{P}_b/\mathcal{P}_f$ (Kad *et al.*, 2008). This is shown in comparison to the theoretical curve in Fig. 16A, and exhibits excellent agreement. Other experimentally observable quantities are compared in the remaining panels of Fig. 16: B) the mean run length z_{run} along the actin before termination, from (Baker *et al.*, 2004; Clemen *et al.*, 2005; Pierobon *et al.*, 2009; Sakamoto *et al.*, 2000); C) the mean velocity $v_{\text{run}} = z_{\text{run}}/t_{\text{run}}$, where t_{run} is the mean run duration, from (Clemen *et al.*, 2005; Gebhardt *et al.*, 2006; Kad *et al.*, 2008; Mehta *et al.*, 1999; Uemura *et al.*, 2004b). The theoretical expressions for these are:

$$z_{\text{run}} = v_{\text{run}}t_{\text{run}}, \quad v_{\text{run}} \approx \frac{\Delta}{t_{d1}} \left(\frac{1}{1 + b\alpha} - \frac{\alpha}{g(b + \alpha)} \right),$$
$$t_{\text{run}} \approx \frac{gt_{d1}^2}{t_{Lb} + gt_{Tb}}. \qquad (64)$$

The theory curves in Fig. 16 are simultaneous best-fits with only two free parameters, ν_c and b, and overall show that the theory quantitatively captures the main features of the motor dynamics (within experimental uncertainties evident in the scatter of data points collected under different buffer and ATP conditions). Equally important, the theory provides insights into how motor properties depend on both kinetic parameters (the gating ratio g,
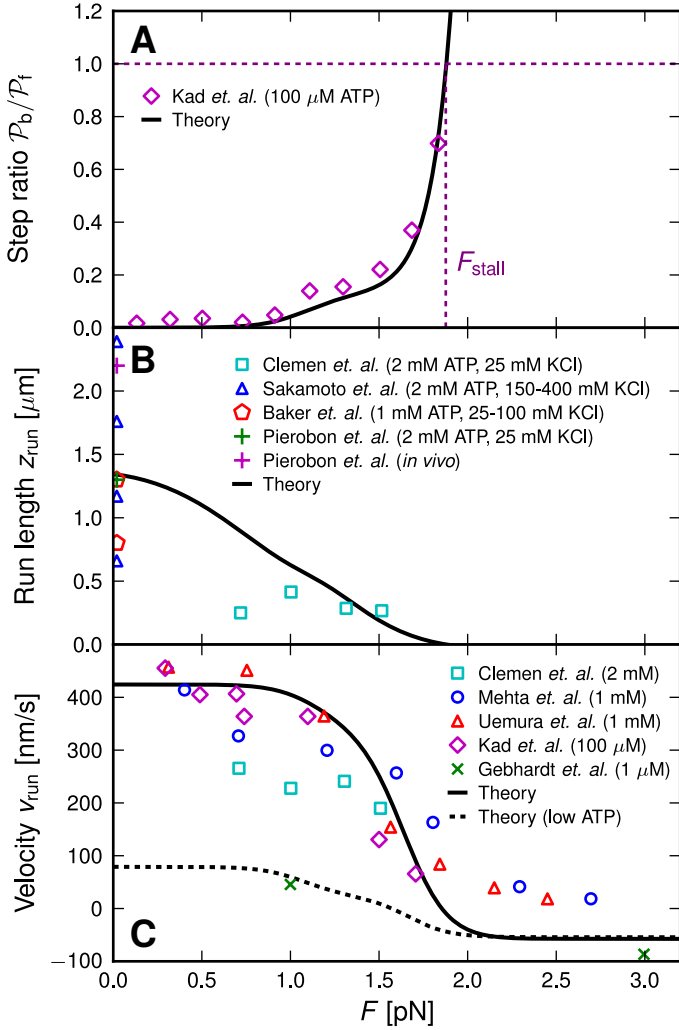
FIG. 16 Best-fit theoretical results (solid curves) from the
coarse-grained polymer theory for myosin V (Hinczewski
et al., 2013) compared to experimental results (symbols) for
the following dynamical quantities: **A.** the ratio of backward
to forward steps, $\mathcal{P}_b/\mathcal{P}_f$; **B.** the mean run length $z_{\rm run}$ on
actin before detachment; **C.** mean velocity $v_{\rm run}$. The sources
of the experimental data are listed in the legends (Baker *et
al.*, 2004; Clemen *et al.*, 2005; Gebhardt *et al.*, 2006; Kad *et
al.*, 2008; Mehta *et al.*, 1999; Pierobon *et al.*, 2009; Sakamoto
et al., 2000; Uemura *et al.*, 2004) with buffer conditions in
parentheses (first value is ATP concentration, second value is
KCl concentration). Where the KCl value is not listed, the
concentration is 25 mM. All experiments / theory are for satu-
rating ATP conditions except (Gebhardt *et al.*, 2006) in panel
C, where the theory has a modified t_{d1}^{-1} detachment rate to
account for low ATP. Adapted from (Hinczewski *et al.*, 2013).

the reduction in binding rates described by b due to un-
favorable lever arm conformations), and structural pa-
rameters (L, l_p, θ_c, and ν_c that control the distribution
during the diffusive search). This allows clear connec-
tions to mutation experiments that perturb the latter,
for example by extending or shortening lever arm length
to change L (Oke *et al.*, 2010; Sakamoto *et al.*, 2005)

(see (Hathcock *et al.*, 2019) for a fuller discussion).

VIII. SIMULATIONS USING COARSE-GRAINED MODELS

Several models that can be simulated readily (Alhad-
eff and Warshel, 2017; Goldtzvik *et al.*, 2018b; Hyeon
and Onuchic, 2007a,b; Mugnai and Thirumalai, 2017;
Mukherjee *et al.*, 2017; Mukherjee and Warshel, 2013;
Nam and Epureanu, 2016; Tehver and Thirumalai, 2010;
Zhang *et al.*, 2017; Zhang and Thirumalai, 2012) have
been introduced in order address a variety of issues re-
lated to stepping kinetics. These include but are not
restricted to the mechanism of stepping of conventional
kinesis on microtubule, gating mechanism, and allosteric
transitions by which the motor heads communicate with
each other and the cytoskeletal filaments. The simula-
tions have to be based on coarse-grained models (Hyeon
and Thirumalai, 2011) because of the long time scales
and interplay of multiple length scales involved in the
motor motility.

Rather than survey the findings in all these studies,
which vary greatly in both details and foci, we describe
a particularly illuminating minimal CG mechanochemi-
cal model (Craig and Linke, 2009) for myosin V, which
beautifully illustrates strain-mediated gating mechanism
(required for maintaining processivity) as well as char-
acteristics such as speed and stall force. The CG model
(see Fig. 17) , which takes the architecture of myosin V
in account, was constructed using the following assump-
tions. First, the level arm is treated as a semiflexible
polymer by representing the six IQ motifs by three inter-
acting moieties. In this sense the model is similar to the
subsequent analytical polymer model (Hinczewski *et al.*,
2013), described above. Second, the junction between
the head and the adjacent IQ motifs (points 2 and 8 in
Fig. 17) was treated as a semiflexible joint with equilib-
rium angles that depend on the nucleotide state of the
motor. Such an assumption is justified by comparison
to EM images. The forward rotation of the lever arm,
which changes the angle from Θ_A to Θ_B, is taken to be
dependent on phosphate release, a crucial step in the re-
action cycle of myosin V. Third, the joint between the two
lever arms is assumed to fully flexible, which would im-
ply that the tethered head diffuses freely about this joint
(see also (Hinczewski *et al.*, 2013)). It should be noted
that recent experiments suggest that this may not be the
case. It has been pointed out that the angle between the
lever arms is constrained (Andrecka *et al.*, 2015), which
has to be taken into account in describing the diffusive
search (Hathcock *et al.*, 2019). Fourth, the filamentous
actin is treated as a passive one dimensional track with
binding sites that are space $\Delta = 36$ nm apart, as was
also assumed in the analytical theory (Hinczewski *et al.*,
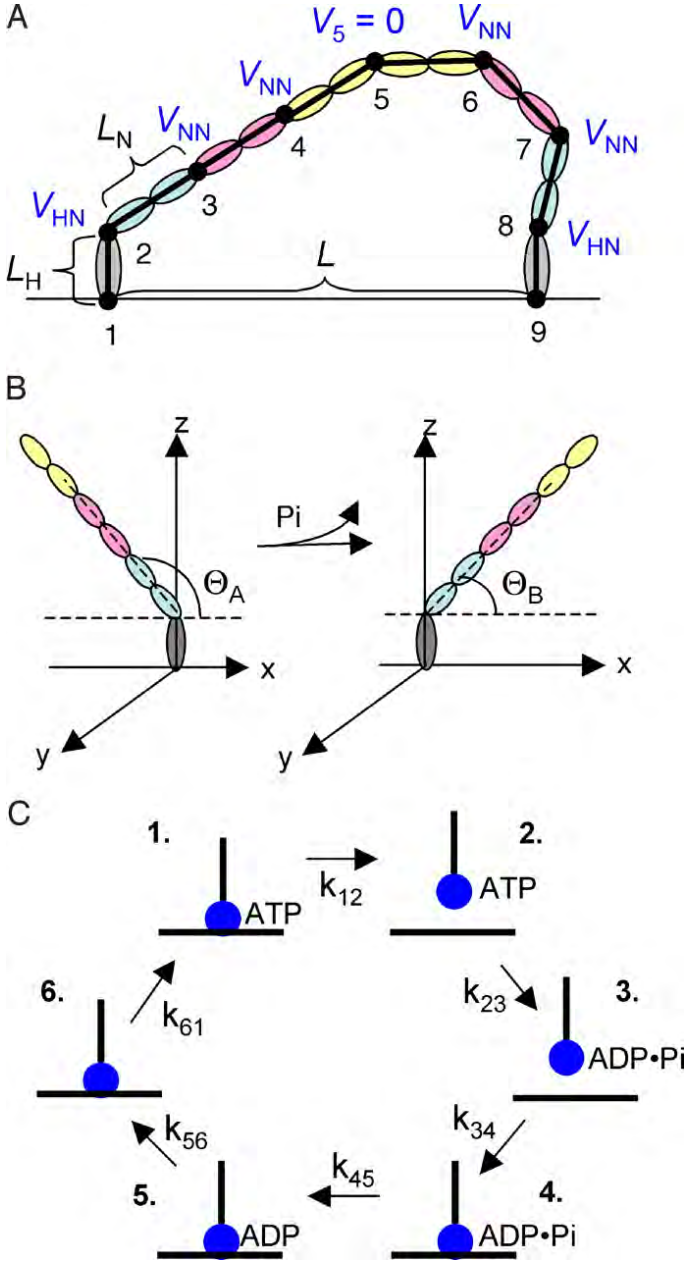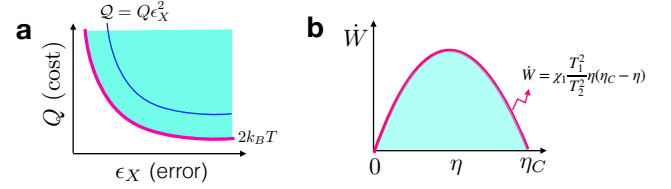2013). More importantly, if the tethered head, which has

FIG. 18 **a.** Cost-error tradeoff relation and its physical bound, $\mathcal{Q} = Q\epsilon_X^2 > 2k_BT$. Accessible region for $\mathcal{Q}$ is in cyan. **b.** Power-efficiency tradeoff. Accessible region for power ($\dot{W}$) is demarcated in cyan.

undergone hydrolysis, diffuses close to a binding site it interacts with the binding site with an attractive electrostatic interaction in order to complete a step.

A potential energy function based on the mechanical model, which can be simulated using Brownian dynamics, is coupled to the catalytic cycle in one of the motor heads. In order to produce realistic dynamics various rates in the cycle were taken from experiments (see Table 1 in (Craig and Linke, 2009)). The simulations were successful in reproducing the run length distribution and the value of the stall force ($f_S \approx 2 - 3$ pN). One of the advantages of the CG simulations is that the stiffness of the head-neck and neck-neck joints, encoded by the terms V_{HN} and V_{NN} could be changed to assess the effect on the motor properties. For example, they discovered that the value of f_S depends on the stiffness of the lever arm. It has to be stiff but not overly so in order to reproduce the experimental data, which was later confirmed theoretically (Hinczewski *et al.*, 2013).

The strategies used in the models described in this section and the previous one are the following. First, the domains that execute mechanical movements are modeled using available structural data. Second, the mechanical model is coupled to the catalytic cycle, which allows one to predict the dependence of measurable quantities as a function of control parameters such as ATP concentration and external load. The level of coarse-graining in the first step is largely guided by intuition. In (Hinczewski *et al.*, 2013) the use of polymer representation afforded analytic solution whereas by discretizing the level arm using discrete connected links Craig and Linke (Craig and Linke, 2009) had to resort to numerical simulations. It is this general strategy that is likely to be successful in tackling the nuances of dynein stepping and perhaps motor functions *in vivo*.

FIG. 17 Illustrating a general strategy to construct and simulate a coarse-grained model for myosin V (Craig and Linke, 2009). The strategy involves making a coarse-grained model based on the structure (see (A) and (B)), which is coupled to the enzyme chemistry given in (C). (A) The lever arm is represented using three rigid segments that are connected to each other. The two lever arms meet at point 5, which rotates freely during the stepping process. (B) The angle between the lever arm and the head (points 2 and 8) is assumed to change from Θ_A to Θ_B upon phosphate release. (C) The reaction cycle with various rates indicated in the figure is coupled to the mechanical model. Brownian dynamics simulations were used to calculate the observable quantities. The figure was reproduced from (Craig and Linke, 2009).

IX. COST-PRECISION TRADE-OFF AND EFFICIENCY OF MOLECULAR MOTORS

1. Cost-precision trade-off and its physical bound of molecular motors

From the perspective of thermodynamics, biological systems are clearly in non-equilibrium, which means energy is constantly injected and dissipated as heat. Because they are subject to incessant thermal and non-thermal fluctuations, cellular processes are inherently stochastic and error-prone. Thus, a plethora of energy-consuming machineries have evolved to fix any error that may be deleterious to biological functions. In the presence of large fluctuations inherent to cellular processes, harnessing energy into precise motion and suppressing the uncertainty are critical for accuracy in cellular computation. Trade-off relations between the energetic cost and information processing have been a recurring theme for many decades in biology (Alberts *et al.*, 2008; Banerjee *et al.*, 2017; Bennett, 1982; Ehrenberg and Blomberg, 1980; Hopfield, 1974; Lan *et al.*, 2012; Mehta and Schwab, 2012). Recently, a concise and fundamental relationship relating cost-precision trade-off and its physical bound, which is called the thermodynamic uncertainty relation (TUR), was first conjectured by Barato and Seifert (Barato and Seifert, 2015) and extensively studied in the statistical physics community.

Barato and Seifert (Barato and Seifert, 2015) formulated the TUR, such that a product ($\mathcal{Q}$) between the heat dissipation ($Q(t)$) and the square of relative error associated with a time-intergrated output observable $X(t)$ of the process, $\epsilon_X^2(t) = \langle \delta X^2 \rangle / \langle X \rangle^2$, is independent of measurement time. Based on numerical results that extensively sampled the rate constants k_{ij} defining the diverse kinetic networks and the linear response theory. They further conjectured that $\mathcal{Q}$ cannot be smaller than $2k_B T$ for any chemical kinetic network described by Markov jump processes, which is succinctly written as,

$$\mathcal{Q} = Q(t) \times \epsilon_X^2(t) \geq 2k_B T. \qquad (65)$$

The trade-off parameter $\mathcal{Q}$ quantifies the energetic cost for a given error and is bounded below by $2k_B T$. The time-integrated output observable $X(t)$ can be selected such that it can best represent the dynamic process of interest. For enzyme reactions that catalyze substrate to product, $X(t)$ could be the product concentration, $c(t)$. For molecular motors moving along one-dimensional track, displacement (travel distance) $x(t)$ is a natural output observable to represent their dynamic processes.

For motors, the relative error associated with the motor displacement decreases with time as $\sqrt{\langle (\delta x(t))^2 \rangle} / \langle x(t) \rangle \propto 1/\sqrt{t}$. Thus, if one were to decide the displacement of a motor precisely, a longer time trace should be generated, which demands more free energy injection (ATP hydrolysis) and heat dissipation. The greater the heat dissipated from the process, the smaller is the error. For molecular motors with output observable $x(t)$, Eq.65 can be written in terms of three quantities (heat dissipation rate $\dot{Q}$, diffusivity D, velocity V) that depend on control parameters such ATP concentration and external load.

$$\mathcal{Q} = Q(t) \times \frac{\langle \delta x(t)^2 \rangle}{\langle x(t) \rangle^2} = \dot{Q} \frac{2D}{V^2} \geq 2k_B T. \qquad (66)$$

Notice that $\mathcal{Q}$ depends on a specific type of motor as well as the conditions of [ATP] and f.

Since the Barato and Seifert's original conjecture, there has been impressive progress in the field. TUR has been reinterpreted as the inequality relation between generalized current and total entropy production rate ($\sigma^{tot} = dS/dt$), which can be written as

$$\sigma^{tot} \frac{\text{Var}(j)}{\langle j \rangle^2} \geq 2. \qquad (67)$$

General and elegant proofs for TUR have not only been given for the case of Markov jump processes on kinetic networks by employing the large deviation theory (Gingrich *et al.*, 2016), but also can be deduced from the equality relation for the Fano factor of entropy production for over-damped Langevin processes (Pigolotti *et al.*, 2017).

More recently, Dechant and Sasa (Dechant and Sasa, 2018b) have generalized Eq.67 to underdamped processes in the following form

$$\sigma^{tot} \geq B \langle j \rangle^2. \qquad (68)$$

where $B(> 0)$ is a model dependent parameter, which can be reduced to $2/\text{Var}(j)$ for over-damped Langevin systems or Markov jump processes on networks. In fact, right hand side of the inequality is always greater than zero, namely, it recovers the second law of thermodynamics $\sigma^{tot} \geq 0$. Therefore, one interpretation of Eq.68 is that it provides tighter bound to the entropy production in terms of the square of generalized current. Although B is not specific but model dependent, Dechant *et al.*(Dechant and Sasa, 2018b) employed the above relation to derive a power-efficiency trade-off for engine operating between two heat sources with temperature $T_1 > T_2$.

$$\dot{W} \leq \chi_1 \frac{T_1^2}{T_2^2} \eta(\eta_C - \eta). \qquad (69)$$

where $\eta = \dot{W}/\dot{Q}_1$ and $\eta_C = 1 - T_2/T_1$ are the thermodynamic efficiency and Carnot efficiency. The inequality gives the upper bound to the power generated using the two heat sources. But, when η approaches to η_C, $\dot{W}$ approaches to 0 as well if the model-dependent parameter χ_1 is finite.

For pedagogical purpose, it is worthwhile to consider simple examples that can demonstrate the significance of the physical bound of TUR.

(i) In the context of the foregoing one-state hopping model to which dynamics of molecular motors can be mapped, the rate of heat dissipation from the process is bounded as,

$$\frac{\dot{Q}}{k_B T} = (u - w) \log \frac{u}{w} \geq \frac{2(u - w)^2}{(u + w)} \qquad (70)$$

where the inequality was discussed in Ref. (Shiraishi *et al.*, 2016). Given a single step displacement d_0, the velocity and diffusivity of the particle moving along the reaction coordinate is $V = d_0(u - w)$ and $D = (d_0^2/2)(u + w)$, respectively. Therefore, inserting the expressions of $\dot{Q}$, V, and D to Eq.66 it is easy to see that $\mathcal{Q} \geq 2k_B T$. The lower bound of $\mathcal{Q}$ in this model is attained when $u = w$, which corresponds to the detailed balance (equilibrium) condition. Provided that $u = u^*[S]$ and $v = v^*[P]$, namely, the forward and backward rate constants vary with the substrate (S) and product (P) concentrations, the detailed balance condition is attained when $[S]/[P] = [S]_{eq}/[P]_{eq}$ with $u^*/v^* = [P]_{eq}/[S]_{eq}$.

(ii) Hyeon *et al.* (Hyeon and Hwang, 2017) studied the over-damped Lagevin motion on tilted washboard potential, which obeys

$$\gamma \dot{x}(t) = f - U'(x) + \xi(t) \qquad (71)$$

with $U(x + L) = U(x)$ and $\langle \xi(t) \rangle = 0$, $\langle \xi(t) \xi(t') \rangle = 2\gamma k_B T \delta(t - t')$, where the driving of quasi-particle on the potential is controlled by the non-conservative force f. They showed that TUR parameter $\mathcal{Q}(f)$ for this problem is a non-monotonic function of f, attaining its physical bound $2k_B T$ at both $f \ll |U'(x)|$ (near equilibrium) and $f \gg |U'(x)|$ (far from equilibrium). For $f \ll |U'(x)|$, the quasi-particle undergoes diffusion on a rough surface. On the other hand, for $f \gg |U'(x)|$ the particle slides along a smooth gradient, without feeling the effect of confining potential, dissipating energy with a rate $\dot{Q} = \gamma V^2$ against friction. Since the diffusivity of particle in this case follows the Stokes-Einstein $D = k_B T/\gamma$, together with the velocity V one can obtain $\mathcal{Q} = 2k_B T$. $\mathcal{Q}(f)$ reaches its maximum value near the critical point where the potential barrier confining the particle is about to vanish.

(iii) As far as the specific models are concerned as in (i) and (ii), the minimal uncertainty condition $(\mathcal{Q}_{\min} = 2 k_B T)$ is attained not only under the detailed balance condition but when there is no confining potential. It has been suggested that $\mathcal{Q}_{\min}$ is attained when the heat dissipated from the process is normally distributed as $P(Q) \sim e^{-(Q - \langle Q \rangle)^2/(2\langle \delta Q^2 \rangle)}$ and that TUR measures the deviation of heat distribution from Gaussianity (Hyeon and Hwang, 2017).

2. Transport Efficiency

There are a number of ways to assess the "efficiency" of engines or machines (Brown and Sivak, 2017). Historically, the efficiency of heat engines has been discussed in terms of the thermodynamic efficiency, the aim of which is to maximize the amount of work extracted from two heat sources with different temperatures (Callen, 1985). For non-equilibrium machines driven by chemical forces that are constantly regulated without shortage in the live cell, the power production could be a more pertinent quantity to maximize. Meanwhile, for transport motors in the cell, the TUR parameter $\mathcal{Q}$ can be used to assess the efficiency of suppressing the uncertainty in dynamical process by means of energy consumption, and thus is quite pertinent for evaluating the transport efficiency of a motor (or motors) (Dechant and Sasa, 2018a). The connection between $\mathcal{Q}$ and the transport efficiency is clear. If a motor transports cargos at a high speed $(V \sim \langle x(t) \rangle/t)$ with small fluctuations $(D \sim \langle \delta x(t)^2 \rangle/t)$ (which leads to punctual delivery to a target site) but consuming only a small amount of energy $(\dot{Q})$, such a motor would be considered efficient for cargo transport. A motor efficient in the cargo transport would be characterized by a small $\mathcal{Q}$ with its minimal bound $2k_B T$, or as originally suggested by Dechant and Sasa (Dechant and Sasa, 2018a), one can consider using the definition $\eta_T = 2k_B T/\mathcal{Q}$ which is bounded between 0 and 1.

$\mathcal{Q}$ can be used to assess the "transport efficiency" of biological nanomachines and to study how it changes with varying conditions of f and [ATP]. To evaluate $\mathcal{Q}$, measurement should be first carried out for $\dot{Q}$, D, and V (see Eq.66). While V and D are straightforward to calculate $(V = \lim_{t \to \infty} dx(t)/dt$ and $D = \lim_{t \to \infty} (1/2) d(\delta x(t))^2/dt)$, experimental measurement of $\dot{Q}$ may be nontrivial. Although there are some reports on direct measurements of heat dissipation rate at single cell level (Rodenfels *et al.*, 2019; Song *et al.*, 2019), direct measurement of heat dissipation at single molecule level is not yet known. Nevertheless, $\dot{Q}$ be estimated by considering a *physically suitable minimal* kinetic network model. For a given cyclic kinetic network defined by multiple chemical states $(i = 1, 2, \ldots N)$ and transition rates $(\{k_{ij}\})$ connecting them, there are straightforward methods (Hwang and Hyeon, 2017; Koza, 1999; Lebowitz and Spohn, 1999) to associate the measured V and D with $\{k_{ij}\}$. As long as all the rate constants $\{k_{ij}\}$ defining the kinetic network are known, it is then straightforward to calculate $\dot{Q}$ value as discussed in details in the early part of this review.

For conventional kinesin, whose chemomechanical properties were extensively studied by several groups, motility data for varying [ATP] and load conditions are available in the literature. The double-cycle network model, depicted in Fig.8b, with the form of rate constant for each edge can be used to analyze those data,
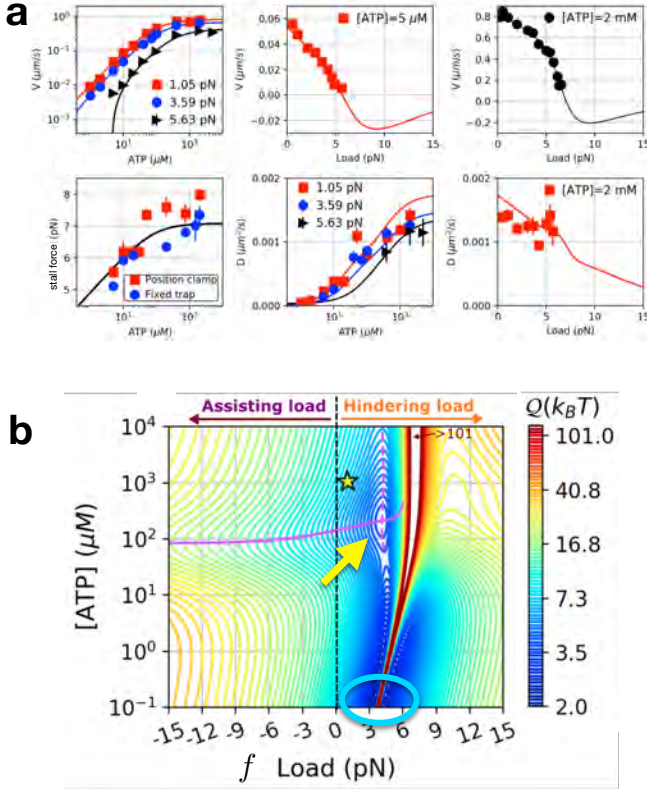
FIG. 19 **a**. Analysis of motility data to determine the parameters for rate constants ($\{k_{ij}\}$ in Eq.48) for double cycle network model. **b**. Diagram of $\mathcal{Q}$ as a function of [ATP] and f.

which allows one to determine the dependence of $\{k_{ij}\}$ on [ATP] and f and finally to build a diagram of $\mathcal{Q}(f, [\text{ATP}])$ as shown in Fig.19. The TUR diagram, $\mathcal{Q}([ATP], f)$ (Fig.19), exhibit several features that are worthwhile to explore:

(i) In terms of the load direction, assisting ($f < 0$) and hindering ($f > 0$), $\mathcal{Q}([ATP], f)$ is asymmetric. This result differs from that of the one-state hopping example for TUR calculated in Fig.20a. The fundamental difference arises from the asymmetric effect of the force on the kinetic rate, particularly on k_{25} and k_{52} with $\theta \neq 1/2$.

(ii) $\mathcal{Q}$ is locally minimized first at a condition of small load ($f \gtrsim 0$) and low [ATP] (indicated by cyan ellipse), and second at $f \approx 3$ pN and [ATP] ≈ 300 μM (indicated by the yellow arrow). The first minimum is closer to 2 $k_B T$ bound; yet, this minimum was attained near the detailed balance condition where [ATP] concentration is small and balanced with [ADP] and [P_i]. It does not have much of biological relevance given that molecular motor works out of equilibrium. In fact, the second local minimum is of particular interest given that the condition is not so far from the cellular condition [ATP] ≈ 1 mM and $f \approx 1$ pN, indicated by yellow arrow (Here, note that $f \approx 1$ pN is a rough estimate of cellular environment
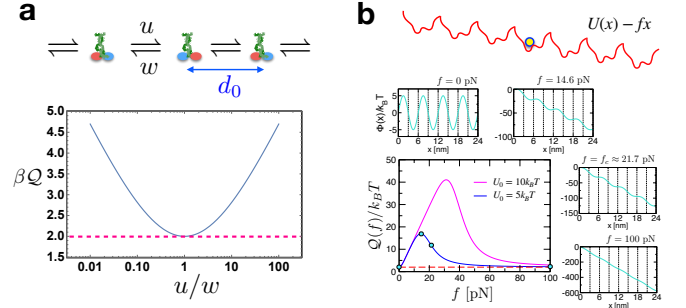

FIG. 20 Two case studies for TUR. **a**. One-state hopping model with forward and backward rate constants u and w. **b**. Brownian motion in tilted wash-board potential. $\mathcal{Q}(f)$ was evaluated as a function f using a specific periodic function, $U(x) = U_0 \sin 2\pi x/L$ with $L = 6$ nm for $U_0 = 5$ and 10 $k_B T$.

replete with obstacles such as cytoskeletal filaments and road blocks). The second local minimum is the very point at which the transport efficiency defined in terms of $\mathcal{Q}$ is optimized.

(iii) The high $\mathcal{Q}$ region ($\mathcal{Q} > 100$ $k_B T$) at $f \approx 3 - 7$ pN is due to the stall condition. As explained in detail in rationalizing the double-cycle network model, energy should be still consumed and heat should be dissipated ($\dot{Q} > 0$) at stall condition ($V \approx 0$). This particular condition renders $\mathcal{Q}$ divergent at stall conditions.

Finally, the structure of $\mathcal{Q}(f, [\text{ATP}])$ is sensitive to the design of motor structure as well as the motor type. First, the diagram of $\mathcal{Q}(f, [\text{ATP}])$ for a kinesin construct, Kin6AA, whose necklinker is engineered longer than that of wild-type kinesin-1 via insertion of six amino-acids (AEQKLT) (Clancy *et al.*, 2011) exhibits great deviation from that of the WT (Fig.21a). In all, the values of $\mathcal{Q}$ were increased, the stall forces were reduced, and the local minimum observed in the WT is missing. KIF17 and KIF3AB (Milic *et al.*, 2017) show qualitatively similar structure of $\mathcal{Q}(f, [\text{ATP}])$; however there are different in terms of quantitative details from that of WT. Next, myosin V, dynein, and F_1-ATPase were analyzed to calculate $\mathcal{Q}(f, [\text{ATP}])$. Bierbaum and Lipowsky (Bierbaum and Lipowsky, 2011) employed a tri-cyclic network model to describe the chemomechanics of myosin V in which cycles for forward steps, energy-consuming futile steps, and force-induced mechanical slippage steps were considered. The resulting $\mathcal{Q}(f, [\text{ATP}])$ (Fig.21d) shows ATP-insensitive stall condition where $\mathcal{Q}$ is divergent with no local minimum as in WT kinesin-1. For dynein (Fig.21e), the uni-cyclic chemomechanical network model adopted for construction of $\mathcal{Q}(f, [\text{ATP}])$ is in principle not satisfactory, since it would be able to account for the physically correct behavior of dynein dynamics at stall and superstall conditions. In particular, $\mathcal{Q}(f, [\text{ATP}])$ diagram shows minimization to $2k_B T$ at the stall. Nevertheless, a suboptimal local minimum is found at $f \approx 3$ pN and
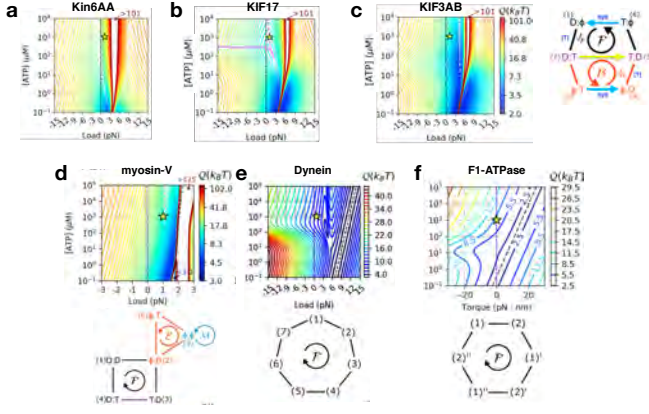
FIG. 21 $\mathcal{Q}(f, [\text{ATP}])$ for **a**. Kin6AA (kinesin-1 mutant); **b**. KIF17; **c**. KIF3AB; **d**. myosin V; **e**. Dynein; **f**. F$_1$ATPase. The diagrams were calculated by directly analyzing the motility data available in (Clancy *et al.*, 2011) for **a**, (Milic *et al.*, 2017) for **b**, **c**, and using the same double-cycle network model as that of kinesin-1. For other motors in **d**, **e**, and **f**, the kinetic network and rate constants were used as in the Ref. (Bierbaum and Lipowsky, 2011) for myosin V, Ref. (Šarlah and Vilfan, 2014) for dynein, and Ref.(Gerritsma and Gaspard, 2010) for F$_1$-ATPase.

$[\text{ATP}] \approx 300 \; \mu M$. Finally, $\mathcal{Q}(\tau, [\text{ATP}])$ (torque τ instead of load f) for F$_1$ATPase is shown in Fig.21f calculated based on uni-cyclic kinetic network model (Gerritsma and Gaspard, 2010). For F$_1$-ATPase, which shows near-reversibility and hence characterized with high efficiency (Toyabe *et al.*, 2010), the use of uni-cyclic network model would be reasonable although this conclusion should be reached with care (see (Sumi and Klumpp, 2019)).

Some of the biological motors, kinesin family and dynein, studied here are found to be semi-optimized in terms of $\mathcal{Q}$ under the cellular condition, which alludes to the role of evolutionary pressure that has shaped the present forms of molecular motors in the cell. In addition, the efficiency quantified in terms of $\mathcal{Q}$ for various molecular machines presented here are ranged between 7 and 20 $k_B T$ (Fig.22). Given that all these machines function out-of-equilibrium, it is of great interest to discover that the value of $\mathcal{Q}$ calculated at the working cellular condition is not significantly far from its physical bound 2 $k_B T$. This may arise from the fact that the molecular machine analyzed here are tightly coupled machine, meaning that ATP hydrolysis is almost always transduced to a machanical step even though energy-wasting futile steps still remain as a possibility.

As long as the underlying mechanism, which offers a clear model for chemomechanical kinetic network, is known, TUR can be studied for any time trace generated from cyclic process at NESS. Other energy-intensive processes, for example, error-correction processes, circadian cycle, and chaperonin action, would exhibit high $\mathcal{Q}$.
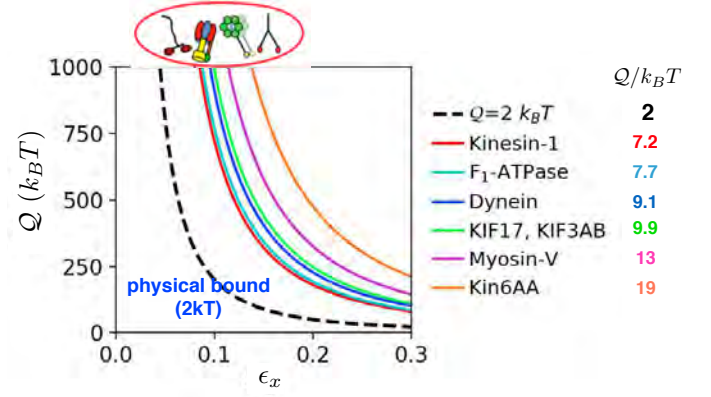

FIG. 22 $\mathcal{Q}$ values for various transport motors calculated at $[\text{ATP}] = 1$ mM and $f = 1$ pN (except for F$_1$-ATPase calculated at $[\text{ATP}] = 1$ mM and $\tau = 0$ pN·nm)

X. MOLECULAR CHAPERONES

Chaperones are another class of cellular molecular machines that expend free energy change associated with ATP binding and catalysis to facilitate the folding of certain proteins and RNA that cannot fold spontaneously under cellular conditions. Because they serve a key function in maintaining protein homeostasis they are deemed as essential for the survival of the organisms. Among a large class of molecular chaperones, the function of the heat-shock GroEL-GroES machinery found in *E.Coli*, referred to as chaperonin, is now quantitatively understood thanks to experimental and theoretical advances. The *in vivo* function of GroES-GroEL machinery in *E. Coli.* is to rescue substrate proteins that are otherwise destined for aggregation. Just like molecular motors, GroEL undergoes a catalytic cycle involving a series of large scale structural changes in response to ATP binding, hydrolysis, and release of ADP and phosphate. GroEL/GroES machine anneals the population of misfolded proteins driving them to the folded state by repeatedly going through rounds of the catalytic cycle, which has been referred to as the Iterative Annealing Mechanism (IAM) (Tehver and Thirumalai, 2008; Todd *et al.*, 1996).

A minimal kinetic network model of chaperonin-assisted protein folding, can be constructed, based on considerable experimental evidence, by assuming that a protein exists either in an intermediate (I), misfolded (M), or folded (native) (N) state as illustrated in Fig.23. The unfolded state, a transient state right after the protein is synthesized, collapses to the I state, and it further undergoes a spontaneous folding process via the kinetic pathways with rates denoted by k_{IM} and k_{IN}. Only a fraction (Φ) of the entire population folds correctly to the N state and the remaining fraction ($1 - \Phi$) is misfolded to the M state. The process of the initial population of molecules with Φ $(1 - \Phi)$ reaching the folded (mis-
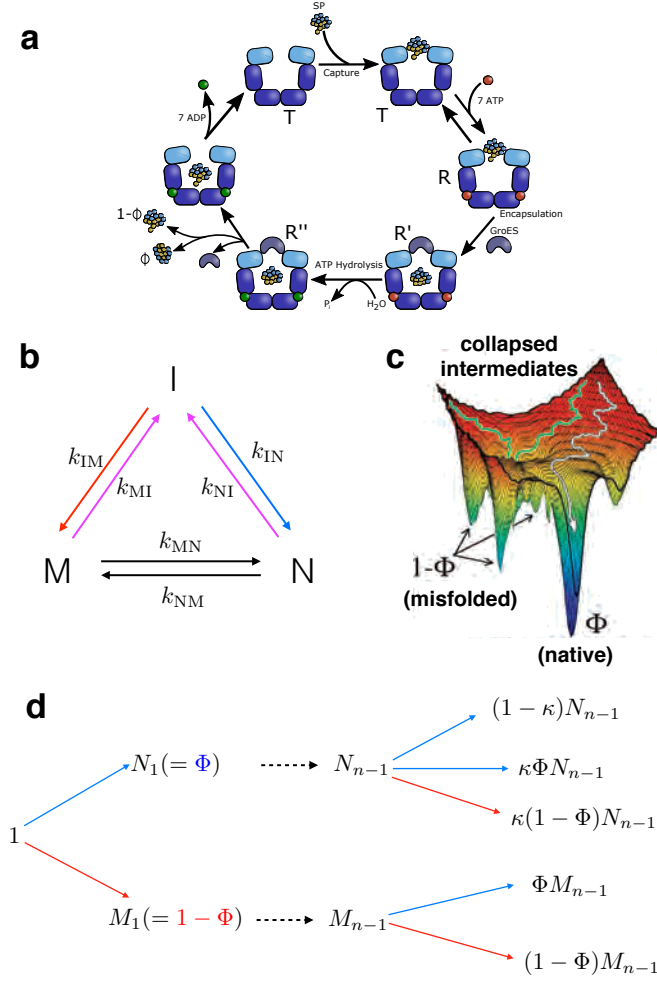
FIG. 23 Chaperone-assisted folding of substrate. **a**. Schematic of the iterative annealing mechanism (IAM) illustrated for the hemicycle of GroEL. The $T \rightleftharpoons R$ transition starts when ATP and the substrate protein (SP) bind. GroES binding and ATP hydrolysis engenders the $R' \rightarrow R''$ transition with a fraction of the SP partitioning to the folded structure. Subsequently ADP and Pi and the SP (folded or not) are released and the $R'' \rightarrow T$ transition completes the cycle. In the presence of the SP the machine turns over in about a second and ADP release is accelerated by about a hundred fold. The rapid turnover comports well with the predictions of IAM (Todd *et al.*, 1996). **b**. A minimal kinetic network model for chaperone-assisted folding of a substrate molecule. Transitions between three manifolds of collapsed intermediate (I), misfolded states (M), and native state (N) are represented in terms of rate constants. Folding, misfolding, and chaperone assisted unfolding pathways are depicted in blue, red, and magenta, respectively. **c**. Schematic of folding landscape of substrate molecules. Upon spontaneous folding, the ensemble of intermediate states collapsed from unfolded ensemble reach the native and misfolded basins of attractions with the proportion of Φ and $1-\Phi$, respectively. **d**. Schematic of the generalized iterative annealing mechanism of chaperone-assisted substrate folding, from which the recursion relation for the native yield from n-th annealing process was derived. N_i and M_i denote the proportion of native and misfolded states from the i-th annealing process. The blue and red arrows represent the pathways leading to the native and misfolded states, respectively.

folded) state is termed the Kinetic Partitioning Mechanism (KPM). The KPM, which was first theoretically proposed (Guo and Thirumalai, 1995), has been used to explain the folding of of proteins (Kiefhaber, 1995) as well as RNA (Pan *et al.*, 1997; Thirumalai and Hyeon, 2005). Typically, substrate proteins (SPs or ribozymes) that require assistance from chaperonin action are characterized by extremely small values of Φ ($\ll 1$), which implies that the majority of population without chaperones are trapped in the misfolded states, which could potentially aggregate unless rescued by the chaperones.

The function of the GroEL machine is quantitatively explained by the Iterative Annealing Mechanism (IAM) according to which the chaperone recognizes and acts on misfolded substrate preferentially. Typically, the SPs have exposed hydrophobic residues, which make them prone to aggregation unless recognized by the GroEL-GoES machine. When the SPs bind to GroEL they become disordered as a result of domain movements in GroEL, which imparts a mechanical force (≈ 10pN) that is sufficiently large to unfold (at least partially) the SPs (Thirumalai and Lorimer, 2001). As the catalytic cycle proceeds the SPs are encapsulated for a brief period (roughly about two seconds) during which a small fraction folds rapidly during the time they are in the cavity. It is worth remarking that if they fold (the probability being Φ), they do so while being encapsulated in the cavity of GroEL. When the catalytic cycle is complete, the SP is ejected from the cavity regardless of whether it is folded or not. If the SP is misfolded then it once again recognized by GroEL and the cycle is iterated until sufficient yield of the folded state is obtained. A key requirement of GroEL-SP interaction is that hydrophobic residues of the SPs must be exposed, which does not typically occur in folded states. It this ability of chaperone not to recognize native proteins that enables the GroEL-GroES machinery to drive the misfolded states to the native state over repeated iterations of the catalytic cycle. More specifically, when the annealing process, corresponding to one catalytic cycle, is iterated n times, the total amount of population that reaches the native state (or native yield) grows as,

$$N_n = 1 - (1 - \Phi)^n \qquad (72)$$

with $n \geq 1$. As $n \rightarrow \infty$, $N_n \rightarrow 1$.

While it is known that GroEL only recognizes misfolded SPs, the IAM concept has to be generalized to RNA chaperones, which act on the both N and M states, but more favorably on M (Bhaskaran and Russell, 2007; Chakrabarti *et al.*, 2017). If the proportion of N identified by chaperones in comparison with the M state is defined as κ ($0 \leq \kappa \leq 1$), the total yield of native state after n rounds of folding (annealing) in the presence of chaperone can be calculated using the following mathematical formulation:

(i) Let N_n and M_n be the yields of native and misfolded state, respectively, after the n-th round of the annealing process. Note that the total amount of substrate proteins is conserved at all times, which implies that $N_n + M_n = 1$ for all n. In the first round of annealing, a fraction $N_1 (= \Phi)$ folds to the N state and $M_1 (= 1 - \Phi)$ partitions to the M state;

(ii) In the n-th round of annealing, chaperones recognize N_{n-1} and M_{n-1} differentially by the factor κ. Whereas $(1 - \kappa)N_{n-1}$ is left unrecognized by chaperones, κN_{n-1} is unfolded and Φ of them refold to yield $\kappa \Phi N_{n-1}$ native states and $\kappa(1-\Phi)N_{n-1}$ misfolded states. On the other hand, the entire population of M_{n-1} is unfolded and Φ of them refold to yield ΦM_{n-1} native states and $(1 - \Phi)M_{n-1}$ misfolded states. Therefore, after the $(n-1)$-th round of chaperone action, the native yield N_n is determined as $N_n = 1 - M_n = 1 - (1-\Phi)M_{n-1} - \kappa(1-\Phi)N_{n-1}$;

(iii) From the resulting recursion relation of $N_n = (1-\kappa)(1-\Phi)N_{n-1} + \Phi$ with $N_1 = \Phi$, we obtain the following expression,

$$N_n = \Phi \frac{1 - (1-\kappa)^n (1-\Phi)^n}{\kappa + (1-\kappa)\Phi}, \qquad (73)$$

native yield after n iterations. After a sufficient number of iterations ($n \to \infty$), the system reaches steady state, $N_\infty \to \Phi/(\kappa + (1-\kappa)\Phi)$. In the case of GroEL only the misfolded state is recognized by chaperones ($\kappa = 0$). Therefore, the GroEL-GroES machinery drives the entire population of substrate proteins to the native state $N_\infty \to 1$ (Eq.72).

As described for molecular motors, the thermodynamic aspects of chaperonin-assisted folding can be succinctly captured by mapping the folding process of substrate molecule onto a kinetic network model. A uni-cyclic reversible kinetic network model consisting of the above-mentioned three states (I, M, and N) suffices to capture the non-equilibrium nature of chaperone-assisted protein folding. The relevant master Equation describing the network is,

$$\partial_t \mathbf{P}(t) = \mathcal{W}\mathbf{P}(t) \qquad (74)$$

where $\mathbf{P}(t) = (P_I(t), P_N(t), P_M(t))^T$ and,

$$\mathcal{W} = \begin{pmatrix} -(k_{IN} + k_{IM}) & k_{NI} & k_{MI} \\ k_{IN} & -(k_{NI} + k_{NM}) & k_{MN} \\ k_{IM} & k_{NM} & -(k_{MI} + k_{MN}) \end{pmatrix}$$

The probability $P_i(t)$ of each state i evolves with time as,

$$\mathbf{P}(t) = \mathbf{P}^{ss} + c_1 \vec{u}_1 e^{-\lambda_1 t} + c_2 \vec{u}_2 e^{-\lambda_2 t} \qquad (75)$$

with $\lambda_2 > \lambda_1 > 0$, and it reaches the steady state value P_i^{ss} at $t \to \infty$. The steady state population P_i^{ss} and steady state current along the reaction cycle

($J = k_{ij}P_i^{ss} - k_{ji}P_j^{ss}$) can be expressed solely using the rate constants, as already discussed in depth in the foregoing section describing the general aspects of the various cycles. Because the function of the chaperones is to promote the formation of the folded state our primary interest is the steady state yield of the native state, which is given by,

$$P_N^{ss} = \frac{k_{MI}([C],[T])k_{IN} + k_{MN}(k_{IM} + k_{IN})}{\Sigma([C],[T])}, \qquad (76)$$

where

$$\Sigma([C],[T]) =$$
$$k_{MI}([C],[T])k_{NI}([C],[T])$$
$$+ k_{NI}([C],[T])(k_{MN} + k_{IM}) + k_{MI}([C],[T])(k_{IN} + k_{NM})$$
$$+ (k_{IM} + k_{IN})(k_{NM} + k_{MN}).$$
$$\qquad (77)$$

The partition factor Φ can be expressed in terms of rate constants as $\Phi = k_{\mathrm{IN}}/(k_{\mathrm{IN}} + k_{\mathrm{IM}})$. Because of chaperone action, k_{IN} and k_{MI} can be significant (in particular $k_{\mathrm{IN}} \gg k_{\mathrm{NI}}$), whereas k_{NM} and k_{MN} are negligible. Under this condition, the native yield simplifies to:

$$P_N^{ss} \simeq \frac{k_{\mathrm{MI}}k_{\mathrm{IN}}}{k_{\mathrm{MI}}k_{\mathrm{NI}} + k_{\mathrm{NI}}k_{\mathrm{IM}} + k_{\mathrm{MI}}k_{\mathrm{IN}}}$$
$$\simeq \frac{k_{\mathrm{IN}}}{\dfrac{k_{\mathrm{NI}}}{k_{\mathrm{MI}}}k_{\mathrm{IM}} + k_{\mathrm{IN}}} = \frac{\Phi}{\kappa + (1-\kappa)\Phi} = N_\infty \qquad (78)$$

In the absence of either of chaperone or ATP which redistributes the population of proteins into NESS, $[C]$, $[T]$-dependent rate constants vanish ($k_{ij}([C],[T]) = 0$). In this case, the steady state population of N state becomes

$$P_N^{ss}([C] = 0 \text{ or } [T] = 0) = \frac{1}{1 + k_{NM}/k_{MN}}$$
$$= \frac{1}{1 + e^{-\Delta G_{NM}/k_B T}}$$
$$= P_N^{eq} \qquad (79)$$

Replacing the two $[C]$, $[T]$-dependent rate constants (k_{IN} and k_{IM}) to zero is tantamount to blocking the chaperone action and placing the M and N states in isolation. As long as they are isolated for long enough time greater than k_{NM}^{-1} and k_{MN}^{-1}, the system would finally reach the equilibrium native yield (P_N^{eq}), as dictated in Eq.79 assuming that aggregation reaction can be neglected. However, equilibrating M and N states via the transition paths of $M \rightleftharpoons N$ is impractical in the light of the time scale of cellular processes because k_{NM}^{-1} and k_{MN}^{-1} would far exceed the biologically meaningful time scale. These arguments suggest that chaperones drive the substrates out of equilibrium. In the process, they optimize the yield of the folded SPs or RNAs per unit time, which we discuss further below.
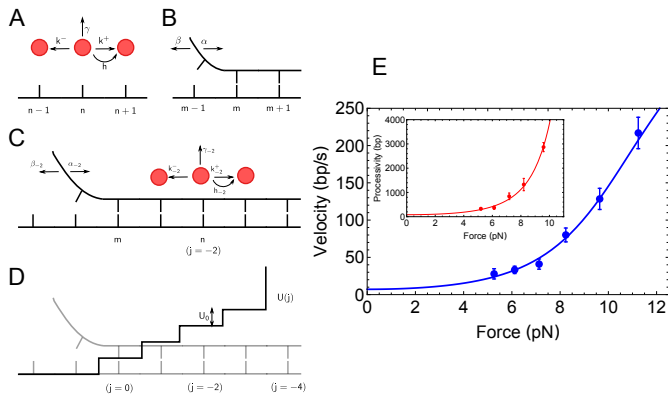
FIG. 24 Helicase model. Panels (A-D): Schematic of a model showing the unwinding of nucleic acids by helicases. (A) Translocation process of a helicase, with k^+, k^-, and γ being the rates of forward, backward, detachment, respectively; n is the position on the nucleic acid. (B) Opening (α) and closing (β) rates of a base pair. (C) Interactions between helicase and nucleic acids modify the stepping kinetics of the helicase and and the opening and closing rates of the base pair. The base pair location is m and $j = m - n$. (D) A model for the interaction with U_0 being the strength. For a passive helicase U_0 is zero. Panel (E): Simultaneous fits of the force-dependence of the velocity and processivity (given in the inset) of T7 helicase as a function of an external load. The blue and red lines are the results using the theory described elsewhere (Chakrabarti *et al.*, 2019) and the data points in circles are taken from experiments (Johnson *et al.*, 2007).

XI. UNIVERSAL CHARACTERISTICS OF HELICASES

Helicases, which are molecular motors found in all organisms, unwind double stranded (ds) DNA and dsRNA when they encounter a junction between single strand (ss)-ds nucleic acids (Delagoutte and Von Hippel, 2002, 2003; Lohman, 1992; Lohman and Bjornson, 1996). Separation of dsDNA strands is required for DNA replication as well as DNA repair. Malfunction of helicases causes genomic instability and are also implicated in cancer. In addition, certain RNA chaperones are also deemed to have helicase activity, which means they are able to unwind helices in RNA in order to facilitate its folding (Mohr *et al.*, 2002; Russell *et al.*, 2013). We refer the reader to a number of articles that describe a variety of cellular functions associated with helicases (Bianco and Kowalczykowski, 2000; Bustamante *et al.*, 2011; Dong *et al.*, 1995; Lohman *et al.*, 2008; Marians, 2000; Pang *et al.*, 2002; Pyle, 2008; Rocak and Linder, 2004; Velankar *et al.*, 1999; Venkatesan *et al.*, 1982). Helicases, classified into six super families based on their sequences (Gorbalenya and Koonin, 1993; Iyer *et al.*, 2004), are described as active or passive (Lohman, 1992). Active he-

licases destabilize the base pairs of the dsDNA, perhaps by exerting a force, thus separating the two strands. If the helicase is passive then it binds to the ssDNA whenever thermal fluctuations transiently open the base pairs. As the strands of the dsDNA are separated the helicase translocates along the ssDNA. This strand separation and translocation are intimately related. Several ensemble experiments have provided glimpses into the stepping mechanism (Ali and Lohman, 1997; Jeong *et al.*, 2004; Levin and Patel, 2002; Lucius *et al.*, 2003; Wong and Lohman, 1992) of helicases. In addition, these experiments have also been insightful in deciphering how they interact with single strand–double strand (ss-ds) junctions, and how often they dissociate from their track. The most detailed picture of the functions of a number of helicases have come from single molecule laser optical tweezers (LOT) and Magnetic Tweezers (MT) experiments. These experiments provided the kinetics of stepping and nucleic acid unwinding (Dessinges *et al.*, 2004; Lionnet *et al.*, 2006; Patel and Picha, 2000; Perkins *et al.*, 2004). Such measurements are instrumental in not only formulating theories and simulations but also in refining them as additional high precision experiments become available.

In a remarkable and influential paper, Betterton and Julicher (BJ) (Betterton and Jülicher, 2003, 2005a,b) presented a theory, which describes quantitatively the coupling between translocation and strand separation. The framework used in this theory, which is another illustration of the SKM, has been most instrumental in understanding the differences between active and passive helicases. The BJ model assumes that the helicase moves forward (backward) at a rate k^+ (k^-) when it is very far from the ss-ds junction. This aspect of the motor movement is similar to Fig. 5 with $\gamma=0$. When the motor is very far from the ss-ds junction the helicase merely translocates along the ss nucleic acids. Similarly, in isolation a nucleic acid base opens at a rate α and closes at a rate β. Depending on whether it is a AT or a GC base pair the rates are different but the inequality $\beta \gg \alpha$ holds in both cases. Since isolated base pair opening rates are due to thermal fluctuations they satisfy $\frac{\alpha}{\beta} = e^{-\Delta G_{bp}/k_B T}$ where ΔG_{bp} is stability of the base pair. However, interactions with base pairs modify these rates. Let n be the position of the helicase on the track and m be the location of the ss-ds junction (Fig. 24). Upon approaching the ds-ss junction the helicase interacts with NA, which was modeled using a variety of potentials - all based on some combination square well like potentials. Passive helicases, characterized by $U_0 \to \infty$, opportunistically step when the base is open. For active helicases, which forcibly rupture the base pair interactions, U_0 is finite but depends on $j = m - n$. The rates k^+ and k^- and α and β are modified when the helicase interacts with the NA (Betterton and Jülicher, 2005b). To describe the action of the helicase one has to keep track of its position on the

track as well the ss-ds location. Helicase-NA interactions modify all the relevant rates making them position (j) dependent. Using the model, with detachment rate set to zero, the velocity of the helicase assuming that it steps one base pair at a time, is given by,

$$V = \frac{1}{2} \sum_j \left(k_j^+ + \alpha_j - k_j^- - \beta_j \right) P_j, \qquad (80)$$

where P_j is the probability of observing the helicase and junction that are separated separated by j, α_j is the rate at which the junction opens when the helicase and junction are at separation j. The corresponding rates for base pair closing, forward, and backward stepping rates of the helicase are β_j, k_j^+, and k_j^-, respectively.

Although experiments (see for example (Manosas *et al.*, 2010)) have been analyzed using the theory sketched above the BJ model has to be generalized in order to make precise comparison with experiments. (1) To account for the average helicase processivity ($\langle \delta \rangle m$), which was not addressed in the original formulation, the consequences of detachment of the motor have to be considered. As pointed out earlier, almost all of the chemical kinetics models ignore detachment, which of course is unrealistic because the run-length in motors or number of base pairs that are disrupted if finite. (2) A theory that allows for arbitrary step size (s) is needed instead of assuming that s is unity. Helicases, such as PcrA and NS3 interact with and possibly destabilize several base pairs that are down stream of the ss-ds junction (Cheng *et al.*, 2007; Velankar *et al.*, 1999). In other words, step size s exceeds unity. (3) Experiments also apply external force and measure the changes in the processivity and velocity as a function of force, f. A viable theory should produce tractable expressions for both the mean velocity and $\langle \delta m \rangle$ as a function of of quantities. (4) Finally, how the sequence of the NA affects $\langle \delta m \rangle$ and V needs to be considered in order to draw general conclusions.

An analytically solvable model that accounts for the first three effects stated above has been proposed recently (Chakrabarti *et al.*, 2019), which was preceded by a less general theory (s was set to unity) (Pincus *et al.*, 2015) that investigate sequence effects ob the mean velocity and $\langle \delta \rangle m$. These studies produced a number of unexpected predictions for the helicase velocity and processivity as a function of external force and DNA sequence. (i) It was predicted that, regardless of the underlying architecture and unwinding kinetics of the helicase or the precise DNA sequence, processivity exhibits a universal increase with applied external force. This finding, which has subsequently has been validated experimentally (Bagchi *et al.*, 2018; Li *et al.*, 2016), was used to suggest that helicases may have evolved to maximize processivity rather than velocity. (ii) The theory, which quantitatively accounts for the experiments for force-dependent V and $\langle \delta m \rangle$ for T7 replisome (Fig. 6 from BJ 2019), shows that $s = 2$

base pairs. (iii) Normally, in analyzing experimental data back NTP-dependent stepping rates are neglected. This is justified while the helicase translocates along ss NA. For instance, in T7 helicase the ratio of the forward to back stepping rate is ~ 270, a value that is not that different from kinesin at zero resistive force. The probability of back-stepping is $\approx 0.3\%$. However, when T7 unwinds dsDNA the probability of back-stepping increases to 26%. This estimate is not that dissimilar to observations in XPD helicases belonging to a different (SF2) family, where it was shown that at 1mM ATP concentration the back-stepping probability is $\sim 10\%$. (iii) Interestingly, many helicases do not function *in vitro* unless an external force is applied. Under *in vivo* conditions partner proteins that bind to single strands and impart a force at the ss-ds junction to destabilize the base pairs (Pincus *et al.*, 2015) are needed. That this is the case has been shown for UvrD helicase (Comstock *et al.*, 2015), which behaves as a processive motor only when 2pN force is applied. It was noted previously that even though these associated proteins may or may not increase the unwinding velocity of a helicase they should universally increase the processivity of the helicase.

The theories for helicase stepping are not complete because they do not resolve many of the challenging problems. First, there is no quantitative explanation for the very broad velocity distribution (Johnson *et al.*, 2007) measured during unwinding of the weakly active ring-shaped T7 helicase. It is challenging to calculate $P(v)$ for the recent model (Chakrabarti *et al.*, 2019), which is the minimal that accounts for the F-dependent mean velocity and δm accurately. Second, there is very little understanding of the structural basis of the universal increase in $\langle \delta m \rangle$ as a function of F and the underlying dramatic variations in sequence and architectures of the motor. Perhaps, carefully designed simulations might shed light on this issue (Ma *et al.*, 2018; Yu *et al.*, 2006). Third, a recurring theme in many aspects of motors is that the dynamics is heterogeneous, exhibiting characteristics reminiscent of glasses. It was discovered that there are substantial molecule-to-molecule variations in the unwinding speed of *E. Coli*. RecBCD helicase even if all the enzymes are prepared under the same condition. To account for this observation it has been suggested that the functional landscape is likely partitioned into a number of metastable states and the initial preparation quenches the enzyme into a specific substate (Kirkpatrick and Thirumalai, 2015) The helicase ergodically explores all the conformations within a single metastable state but the transitions to other states could only be achieved by resetting the ATP concentration (Liu *et al.*, 2013). The emergence and relevance of glass-like heterogeneous behavior, under ambient conditions, is not understood theoretically, and remains a challenge not only in the context of helicases but also in other biological systems as well (Altschuler and Wu, 2010; Hyeon *et al.*, 2012; Solomatin

XII. DISCUSSION

The combination of excellent experiments and few theoretical approaches touched on here have greatly advanced our understanding of how biological machines work. However, we still do not have a complete understanding of how these machines work even in *in vitro* conditions. This should not be a surprise because some believe that the link between allosteric communication in hemoglobin (Hb) and oxygen transport is not fully understood despite over fifty years of intensive study. It is unclear if at present there is an analogue of Hb, considered the hydrogen molecule for allostery, in molecular machines. The investigation of these machines, one at a time *in vitro*, seems to raise unsolved problems in each individual case. We outline a few of the challenging problems. Clearly, the list is far from being exhaustive.

A. Sometimes Details Matter

Here, we have only described coarse-grained theoretical methods, which are impervious to the molecular details. There are several examples (we mention two) in which spectacular changes occur by a single or few amino acid substitutions. (i) About twenty years ago, Endow and Higuchi (Endow and Higuchi, 2000) made a single amino acid substitution in the neck linker (NL) region of a Ncd, a motor that is related to kinesin. The wild type Ncd walks towards the minus end of the microtubule in contrast to conventional kinesin. Upon substituting an asparagine (a polar amino acid residue) to lysine (positively charged) in the NL, an element that is responsible for the motor to walk predominantly on a single protofilament of the MT, it was found that Ncd moves in both the plus and minus direction on the MT. (ii) Myosin VI, unlike myosin V, walks towards the filamentous actin minus end and is the only known member belonging to the myosin super family with this property. In humans there are reports of three mutations that cause deafness. A point missense mutation (replacement of aspartic acid by tyrosine (D179Y) in the so-called U50 domain of the motor) leads to deafness in mouse (Hertzano *et al.*, 2008). It has been suggested (Pylypenko *et al.*, 2015), using a variety of experimental methods, that D179Y mutant leads to a premature release of the phosphate P_i, a product of ATP hydrolysis, from the detached head, thus preventing it from rapidly binding to F-actin. In the meanwhile, ATP does bind to the leading head, which detaches the motor from actin, thus preventing processive motion. Besides these examples there are many others, such as the link between mutations in β-cardiac myosin and hypertrophic cardiomyopathy. None of these obser-

vations are amenable to theoretical treatments exposed here, in which molecular details are ignored. Detailed simulations, if possible, could provide biophysical insights but linking such studies to functions is a daunting task. These anecdotal examples should remind us that in the search of principles of generality in biology one should not forget that molecular details matter and could dramatically influence functions.

B. Molecular basis of ATP hydrolysis:

In all the molecular machines hydrolysis of ATP is required to generate mechanical movements. However, the detailed molecular mechanisms by which the chemical free energy release from ATP (or GTP) hydrolysis is converted to conformational changes in the motor protein and eventual directed motion are remain unsolved. Elucidating these mechanisms remains an active area of research, for example using hybrid quantum mechanical / molecular mechanical (QM/MM) approaches to describe the ATP hydrolysis reaction in the active site of myosin (Grigorenko *et al.*, 2007; Kiani and Fischer, 2014). While these have provided insights into the first few steps of the energetic mechanism, the full molecular picture of what transpires in the myosin head post-hydrolysis, eventually culminating in a power stroke after rebinding to F-actin, remains incomplete. There has been an intriguing suggestion (Ross, 2006), supported by QM/MM calculations (Kamerlin and Warshel, 2009; Takahashi *et al.*, 2017), that electrostatics may play a key role: in the immediate aftermath of hydrolysis both ADP and P_i remain in the binding pocket, despite the strong Coulombic repulsion between them. One can imagine the following scenario: the separation of these negatively charged ions in the pocket may deform the surrounding protein, similar to cocking a spring. These deformations propagate allosterically into conformational changes that induce greater affinity between the myosin head and actin. A later round of allosteric changes upon actin rebinding leads to freeing the phosphate from the pocket and thus relieving the repulsive interaction: the "spring" is released, initiating a sequence of conformational rearrangements that ends in the power stroke. Considerable work will be needed to link how the molecular details of ATP hydrolysis, on a molecular length scale, is magnified to motility on several nm length scale.

C. Efficiency and optimality

A naive assessment of efficiency may be made by estimating the theoretical stall force, based on the available free energy due to ATP hydrolysis, and comparing it to the measured stall force (Kolomeisky and Fisher, 2007). Consider F_1-ATPase for example. This is part

of the F_0F_1-ATP synthase responsible for synthesizing ATP. This rotary motor undergoes precise $120°$ rotations in the absence of an external applied torque (τ_{tor}) at low ATP concentrations. By controlling τ_{tor} using the electro rotation method and the chemical potential by choosing appropriate ATP, ADP, and P_i concentrations, it is possible to measure the probability (p_s) of rotation in the synthetic direction (ADP and P_i are consumed to generate ATP and the probability p_h in the reverse hydrolytic direction could be measured (see for example (Toyabe *et al.*, 2011)). From the linear dependence of $k_B T ln \frac{p_s}{p_h}$ on it was found that the output energy at stall is roughly equal to the chemical potential. This implies that F_1-ATPase operates at near 100% efficiency.

For myosin motors, which take roughly a d=36 nm step, the maximum force that can be exerted is $f_{\max} \approx \frac{\Delta G_{ATP}}{d}$, which is approximately 2.5 pN assuming that $\Delta G_{ATP} \approx 22 k_B T$. The measured stall force (f_{stall}) is roughly in this ball park, which suggests that myosin motors operate efficiently ($\eta = \frac{f_{\text{stall}}}{f_{\max}}$ is very high). A similar argument for kinesin yields (d=8.1nm) $f_{\max} \approx$ 12pN whereas measurements report values close to 8 pN. Thus, there is about a 30-35% decrease in η for kinesin motors. A precise computation of efficiency should be undertaken by considering the network for a given motor that captures many aspects of motor motility. The arguments given above hold roughly if the motility can be described in a periodic one dimensional tilted potential with two equivalent sites with a transition state, which is close to the initial site. Treatments (Golubeva *et al.*, 2012; Schmiedl and Seifert, 2008; Seifert, 2011a; Wagoner and Dill, 2019) using more elaborate models indicate that the the motor efficiency would be much less and would depend on the details of the network dynamics.

A question that is related to efficiency is optimal performance. In the context of the biological machines discussed here, performance should be measured by velocity of movement, processivity, and for molecular chaperones the time-dependent production of folded state. As mentioned above, studies in the last decade have addressed how biological machines might optimize speed by considering models that are used to analyze force-velocity curves in motors. One of the lessons is that speed might be optimized if the motor takes many sub-steps instead of a single step (Wagoner and Dill, 2016). These studies have not considered processivity (run length) as a function of ATP and external force. For helicases, it appears that maximization of velocity is not as relevant as optimization of processivity. In addition, for GroEL it appears that the rate of production of the folded state per unit time is maximized even at the consumption of lavish amount of energy, which would render this machine highly inefficient. Whether questions pertaining to optimal performance, given available free energy, must consider simultaneously many functional requirements remain an open problem. It is possible that optimality could depend on specific function carried out by a class of machines.

D. Specificity versus Promiscuity

The *E. coli* chaperonin has evolved to be a promiscuous machine in that it facilitates the folding of a variety of misfolded substrate proteins (even those that are not in the *E. coli* proteome) that are unrelated by sequence, size, or the structure of the folded state. Using directed evolution methods a mutated GroEL/GroES, referred to as GroEL$_{3-1}$ was constructed (Wang *et al.*, 2002). The altered GroEL$_{3-1}$ contained a single mutation in GroES (tyrosine was replaced by histidine) and two mutations (valine was substituted for alanine and glycine for aspartic acid) in GroEL. It was found that GroEL$_{3-1}$ had enhanced ATPase activity compared to the wild type. More importantly, it was found that GroEL$_{3-1}$, with a highly polar environment in the cavity compared to the wild-type, dramatically increased the folding of green fluorescent protein (GFP). However, the enhanced specificity of folding GFP with ease came at the expense of substantial reduction in the capacity of GroEL$_{3-1}$ to facilitate the folding of several other proteins. Thus, in this instance nature has solved the tension between specificity and promiscuity by evolving an all purpose *E. coli* chaperonin that can process the folding of a large class of proteins, albeit not as efficiently. In eukaryotes there has been a great expansion in the number of chaperone classes possibly to enable the larger and more complex proteome. This example suggests that evolutionary constraints might have to a part of addressing issues related to optimality. This might imply that the simple network used to explain the out of equilibrium performance of GroEL/GroES has to expanded to describe optimality in the functions of chaperone networks in eukaryotes.

E. Biological complexity

We circle back to Fig. 1, which is a schematic illustration of transport of melanosomes (vesicles containing the light absorbing pigment melanin found in amphibians). The vesicles are either dispersed throughout the cytosol or aggregate near the cell center. The transportation of melanosomes is clearly complex and is controlled by the interplay of multiple motors involving kinesin-2, a plus end directed MT motor, and dynein, that walks towards the minus end of the MT. In addition, actin bound myosin V is also involved in the transport. It is suspected that a low number of motors (about 1-2 kinesin-2 motors and roughly 1-3 dyneins) move melanosomes during aggregation (Levi *et al.*, 2006). In contrast, pigment dispersion is mediated by kinesin-2 as well as assistance by

myosin V. Because both these motors compete for the same attachment site (p150Glue on dynactin, see Fig. 1) it follows (Levi *et al.*, 2006) that myosin V is released during aggregation and vesicle transport is dominated by dynein. The need to change movement of melanosomes, powered by multiple motors, is thus determined by function, which in this case is related to their dispersion or aggregation.

Although individual motors predominantly move unidirectionally on cytoskeletal filamentous tracks there are reports that motors could change directions as well. A very impressive *in vitro* illustration of bidirectional motility of the complex of dynein with dynactin (a complex that is attached to the cargo and activates dynein (Fig. 1) was reported sometime ago (Ross *et al.*, 2006). It was found that the complex moves processively in both directions on MT with ATP-dependent velocities that are not significantly different in either direction (Ross *et al.*, 2006). Although the authors provide a qualitative picture of the mechanism of bidirectional transport a theory for such unexpected behavior is lacking. It is not even clear, at least to us, the level of coarse-graining needed to construct such a theory, which is clearly needed to unveil the complexity of vesicle transport.

We have not discussed how groups of motors move on cytoskeletal structures. It is likely that under most physiological conditions motors operate in groups. As is often the case in physics, the consequences of collective effects of motors cannot be easily described using the behavior of individual motors alone. Typically, new and unexpected behaviors emerge. In order to shed light on this important topic, several theoretical studies (Beeg *et al.*, 2008; Campas *et al.*, 2006; Guerin *et al.*, 2010; Klumpp and Lipowsky, 2005; Malgaretti *et al.*, 2017; McLaughlin *et al.*, 2016) have been proposed. Some of the new effects that are predicted to emerge include not only tug-of-war but also bidirectional movements, spontaneous oscillations, and positive as well as negative cooperativity. Although considerable work has already been done in this field understanding *in vivo* operation of motors will require scrutinizing how groups of different kinds of motors function.

XIII. A FINAL REMARK

The eventual goals of understanding biology through the lenses of physics are to create theoretical tools that are capable of describing biological functions under crowded and noisy cellular conditions, and in the process discover general physical principles that control life processes. It is likely that as the scale at which living systems are examined increases it may be possible to describe cellular processes using functional modules (Hartwell *et al.*, 1999), which is a coarse grained view of biology. However, it would be hard to anticipate the functions of such modules from their components, which are the molecules of life. Furthermore, interactions between modules could lead to new functions not encoded in isolated modules, as illustrated by many examples described in our perspective. After all "more is different" (Anderson, 1972). Because "Nature is an excellent tinkerer, not an engineer" (a quote attributed to Francois Jacob), and tinkering in biology involves stochastically altering existing modules to evolve new functions without time constraints or any ultimate design as a goal (Jacob, 1977), it is likely that concepts in many fields of science would have to be used to develop an integrated view of biology.

REFERENCES

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter (2008), *Molecular Biology of the Cell*, 5th ed. (Garland Science).

Alhadeff, R., and A. Warshel (2017), Proc. Natl. Acad. Sci. **114** (39), 10426.

Ali, J. A., and T. M. Lohman (1997), Science **275** (5298), 377.

Altman, D., H. L. Sweeney, and J. A. Spudich (2004), Cell **116** (5), 737.

Altschuler, S. J., and L. F. Wu (2010), Cell **141**, 559.

Anderson, P. W. (1972), Science **177** (4047), 393.

Andrecka, J., J. O. Arroyo, Y. Takagi, G. de Wit, A. Fineberg, L. MacKinnon, G. Young, J. R. Sellers, and P. Kukura (2015), eLife **4**, e05413.

Ariga, T., M. Tomishige, and D. Mizuno (2018), Phys. Rev. Lett. **121** (21), 218101.

Astumian, R. D., and M. Bier (1996), Biophys. J. **70** (2), 637.

Bagchi, D., M. Manosas, W. Zhang, K. A. Manthei, S. Hodeib, B. Ducos, J. L. Keck, and V. Croquette (2018), Nucleic acids research **46** (16), 8500.

Bagshaw, C., and D. Trentham (1973), Biochem. J. **133** (2), 323.

Baker, J. E., E. B. Krementsova, G. G. Kennedy, A. Armstrong, K. M. Trybus, and D. M. Warshaw (2004), Proc. Natl. Acad. Sci. **101** (15), 5542.

Banerjee, K., A. B. Kolomeisky, and O. A. Igoshin (2017), Proc. Natl. Acad. Sci. **114** (20), 5183.

Barato, A. C., and U. Seifert (2015), Phys. Rev. Lett. **114** (15), 158101.

Battle, C., C. P. Broedersz, N. Fakhri, V. F. Geyer, J. Howard, C. F. Schmidt, and F. C. MacKintosh (2016), Science **352** (6285), 604.

Beeg, J., S. Klumpp, R. Dimova, R. S. Gracia, E. Unger, and R. Lipowsky (2008), Biophys. J. **94**, 532.

Bell, G. I. (1978), Science **200** (4342), 618.

Belyy, V., N. L. Hendel, A. Chien, and A. Yildiz (2014a), Nat. Comm. **5**, 5544.

Belyy, V., N. L. Hendel, A. Chien, and A. Yildiz (2014b), Nat. Comm. **5**, 5544.

Bennett, C. H. (1982), International Journal of Theoretical Physics **21** (12), 905.

Betterton, M. D., and F. Jülicher (2003), Phys. Rev. Lett. **91** (25), 258103.

Betterton, M. D., and F. Jülicher (2005a), Phys Rev E **71** (1 Pt 1), 011904.

Betterton, M. D., and F. Jülicher (2005b), J. Phys.: Condens. Matter **17** (47), S3851.

Bhabha, G., H.-C. Cheng, N. Zhang, A. Moeller, M. Liao, J. A. Speir, Y. Cheng, and R. D. Vale (2014), CELL **159** (4), 857.

Bhaskaran, H., and R. Russell (2007), Nature **449** (7165), 1014.

Bianco, P. R., and S. C. Kowalczykowski (2000), Nature **405** (6784), 368.

Bierbaum, V., and R. Lipowsky (2011), Biophys. J. **100** (7), 1747.

Block, S. M. (2007), Biophys. J. **92** (9), 2986.

Block, S. M., C. L. Asbury, J. W. Shaevitz, and M. J. Lang (2003), Proc. Natl. Acad. Sci. **100** (5), 2351.

Block, S. M., L. S. Goldstein, and B. J. Schnapp (1990), Nature **348** (6299), 348.

Brady, S. T. (1985), Nature **317** (6032), 73.

Brown, A. I., and D. A. Sivak (2017), Proc. Natl. Acad. Sci. **114** (42), 11057.

Bustamante, C., W. Cheng, and Y. X. Mejia (2011), Cell **144** (4), 480.

Callen, H. B. (1985), *Thermodynamics and an Introduction to Thermostatistics* (Wiley).

Campas, O., Y. Kafri, K. B. Zeldovich, J. Casademunt, and J. F. Joanny (2006), Phys. Rev. Lett. **97** (3), 038101.

Caporizzo, M. A., C. E. Fishman, O. Sato, R. M. Jamiolkowski, M. Ikebe, and Y. E. Goldman (2018), Biophys. J. **114** (6), 1400.

Carter, N. J., and R. Cross (2005), Nature **435** (7040), 308.

Chakrabarti, S., C. Hyeon, X. Ye, G. H. Lorimer, and D. Thirumalai (2017), Proc. Natl. Acad. Sci. **114** (51), E10919.

Chakrabarti, S., C. Jarzynski, and D. Thirumalai (2019), Biophys. J. 10.1016/j.bpj.2019.07.021.

Cheng, W., S. Dumont, I. Tinoco, and C. Bustamante (2007), Proc. Natl. Acad. Sci. **104** (35).

Clancy, B. E., W. M. Behnke-Parks, J. O. Andreasson, S. S. Rosenfeld, and S. M. Block (2011), Nat. Struct. Mol. Biol. **18** (9), 1020.

Clemen, A. E. M., M. Vilfan, J. Jaud, J. S. Zhang, M. Barmann, and M. Rief (2005), Biophys. J. **88** (6), 4402.

Comstock, M. J., K. D. Whitley, H. Jia, J. Sokoloski, T. M. Lohman, T. Ha, and Y. R. Chemla (2015), Science **348** (6232), 352.

Craig, E. M., and H. Linke (2009), Proc. Natl. Acad. Sci. **106** (43), 18261.

Crooks, G. E. (1998), J. Stat. Phys. **90** (5-6), 1481.

De la Cruz, E. M., E. M. Ostap, and H. L. Sweeney (2001), J. Biol. Chem. **276** (34), 32373.

De la Cruz, Enrique M M., A. L. Wells, S. S. Rosenfeld, E. M. Ostap, and H. L. Sweeney (1999), Proc. Natl. Acad. Sci. **96** (24), 13726.

De La Cruz, E. M., and E. M. Ostap (2004), Curr. Opin. Cell Biol. **16** (1), 61.

Dechant, A., and S.-i. Sasa (2018a), Journal of Statistical Mechanics: Theory and Experiment **2018** (6), 063209.

Dechant, A., and S.-i. Sasa (2018b), Phys. Rev. E **97** (6), 062101.

Delagoutte, E., and P. H. Von Hippel (2002), Q. Rev. Biophys. **35** (4), 431.

Delagoutte, E., and P. H. Von Hippel (2003), Q. Rev. Biophys. **36** (1), 1.

Derrida, B. (1983), Journal of statistical physics **31** (3), 433.

Dessinges, M.-N., T. Lionnet, X. G. Xi, D. Bensimon, and V. Croquette (2004), Proc. Natl. Acad. Sci. **101** (17), 6439.

DeWitt, M. A., A. Y. Chang, P. A. Combs, and A. Yildiz (2012), Science **335** (6065), 221.

Dominguez, R., Y. Freyzon, K. M. Trybus, and C. Cohen (1998), Cell **94** (5), 559.

Dong, F., E. P. Gogol, and P. H. von Hippel (1995), J. Biol. Chem. **270** (13), 7462.

Dunn, A. R., P. Chuan, Z. Bryant, and J. A. Spudich (2010), Proc. Natl. Acad. Sci. **107** (17), 7746.

Dunn, A. R., and J. A. Spudich (2007), Nat. Struct. Mol. Biol. **14** (3), 246.

Ehrenberg, M., and C. Blomberg (1980), Biophys. J. **31** (3), 333.

Elber, R., and A. West (2010), Proc. Natl. Acad. Sci. **107** (11), 5001.

Elting, M. W., Z. Bryant, J.-C. Liao, and J. A. Spudich (2011), Biophys. J. **100** (2), 430.

Endow, S., and H. Higuchi (2000), Nature **406** (6798), 913.

Fei, X., X. Ye, N. A. LaRonde, and G. H. Lorimer (2014), Proc. Natl. Acad. Sci. **111** (35), 12775.

Finer, J. T., R. M. Simmons, and J. A. Spudich (1994), Nature **368** (6467), 113.

Fisher, M. E., and A. B. Kolomeisky (1999a), Proc. Natl. Acad. Sci. **96** (12), 6597.

Fisher, M. E., and A. B. Kolomeisky (1999b), Physica A: Statistical Mechanics and its Applications **274** (1), 241.

Fisher, M. E., and A. B. Kolomeisky (2001), Proc. Natl. Acad. Sci. **98** (14), 7748.

Ge, H., and H. Qian (2010), Phys. Rev. E **81** (5), 051133.

Gebhardt, J. C. M., A. E.-M. Clemen, J. Jaud, and M. Rief (2006), Proc. Natl. Acad. Sci. **103** (23), 8680.

Gennerich, A., A. P. Carter, S. L. Reck-Peterson, and R. D. Vale (2007), Cell **131** (5), 952.

Gerritsma, E., and P. Gaspard (2010), Biophys. Rev. Lett. **5** (04), 163.

Gibbons, I., and A. Rowe (1965), Science **149** (3682), 424.

Gingrich, T. R., J. M. Horowitz, N. Perunov, and J. L. England (2016), Phys. Rev. Lett. **116** (12), 120601.

Gladrow, J., N. Fakhri, F. MacKintosh, C. Schmidt, and C. Broedersz (2016), Phys. Rev. Lett. **116** (24), 248301.

Goldtzvik, Y., M. Hinczewski, and D. Thirumalai (2019), preprint.

Goldtzvik, Y., M. L. Mugnai, and D. Thirumalai (2018a), Structure **26** (12), 1664.

Goldtzvik, Y., M. L. Mugnai, and D. Thirumalai (2018b), Structure **26** (12), 1664.

Golubeva, N., A. Imparato, and L. Peliti (2012), Euro Phys. Lett. **97** (6), 60005.

Gorbalenya, A. E., and E. V. Koonin (1993), Curr. Opin. Struct. Biol. **3** (3), 419.

Grigorenko, B. L., A. V. Rogov, I. A. Topol, S. K. Burt, H. M. Martinez, and A. V. Nemukhin (2007), Proc. Natl. Acad. Sci. **104** (17), 7057.

Gross, S. P., M. C. Tuma, S. W. Deacon, A. S. Serpinskaya, A. R. Reilein, and V. I. Gelfand (2002), J. Cell. Biol. **156** (5), 855.

Guerin, T., J. Prost, P. Martin, and J.-F. Joanny (2010), Curr. Opin. Cell Biol. **22** (1), 14.

Guo, Z., and D. Thirumalai (1995), Biopolymers: Original Research on Biomolecules **36** (1), 83.

Hackney, D. D. (1988), Proc. Natl. Acad. Sci. **85** (17), 6314.

Hackney, D. D. (2005), Proc. Natl. Acad. Sci. **102** (51), 18338.

Hancock, W. O. (2016), Biophys. J. **110** (6), 1216.

Harada, T., and S.-i. Sasa (2006), Phys. Rev. E **73** (2), 026131.

Hartman, M. A., D. Finan, S. Sivaramakrishnan, and J. A. Spudich (2011), Ann. Rev. Cell Dev. Biol. **27**, 133.

Hartman, M. A., and J. A. Spudich (2012), J. Cell. Sci. **125** (7), 1627.

Hartwell, L. H., J. J. Hopfield, S. Leibler, and A. W. Murray (1999), Nature **402** (6761supp), C47.

Hathcock, D., R. Tehver, M. Hinczewski, and D. Thirumalai (2019), .

Hertzano, R., E. Shalit, A. K. Rzadzinska, A. A. Dror, L. Song, U. Ron, J. T. Tan, A. S. Shitrit, H. Fuchs, T. Hasson, *et al.* (2008), PLoS genetics **4** (10), e1000207.

Hill, T. L. (1966), J. Theor. Biol. **10** (3), 442.

Hill, T. L. (2005a), *Free Energy Transduction and Biochemical Cycle Kinetics* (Dover).

Hill, T. L. (2005b), *Free energy transduction and biochemical cycle kinetics* (Dover Publications, Inc.).

Hill, T. L., and Y.-D. Chen (1975), Proc. Natl. Acad. Sci. **72** (4), 1291.

Hill, T. L., and R. Simmons (1976), Proc. Natl. Acad. Sci. **73** (1), 95.

Hinczewski, M., R. Tehver, and D. Thirumalai (2013), Proc. Natl. Acad. Sci. **110** (43), E4059.

Hirokawa, N., Y. Noda, Y. Tanaka, and S. Niwa (2009), Nat. Rev. Mol. Cell. Biol. **10** (10), 682.

Holzbaur, E. L., and Y. E. Goldman (2010), Curr. Opin. Cell Biol. **22** (1), 4.

Homma, K., and M. Ikebe (2005), J. Biol. Chem. **280** (32), 29381.

Hopfield, J. J. (1974), Proc. Natl. Acad. Sci. **71** (10), 4135.

Howard, J. (2001), *Mechanics of motor proteins and the cytoskeleton* (Sinauer associates Sunderland, MA).

Howard, J. (2006), Current Biology **16** (14), R517.

Howard, J., A. Hudspeth, and R. Vale (1989), Nature **342** (6246), 154.

Hulett, H. (1970), Nature **225** (5239), 1248.

Hwang, W., and C. Hyeon (2017), J. Phys. Chem. Lett. **8** (1), 250.

Hwang, W., M. J. Lang, and M. Karplus (2008), Structure **16** (1), 62.

Hwang, W., M. J. Lang, and M. Karplus (2017), eLife **6**, e28948.

Hyeon, C., and W. Hwang (2017), Phys. Rev. E **96** (1), 012156.

Hyeon, C., S. Klumpp, and J. N. Onuchic (2009), Phys. Chem. Chem. Phys. **11** (24), 4899.

Hyeon, C., J. Lee, J. Yoon, S. Hohng, and D. Thirumalai (2012), Nat. Chem. **4**, 907.

Hyeon, C., and J. N. Onuchic (2007a), Proc. Natl. Acad. Sci. **104** (7), 2175.

Hyeon, C., and J. N. Onuchic (2007b), Proc. Natl. Acad. Sci. **104** (44), 17382.

Hyeon, C., and D. Thirumalai (2011), Nat. Commun. **2**, 487.

Ikezaki, K., T. Komori, M. Sugawa, Y. Arai, S. Nishikawa, A. H. Iwane, and T. Yanagida (2012), Small **8** (19), 3035.

Ikezaki, K., T. Komori, and T. Yanagida (2013), PloS one **8** (3), e58912.

Inoue, A., J. Saito, R. Ikebe, and M. Ikebe (2002), Nature cell biology **4** (4), 302.

Isojima, H., R. Iino, Y. Niitani, H. Noji, and M. Tomishige (2016), Nat. Chem. Biol. **12** (4), 290.

Itoh, H., A. Takahashi, K. Adachi, H. Noji, R. Yasuda, M. Yoshida, and K. Kinosita Jr (2004), Nature **427** (6973), 465.

Itsathitphaisarn, O., R. A. Wing, W. K. Eliason, J. Wang, and T. A. Steitz (2012), Cell **151** (2), 267.

Iyer, L. M., D. D. Leipe, E. V. Koonin, and L. Aravind (2004), J. Struct. Biol. **146** (1-2), 11.

Jacob, F. (1977), Science **196** (4295), 1161.

Jeong, Y.-J., M. K. Levin, and S. S. Patel (2004), Proc. Natl. Acad. Sci. **101** (19), 7264.

Johnson, D. S., L. Bai, B. Y. Smith, S. S. Patel, and M. D. Wang (2007), Cell **129** (7), 1299.

Jülicher, F., A. Ajdari, and J. Prost (1997), Rev. Mod. Phys. **69** (4), 1269.

Kad, N. M., K. M. Trybus, and D. M. Warshaw (2008), J. Biol. Chem. **283** (25), 17477, http://www.jbc.org/content/283/25/17477.full.pdf+html.

Kamerlin, S. C., and A. Warshel (2009), J. Phys. Chem. B **113** (47), 15692.

Kiani, F. A., and S. Fischer (2014), Proc. Natl. Acad. Sci. **111** (29), E2947.

Kiefhaber, T. (1995), Proc. Natl. Acad. Sci. **92** (20), 9029.

Kirkpatrick, T., and D. Thirumalai (2015), Rev. Mod. Phys. **87** (1), 183.

Klumpp, S., and R. Lipowsky (2005), Proc. Natl. Acad. Sci. U. S. A. **102** (48), 17284.

Kodera, N., D. Yamamoto, R. Ishikawa, and T. Ando (2010), Nature **468** (7320), 72.

Kolomeisky, A. B., and M. E. Fisher (2000), Physica A: Statistical Mechanics and its Applications **279** (1), 1.

Kolomeisky, A. B., and M. E. Fisher (2003), Biophys. J. **84** (3), 1642.

Kolomeisky, A. B., and M. E. Fisher (2007), Annu. Rev. Phys. Chem. **58**, 675.

Koza, Z. (1999), Journal of Physics A: Mathematical and General **32** (44), 7637.

Koza, Z. (2002), arXiv preprint cond-mat/0207203.

Lan, G., P. Sartori, S. Neumann, V. Sourjik, and Y. Tu (2012), Nature physics **8** (5), 422.

Lebowitz, J. L., and H. Spohn (1999), J. Stat. Phys. **95** (1-2), 333.

Leibler, S., and D. A. Huse (1993), J. Cell. Biol. **121** (6), 1357.

Levi, V., A. Serpinskaya, E. Gratton, and V. Gelfand (2006), Biophys. J. **90** (1), 318.

Levin, M. K., and S. S. Patel (2002), J. Biol. Chem. **277** (33), 29377.

Li, J.-H., W.-X. Lin, B. Zhang, D.-G. Nong, H.-P. Ju, J.-B. Ma, C.-H. Xu, F.-F. Ye, X. G. Xi, M. Li, *et al.* (2016), Nucleic acids research **44** (9), 4330.

Liepelt, S., and R. Lipowsky (2007a), Phys. Rev. Lett. **98** (25), 258102.

Liepelt, S., and R. Lipowsky (2007b), EPL (Europhysics Letters) **77** (5), 50002.

Lionnet, T., A. Dawid, S. Bigot, F.-X. Barre, O. A. Saleh, F. Heslot, J.-F. Allemand, D. Bensimon, and V. Croquette (2006), Nucleic acids research **34** (15), 4232.

Lipowsky, R., and S. Liepelt (2008), J. Stat. Phys. **130** (1), 39.

Lipowsky, R., S. Liepelt, and A. Valleriani (2009), J. Stat. Phys. **135** (5), 951.

Liu, B., R. J. Baskin, and S. C. Kowalczykowski (2013), Nature **500**, 482.

Liu, C., X. Fu, L. Liu, X. Ren, C. K. Chau, S. Li, L. Xiang, H. Zeng, G. Chen, L.-H. Tang, *et al.* (2011), Science **334** (6053), 238.

Lohman, T. (1992), Molecular microbiology **6** (1), 5.

Lohman, T. M., and K. P. Bjornson (1996), Annual review of biochemistry **65** (1), 169.

Lohman, T. M., E. J. Tomko, and C. G. Wu (2008), Nature Reviews Molecular Cell Biology **9** (5), 391.

Lucius, A. L., N. K. Maluf, C. J. Fischer, and T. M. Lohman (2003), Biophys. J. **85** (4), 2224.

Ma, J., P. Sigler, Z. Xu, and M. Karplus (2000), J. Mol. Biol. **302** (2), 303.

Ma, W., K. D. Whitley, Y. R. Chemla, Z. Luthey-Schulten, and K. Schulten (2018), eLife **7**, e34186.

Maes, C., and M. H. Van Wieren (2003), J. Stat. Phys. **112** (1-2), 329.

Malgaretti, P., I. Pagonabarraga, and J. F. Joanny (2017), Phys. Rev. Lett. **119** (16), 168101.

Manosas, M., X. G. Xi, D. Bensimon, and V. Croquette (2010), Nuc. Acids Res. **38** (16), 5518.

Marians, K. J. (2000), Structure **8** (12), R227.

McLaughlin, R., D. M.R, and A. Kolomeisky (2016), Soft Matter **12** (1), 14.

Mehta, A. D., R. S. Rock, M. Rief, J. A. Spudich, M. S. Mooseker, and R. E. Cheney (1999), Nature **400** (6744), 590.

Mehta, P., and D. J. Schwab (2012), Proc. Natl. Acad. Sci. **109** (44), 17978.

Mickolajczyk, K. J., N. C. Deffenbaugh, J. O. Arroyo, J. Andrecka, P. Kukura, and W. O. Hancock (2015), Proc. Natl. Acad. Sci. **112** (52), E7186.

Milic, B., J. O. Andreasson, D. W. Hogan, and S. M. Block (2017), Proc. Natl. Acad. Sci. **114** (33), E6830.

Mohr, S., J. M. Stryker, and A. M. Lambowitz (2002), Cell **109**, 769.

Moore, J. R., E. B. Krementsova, K. M. Trybus, and D. M. Warshaw (2004), Journal of Muscle Research & Cell Motility **25** (1), 29.

Mugnai, M., M. Caporizzo, Y. Goldman, and D. Thirumalai (2019), bioRxiv , 684696.

Mugnai, M. L., and D. Thirumalai (2017), Proc. Natl. Acad. Sci. **114** (22), E4389.

Mukherjee, S., R. Alhadeff, and A. Warshel (2017), Proc. Natl. Acad. Sci. **114** (9), 2259.

Mukherjee, S., and A. Warshel (2013), Proc. Natl. Acad. Sci. **110** (43), 17326.

Nam, W., and B. I. Epureanu (2016), PloS one **11** (1), e0147676.

Nishikawa, S., I. Arimoto, K. Ikezaki, M. Sugawa, H. Ueno, T. Komori, A. H. Iwane, and T. Yanagida (2010), Cell **142** (6), 879.

Nishiyama, M., H. Higuchi, and T. Yanagida (2002), Nature Cell Biology **4** (10), 790.

Oguchi, Y., S. V. Mikhailenko, T. Ohki, A. O. Olivares, E. M. De la Cruz, and S. Ishiwata (2008), Proc. Natl. Acad. Sci. **105** (22), 7714.

Oguchi, Y., S. V. Mikhailenko, T. Ohki, A. O. Olivares, E. M. De La Cruz, and S. Ishiwata (2010), Nat. Chem. Biol. **6** (4), 300.

Oke, O. A., S. A. Burgess, E. Forgacs, P. J. Knight, T. Sakamoto, J. R. Sellers, H. White, and J. Trinick (2010), Proc. Natl. Acad. Sci. **107** (6), 2509.

Ökten, Z., L. S. Churchman, R. S. Rock, and J. A. Spudich (2004), Nat. Struct. Mol. Biol. **11** (9), 884.

Oppenheim, I., K. E. Shuler, G. H. Weiss, *et al.* (1977), *Stochastic processes in chemical physics* (Mit Press).

Ori-McKenney, K. M., J. Xu, S. P. Gross, and R. B. Vallee (2010), Nature cell biology **12** (12), 1228.

Ortega, A., D. Amorós, and J. G. de La Torre (2011), Biophys. J. **101** (4), 892.

Pan, J., D. Thirumalai, and S. A. Woodson (1997), Journal of molecular biology **273** (1), 7.

Pang, P. S., E. Jankowsky, P. J. Planet, and A. M. Pyle (2002), The EMBO journal **21** (5), 1168.

Patel, S. S., and K. M. Picha (2000), Annual review of biochemistry **69** (1), 651.

Perkins, T. T., H.-W. Li, R. V. Dalal, J. Gelles, and S. M. Block (2004), Biophys. J. **86** (3), 1640.

Pierobon, P., S. Achouri, S. Courty, A. R. Dunn, J. A. Spudich, M. Dahan, and G. Cappello (2009), Biophys. J. **96** (10), 4268.

Pigolotti, S., I. Neri, É. Roldán, and F. Jülicher (2017), Phys. Rev. Lett. **119** (14), 140604.

Pincus, D. L., S. Chakrabarti, and D. Thirumalai (2015), Biophys. J. **109** (2), 220.

Post, P. L., M. J. Tyska, C. B. O'Connell, K. Johung, A. Hayward, and M. S. Mooseker (2002), J. Biol. Chem. **277** (14), 11679.

Purcell, T. J., H. L. Sweeney, and J. A. Spudich (2005), Proc. Natl. Acad. Sci. **102** (39), 13873.

Pyle, A. M. (2008), Annu. Rev. Biophys. **37**, 317.

Pylypenko, O., L. Song, A. Shima, Z. Yang, A. M. Houdusse, and H. L. Sweeney (2015), Proc. Natl. Acad. Sci. **112** (11), E1201.

Qian, H. (2004), Phys. Rev. E **69** (1), 012901.

Qian, H. (2007), Annu. Rev. Phys. Chem. **58**, 113.

Qian, H., and D. A. Beard (2005), Biophys. Chem. **114** (2-3), 213.

Reck-Peterson, S. L., W. B. Redwine, R. D. Vale, and A. P. Carter (2018), Nat. Rev. Mol. Cell. Biol. **19** (6), 382.

Reck-Peterson, S. L., A. Yildiz, A. P. Carter, A. Gennerich, N. Zhang, and R. D. Vale (2006a), Cell **126** (2), 335.

Reck-Peterson, S. L., A. Yildiz, A. P. Carter, A. Gennerich, N. Zhang, and R. D. Vale (2006b), Cell **126** (2), 335.

Rice, S., A. W. Lin, D. Safer, C. L. Hart, N. Naber, B. O. Carragher, S. M. Cain, E. Pechatnikova, E. M. Wilson-Kubalek, M. Whittaker, *et al.* (1999), Nature **402** (6763),

778.

Rocak, S., and P. Linder (2004), Nat. Rev. Mol. Cell Biol. **5** (3), 232.

Rock, R. S., S. E. Rice, A. L. Wells, T. J. Purcell, J. A. Spudich, and H. L. Sweeney (2001), Proc. Natl. Acad. Sci. **98** (24), 13655.

Rodenfels, J., K. M. Neugebauer, and J. Howard (2019), Developmental cell **48** (5), 646.

Rosenfeld, S. S., and H. Lee Sweeney (2004), J. Biol. Chem. **279** (38), 40100.

Ross, J. (2006), J. Phys. Chem. B **110** (13), 6987.

Ross, J. L., K. Wallace, H. Shuman, Y. E. Goldman, and E. L. F. Holzbaur (2006), Nat. Cell. Biol. **8** (6), 562.

Russell, R., I. Jarmoskaite, and A. M. Lambowitz (2013), RNA Biology **10** (1), 44.

Sakamoto, T., I. Amitani, E. Yokota, and T. Ando (2000), Biochem. Biophys. Res. Commun. **272** (2), 586.

Sakamoto, T., A. Yildiz, P. R. Selvin, and J. R. Sellers (2005), Biochemistry **44** (49), 16203.

Sarlah, A., and A. Vilfan (2014), Biophys. J. **107** (3), 662.

Šarlah, A., and A. Vilfan (2014), Biophys. J. **107** (3), 662.

Schief, W. R., R. H. Clark, A. H. Crevenna, and J. Howard (2004), Proc. Natl. Acad. Sci. **101** (5), 1183.

Schmidt, H., R. Zalyte, L. Urnavicius, and A. P. Carter (2015), Nature **518** (7539), 435.

Schmiedl, T., and U. Seifert (2008), EPL (Europhysics Letters) **83** (3), 30005.

Schnakenberg, J. (1976), Rev. Mod. Phys. **48** (4), 571.

Schnitzer, M. J., and S. Block (1995), in *Cold spring harbor symposia on quantitative biology*, Vol. 60 (Cold Spring Harbor Laboratory Press) pp. 793–802.

Schnitzer, M. J., and S. M. Block (1997), Nature **388** (6640), 386.

Schnitzer, M. J., K. Visscher, and S. M. Block (2000), Nature cell biology **2** (10), 718.

Seifert, U. (2005a), Phys. Rev. Lett. **95** (4), 040602.

Seifert, U. (2005b), EPL (Europhysics Letters) **70** (1), 36.

Seifert, U. (2011a), Phys. Rev. Lett. **106** (2), 020601.

Seifert, U. (2011b), Euro. Phys. J. E **34** (3), 26.

Seifert, U. (2012), Rep. Prog. Phys. **75** (12), 126001.

Sellers, J. (2000), Biochem. Biophys. ACTA - Mol. Cell Res. **1496** (1), 3.

Sellers, J. R., and C. Veigel (2010), Nat. Struct. Mol. Biol. **17** (5), 590.

Shiraishi, N., K. Saito, and H. Tasaki (2016), Phys. Rev. Lett. **117** (19), 190601.

Soldati, T., and M. Schliwa (2006), Nature Reviews Molecular Cell Biology **7** (12), 897.

Solomatin, S. V., M. Greenfeld, S. Chu, and D. Herschlag (2010), Nature **463**, 681.

Song, Y., J. O. Park, L. Tanner, Y. Nagano, J. D. Rabinowitz, and S. Y. Shvartsman (2019), Current Biology **29** (12), R566.

Spudich, J. A., and S. Sivaramakrishnan (2010), Nat. Rev. Mol. Cell. Biol. **11** (2), 128.

Stan, G., B. Brooks, and D. Thirumalai (2005), J. Mol. Biol. **350** (4), 817.

Sumi, T., and S. Klumpp (2019), Nano Lett. **19**, 3370.

Sun, Y., and Y. E. Goldman (2011a), Biophys. J. **101** (1), 1.

Sun, Y., and Y. E. Goldman (2011b), Biophys. J. **101** (1), 1.

Sun, Y., O. Sato, F. Ruhnow, M. E. Arsenault, M. Ikebe, and Y. E. Goldman (2010), Nat. Struct. Mol. Biol. **17** (4), 485.

Sweeney, H. L., and A. Houdusse (2010), Annu. Rev. Biophys. **39**, 539.

Tailleur, J., and M. Cates (2008), Phys. Rev. Lett. **100** (21), 218103.

Takahashi, H., S. Umino, Y. Miki, R. Ishizuka, S. Maeda, A. Morita, M. Suzuki, and N. Matubayasi (2017), J. Phys. Chem. B **121** (10), 2279.

Takaki, R., M. L. Mugnai, Y. Goldtzvik, and D. Thirumalai (2019), in preparation.

Tehver, R., and D. Thirumalai (2008), J. Mol. Biol. **377**, 1279.

Tehver, R., and D. Thirumalai (2010), Structure **18** (4), 471.

Thirumalai, D., and B. Ha (1998), *Theoretical and mathematical models in polymer research*, edited by A. Y. Grosberg (Academic Press, Boston).

Thirumalai, D., and C. Hyeon (2005), Biochemistry **44** (13), 4957.

Thirumalai, D., C. Hyeon, P. I. Zhuravlev, and G. H. Lorimer (2019), Chemical reviews **119** (12), 6788.

Thirumalai, D., and G. H. Lorimer (2001), Annu. Rev. Biophys. Biomol. Struct. **30** (1), 245.

Toba, S., T. M. Watanabe, L. Yamaguchi-Okimoto, Y. Y. Toyoshima, and H. Higuchi (2006), Proc. Natl. Acad. Sci. **103** (15), 5741.

Todd, M. J., G. H. Lorimer, and D. Thirumalai (1996), Proc. Natl. Acad. Sci. U. S. A. **93**, 4030.

Tominaga, M., and A. Nakano (2012), Frontiers in plant science **3**, 211.

Toyabe, S., T. Okamoto, T. Watanabe-Nakayama, H. Taketani, S. Kudo, and E. Muneyuki (2010), Phys. Rev. Lett. **104** (19), 198103.

Toyabe, S., T. Watanabe-Nakayama, T. Okamoto, S. Kudo, and E. Muneyuki (2011), Proc. Natl. Acad. Sci. **108** (44), 17951.

Tsygankov, D., A. W. R. Serohijos, N. V. Dokholyan, and T. C. Elston (2011), Biophys. J. **101** (1), 144.

Turina, P., D. Samoray, and P. Gräber (2003), The EMBO Journal **22** (3), 418.

Uemura, S., H. Higuchi, A. O. Olivares, E. M. De La Cruz, and S. Ishiwata (2004a), Nat. Struct. Mol. Biol. **11**, 877.

Uemura, S., H. Higuchi, A. O. Olivares, E. M. De La Cruz, and S. Ishiwata (2004b), Nat. Struct. Mol. Biol. **11** (9), 877.

Vale, R. D. (2003a), Cell **112** (4), 467.

Vale, R. D. (2003b), J Cell Biol **163** (3), 445.

Vale, R. D., and R. A. Milligan (2000), Science **288** (5463), 88.

Vale, R. D., T. S. Reese, and M. P. Sheetz (1985), Cell **42** (1), 39.

Van Kampen, N. (2007), *Stochastic processes in physics and chemistry* (Elsiver).

Veigel, C., S. Schmitz, F. Wang, and J. R. Sellers (2005), Nature cell biology **7** (9), 861.

Veigel, C., F. Wang, M. L. Bartoo, J. R. Sellers, and J. E. Molloy (2002), Nature cell biology **4** (1), 59.

Velankar, S. S., P. Soultanas, M. S. Dillingham, H. S. Subramanya, and D. B. Wigley (1999), Cell **97** (1), 75.

Venkatesan, M., L. Silver, and N. Nossal (1982), J. Biol. Chem. **257** (20), 12426.

Visscher, K., M. J. Schnitzer, and S. M. Block (1999), Nature **400** (6740), 184.

Vu, H. T., S. Chakrabarti, M. Hinczewski, and D. Thirumalai (2016), Phys. Rev. Lett. **117** (7), 078101.

Wagoner, J. A., and K. A. Dill (2016), J. Phys. Chem. B **120** (26), 6327.

Wagoner, J. A., and K. A. Dill (2019), Proc. Natl. Acad. Sci. **116** (13), 5902.

Walker, M. L., S. A. Burgess, J. R. Sellers, F. Wang, J. A. Hammer III, J. Trinick, and P. J. Knight (2000), Nature **405** (6788), 804.

Walter, W. J., V. Beránek, E. Fischermeier, and S. Diez (2012), PLoS One **7** (8), e42218.

Wang, J., C. Herman, K. Tipton, C. Gross, and J. Weissman (2002), Cell **111** (7), 1027.

Wang, Q., D. Jana, M. Diehl, M. S. Cheung, J. Onuchic, and A. Kolomeisky (2018), Proc. Natl. Acad. Sci. **115**, 10052.

Wang, X., X. Zhang, D. Wu, Z. Huang, T. Hou, C. Jian, P. Yu, F. Lu, R. Zhang, T. Sun, *et al.* (2017), eLife **6**, e23908.

Wong, I., and T. M. Lohman (1992), Science **256** (5055), 350.

Yildiz, A., J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin (2003), Science **300** (5628), 2061.

Yildiz, A., H. Park, D. Safer, Z. Yang, L.-Q. Chen, P. R. Selvin, and H. L. Sweeney (2004a), J. Biol. Chem. **279** (36), 37223.

Yildiz, A., M. Tomishige, A. Gennerich, and R. D. Vale (2008), Cell **134** (6), 1030.

Yildiz, A., M. Tomishige, R. D. Vale, and P. R. Selvin (2004b), Science **303** (5658), 676.

Yu, J., T. Ha, and K. Schulten (2006), Biophys. J. **91** (6), 2097.

Zhang, Y., and A. B. Kolomeisky (2018), J. Phys. Chem. B **122** (13, SI), 3272.

Zhang, Z., Y. Goldtzvik, and D. Thirumalai (2017), Proc. Natl. Acad. Sci. **114** (46), E9838.

Zhang, Z., and D. Thirumalai (2012), Structure **20** (4), 628.

Zimmermann, E., and U. Seifert (2012), New J. Phys. **14** (10), 103023.