Structural dynamics of isolated myosin motor domains encode differences in their mechanochemical cycles

Justin R. Porter,[1] Artur Meller,[1] Maxwell I. Zimmerman,[1] Michael J. Greenberg,[1] Gregory R. Bowman[1,2]

[1] Department of Biochemistry & Molecular Biophysics, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
[2] Center for the Science and Engineering of Living Systems, Washington University School in St. Louis, St. Louis, MO, USA

Classification: Biological sciences (biochemistry, biophysics and computational biology)

Keywords: energy landscapes, machine learning, Markov sate models, molecular dynamics, conformational heterogeneity

Author contributions: J.R.P., A.M., M.I.Z., M.J.G., and G.R.B. designed research, J.R.P., M.J.G., and A.M. performed research, J.R.P., A.M., and M.I.Z. built new analytic tools, J.R.P., A.M., M.I.Z., M.J.G., and G.R.B. analyzed data, and J.R.P., A.M., M.J.G., and G.R.B. wrote the paper.

Abstract

Myosin motor domains perform an extraordinary diversity of biological functions despite sharing a common mechanochemical cycle, which couples ATP hydrolysis to mechanical work on an actin filament. Diversity is achieved in part by modulating the thermodynamics and kinetics of the states in this cycle. However, it remains unclear how sequence encodes these differences as biochemically diverse motors are often nearly indistinguishable in high resolution crystal structures. We hypothesized that comparing the conformational distributions of isolated motor domains would reveal signatures of their biochemical differences. Our reasoning was that sequence changes that modulate a motor's preferences for different states of the mechanochemical cycle are also likely to have a systematic effect on the isolated motor's structural preferences. To test this hypothesis, we performed two milliseconds of all-atom molecular dynamics simulations with explicit solvent for twelve myosin motor domains (~700 residues each). Markov state models (MSMs) built from this data captured significant conformational heterogeneity not observed in existing crystal structures, particularly in the nucleotide-binding P-loop. As we hypothesized, the thermodynamics and kinetics of our models predict the experimental duty ratios and ADP release rates of unloaded motors. Importantly, our approach was successful even when a homology model was used as the starting point for simulations, suggesting our methods are applicable to any motor sequence. Therefore, we expect our methods to be of great utility for studying sequence-function relationships in myosin motors, both in health and disease, and likely other proteins.

Significance Statement

Myosin motor domains perform an extraordinary diversity of biological functions, ranging from muscle contraction to intracellular transport. Furthermore, mutation of these domains leads to a wide variety of diseases, including blindness, deafness, and heart disease. However, it is currently impossible to predict how sequence variations impact a motor's function even given a high-resolution crystal structure. Here, we demonstrate that computer simulations of the motions of isolated myosin motor domains successfully predict biochemical properties, such as ADP release rates and duty ratios. This achievement indicates that motors' functional diversity is encoded in their conformational distributions, rather than any single conformation. Furthermore, it opens new opportunities to understand sequence-function relationships and obtain mechanistic insight that may inform the development of new therapies.

Introduction

Myosin motors (Figure 1A) perform an extraordinary diversity of biological functions despite sharing a common mechanochemical cycle. For example, myosin-II motors power muscle contraction, whereas myosin-V motors engage in intracellular transport. This diversity is in part due to differences in myosins' tails and light chain-binding domains, which influence properties like localization and multimerization (1). However, some of this diversity is encoded in the motor domains themselves (2). These differences stem from variations in the tunings of the thermodynamics and kinetics of the individual steps of the myosins' conserved mechanochemical cycle, which couples ATP hydrolysis to actin binding and the swing of a lever arm (3).

Two important and highly variable parameters for motor function are the rate of ADP release, which sets the speed of movement along actin, and the duty ratio, which is the fraction of time a myosin spends attached to actin during one full pass through its mechanochemical cycle. For example, in muscle, myosin-II motors are arranged into multimeric arrays called thick filaments and the individual motors typically have a strong preference for the actin free state (i.e., low duty ratio). These motors quickly detach after pulling on the actin filament to avoid creating drag for other motors in the array, much as a rower quickly removes their oar from the water to minimize drag. In contrast, individual myosin-Va motors have high duty ratios (i.e. prefer the actin-bound state), helping them to processively walk along actin filaments in intracellular transport. Similarly, the speed of myosin movement along actin (in the absence of opposing forces) is set by the rate of ADP dissociation (3), and it varies by four orders of magnitude from ~0.4 s^{-1} for non-muscle myosin-IIb (4) to >2800 s^{-1} for myosin-XI (5).

Unfortunately, inferring the relationship between a motor's sequence and its biochemical properties is not trivial. For example, one cannot simply predict the duty ratio or ADP release rate of a motor based on phylogeny. Myosin-V family members contain both high duty ratio motors, like myosin-Va, (6) and low duty ratio motors, like myosin-Vc (7). Similarly, ADP release rates within the myosin-II family vary from ~0.4 s^{-1} (non-muscle myosin-IIb) (4) to >400 s^{-1} (extraocular myosin-II) (8, 9). Insertions and deletions in the myosin motor domain sequence also convey useful, but typically incomplete, information. For instance, pioneering biochemical work (10) demonstrated a correlation between the length of loop 1 and ADP release rates in myosin-II motors. However, this observation does not explain how other myosin isoforms that have virtually the same loop 1 lengths have ADP release rates that differ by an order of magnitude (11). It is also difficult to predict the effects of mutations implicated in human disease, as the effects cannot be easily predicted from the location of the mutation. For example, an A223T mutation in human β-cardiac myosin causes a dilated cardiomyopathy (12) while an I263T mutation has the opposite effect, resulting in a hypertrophic cardiomyopathy (13), despite being separated by less than 6 Å (14).

Structural studies have provided detailed pictures of many key states in the mechanochemical cycle, but have yet to enable the routine prediction of a motor's biochemical properties from

its sequence. For example, high-resolution structures have illuminated many shared features of myosin motor domains, such as the lever arm swing (15) and conformational rearrangements associated with changes in nucleotide binding (16, 17). They have also revealed the strain-sensing elements of myosin-I motors (18–20) and the binding modes of many small molecules (14, 21, 22). However, the structures of motor domains with vastly different biochemical properties are often nearly indistinguishable. Similarly, computer simulations have begun to reveal aspects of motor function (23–26). However, simulating an individual motor domain (~700 residues) is a huge computational expense, so most simulation studies have been based on less than a microsecond of data. Thus, adding binding partners like actin to simulate the full mechanochemical cycle and infer properties like duty ratio is currently infeasible, especially if one wanted to compare multiple isoforms to infer sequence-function relationships.

Here, we investigate the possibility that the distribution of structures that an isolated motor domain explores correlates with its biochemical properties, allowing the prediction of sequence-function relationships. This hypothesis was inspired by a growing body of work showing that protein dynamics encode function (27, 28), even in the absence of relevant binding partners (29–31). In the case of myosin, we reasoned that as sequence changes modulate motors' preferences for different states of the mechanochemical cycle, they likely also have a systematic effect on the distribution of conformations explored by the motor, even in the absence of binding partners. Therefore, comparing the distribution of conformations that isolated motor domains sample in solution should reveal signatures of their biochemical differences.

To test this hypothesis, we ran an unprecedented two milliseconds of all-atom, explicit solvent molecular dynamics (MD) simulations of twelve myosin motors with diverse but well-established biochemical properties (Figure 1B, Table S1 and S2). Such simulations are adept at identifying excited states, which are lower-probability conformational states that are often invisible to other structural techniques. Indeed, our simulations reveal a surprising degree of conformational heterogeneity, particularly in the highly conserved P-loop (or Walker A motif), a common structural element for nucleotide binding that is highly conserved across myosin motor domains (32). Because of its high conservation, we reasoned that the P-loop would report on the conformation of the nucleotide binding site while still being comparable between motors with otherwise differing sequences. To enable quantitative comparisons, we constructed Markov state models (MSMs) from the MD data for each motor. MSMs are network models of protein free energy landscapes composed of many conformational states and the probabilities of transitioning between these states. They are a powerful means to capture phenomena far beyond the reach of any individual simulation by integrating information from many independent trajectories (33, 34). Analyzing our MSMs, we find they capture sufficient information about myosin motor domains' thermodynamics and kinetics to produce reasonable estimates of duty ratio and ADP release rates. Thus, MD and MSMs constitute a powerful platform for identifying relationships between the sequence of individual motor domains and their mechanochemical cycles.

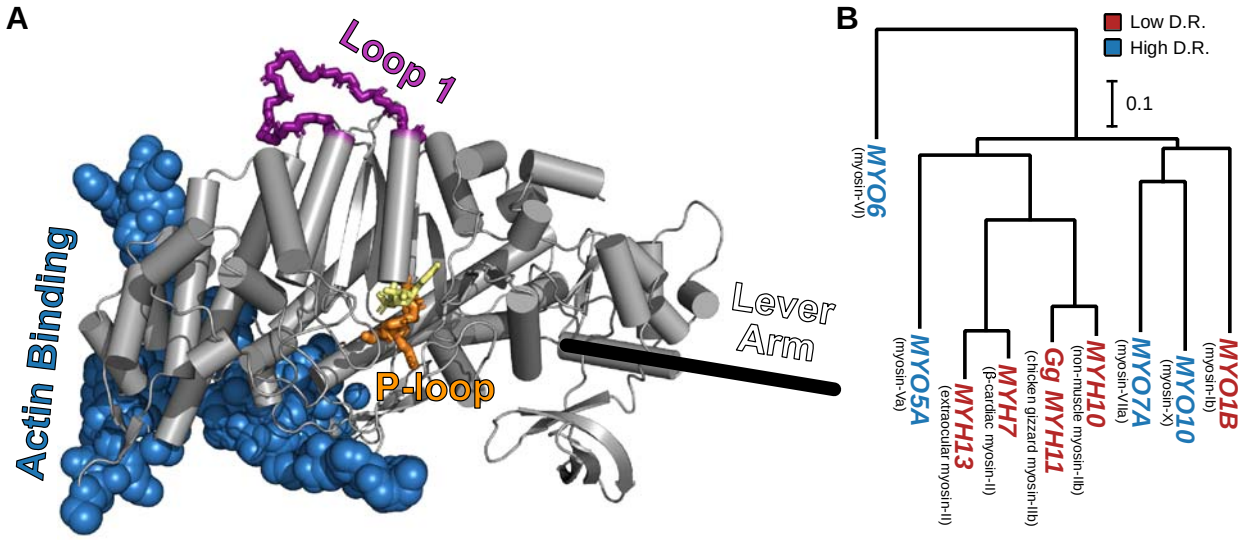


Figure 1: The conserved myosin motor domain fold across a diverse phylogeny of motors. (A), A crystal structure (PDB ID 4PA0) (22) of *Homo sapiens* β-cardiac myosin motor domain as an example of the conserved myosin motor domain fold. We note the structural elements most relevant to our work here (loop 1, in purple backbone sticks, and the P-loop, in orange sticks), along with the actin binding region (blue spheres). For orientation, we include the location of the lever arm (black line) and, to indicate the active site, the estimated location of ADP (yellow sticks). (B) The phylogenetic relationship the various myosin motor domains examined in this work. Except MYH11, all genes are from *Homo sapiens*. Gene names in blue indicate high duty ratio motors and red indicates low duty ratio. Common protein names are indicated as parentheticals to the left of each gene name. Phylogenetic relationships were inferred from the sequence of the motor domain using *k*-mer distances (35).

Results & Discussion

In simulation, the P-loop adopts conformational states that are rare in crystal structures.

We reasoned that any differences between myosin motor domains in nucleotide handling—ADP release rate or duty ratio, for instance—must somehow be manifest at the active site to have an effect. The P-loop is a highly conserved element of the myosin active site that plays an important role in interacting with the phosphates of the ATP substrate (36). Consequently, we reasoned that the P-loop would report on the conformation of the nucleotide binding site while still being comparable between motors whose sequences differ elsewhere in the protein. To assess the degree of conformational heterogeneity captured by crystal structures, we first analyzed structures deposited in the PDB (Figure 2A). We queried the PDB (37) for myosin motor domains (see *Methods*), yielding 114 crystal structures. Using sequence alignments (see *Methods*) we identified the P-loop in each of these models and computed the backbone root mean square deviation (RMSD) of each of these models to a reference structure (β-cardiac myosin, PDB ID 4PA0) (22). We found very little structural diversity among crystal structures, which rarely sample any conformations with P-loop backbone RMSD > 0.6 Å away (Figure 2A).

Then, to assess the capacity of the P-loop to adopt conformations not observed in crystal structures, we used molecular dynamics to simulate the myosin motor domain. These simulations of human β-cardiac myosin (*Hs MYH7*) were performed in the actin-free, nucleotide-free state for roughly a quarter-millisecond in all-atom explicit-solvent detail used to construct an MSM (see *Methods*). All simulations were conducted using the same force fields and conditions that we have previously used to analyze other systems' conformational distributions, including β-lactamases (31, 38, 39), *E. coli* catabolite activator protein (40), Ebola virus nucleoprotein (41), and G-proteins (42). Then, using the MSM, we computed the distribution of backbone RMSDs of the P-loop relative to the reference crystal structure.

In contrast to the relative uniformity among crystal structures, simulations revealed extensive conformational heterogeneity in the P-loop (Figure 2B). Where crystal structures rarely sampled conformations with RMSD >0.6 Å, in simulation we observe broad sampling (i.e. high-probability density) in regions from 0.2 Å RMSD all the way to ~1.5Å RMSD from the starting structure. Only 10 of 114 (9%) crystal structures' conformations were >0.6 Å RMSD from the reference conformation, whereas fully 58% of the distribution observed *in silico* is above 0.6Å RMSD from the reference conformation. These results suggest our simulations may provide mechanistic insight not previously accessible from crystal structures alone.

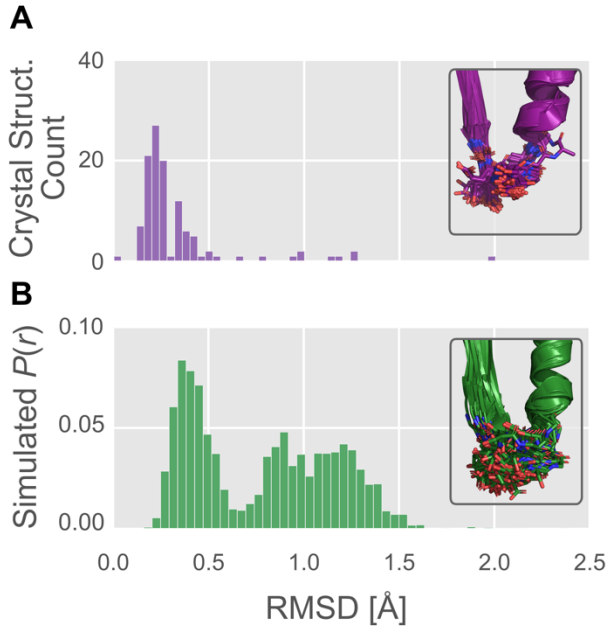


Figure 2: The P-loop conformational distribution observed *in silico* is substantially broader than that found in crystal structures. (A) P-loop conformations in the PDB are largely restricted to backbone RMSD ≤ 0.6 Å to a reference conformation (PDB ID 4PA0). *Inset*, the 114 myosin crystal structures superimposed, with the P-loop shown as sticks. (B) P-loop conformations from simulations of *Hs* β-cardiac myosin frequently explore conformations that are rare or not seen in crystal structures. *Inset*, the 114 most probable P-loop conformations extracted from our simulations of *Hs* β-cardiac myosin.

Simulations suggest that the nucleotide-free motor explores distinct nucleotide-favorable and nucleotide-unfavorable states.

We reasoned that P-loop conformations identified by our simulations might have important implications for motors' nucleotide handling. For example, modulating the relative probabilities of these conformations would provide a facile mechanism by which sequence variation might tune the mechanochemical cycle.

To assess the nucleotide compatibility of the P-loop conformations we observe in simulation, we sought to systematically compare these conformations with crystal structures with and without nucleotide. To do this, we built a map of P-loop conformational space using the dimensionality reduction algorithm Principal Components Analysis (PCA) to learn a low-dimensional representation of the pairwise interatomic distances between P-loop atoms that retains as much of the geometric diversity in the input as possible (see Figures S1-3, and *Methods* for details) (43). We then projected the states of our MSM built from our MYH7 simulations onto two principle components (PCs) to visualize the free energy surface sampled by our simulations (Figure 3A, green level sets). Using the same PCA, we then projected each crystal structure's P-loop conformation into this space, plotting each as a point (Figure 3A, points). Points labeled with PDB IDs represent crystal structures with P-loops >0.6 Å backbone RMSD away from the reference structure 4PA0 used above. We also classified each structure (see *Methods*) as nucleotide-bound (yellow points) or nucleotide-free (purple points). Then, we compared the frequency at which nucleotide-bound and nucleotide-free P-loop conformations were found in various conformations.

This analysis revealed two dominant conformational states that likely constitute nucleotide-favorable and nucleotide-unfavorable states (Figure 3A and B). Once the distribution of P-loop conformations is projected onto two PCs (the green level sets in Figure 3A), we observe two broad minima in the P-loop conformational landscape. We refer to these apparent minima as the upper and lower basin for brevity but recognize that other minima may exist and be obscured by the projection of a high-dimensional space into a low-dimensional space. The lower basin (<0.6 Å RMSD from the reference structure) contains 91% of crystal structures (104/114) and, because 80% (84/105) of these structures are bound to nucleotide, it is highly likely to represent a nucleotide-compatible conformation. In contrast, despite being populated roughly equally in simulation, regions outside the lower basin (≥0.6 Å RMSD) contain only 9% (10/114) of crystal structures. And, because only one (11%) of these structures is nucleotide bound, these regions are significantly depleted in nucleotide-bound structures (odds ratio = 0.03, $p < 1.3 \times 10^{-5}$ by Fisher's exact test), strongly implying that they are less or not at all nucleotide compatible. Interestingly, this single exception (PDB ID 2Y8I, *Dictyostelium discoideum* myosin-II G680V) is a highly perturbed motor that has been shown to have low ATPase activity, low motility and a disordered allosteric network (44, 45), potentially contributing to its aberrant conformation.

To characterize the structural differences between nucleotide-favorable and nucleotide-unfavorable states captured in the simulations, we coarse-grained our MSM into a model with

just five states, called A-E. We used hierarchical clustering to group the thousands of states explored by *Hs* MYH7 into five states based only on their P-loop conformations (see *Methods*). Then, using the assignment of each frame from our simulations to one of these five states, we fit a five-state MSM (Figure 3D, node sizes indicate equilibrium probabilities, arrow weights indicate transition probabilities). The most probable single state is the A state (49%), which encompasses the entire lower basin and, as we will see below, appears to form favorable interactions with nucleotide based on the conformation of the P-loop. The excited, apparently nucleotide-disfavoring conformations in the upper basin are split into 3 states, B-D, which together account for 50% of the equilibrium probability. Thus, β-cardiac myosin spends about equal time in nucleotide-favorable (state A) and nucleotide-unfavorable states (states B-D) in simulations. Finally, state E (1%, too low to be seen clearly in Figure 3A), involves a condensation of the P-loop into an extension of the HF helix, similar to the crystal structure 4L79 (18). The reduced number of states in this MSM allowed us to inspect a small number of high-probability conformations near the mean of each P-loop state, which we took as exemplars of each of the five P-loop states.

Comparing the states of our MSM reveals that the dominant geometrical difference between nucleotide-favorable and nucleotide-unfavorable P-loop states is the orientation of the peptide bond between S180 and G181 (Figure 3C). In the nucleotide-favorable state A (Figure 3D, lower right inset), the S180 backbone carbonyl (shown in pink sticks with a white arrow) is oriented away from the phosphates of the nucleotide, enabling the nucleotide to bind to the active site. In contrast, nucleotide-disfavoring states (labeled B-D in Figure 3D) orient the S180 backbone carbonyl toward the phosphate groups of the nucleotide. This positions the carbonyl oxygen in a way that appears to sterically clash with the phosphates of nucleotide. It also orients the negative end of the carbonyl bond's electric dipole toward the nucleotide binding site and the negatively charged phosphates of ADP and ATP. Taken together, our observations about the geometry of the excited, nucleotide-disfavoring state in the upper basin are consistent with a lowered capacity for nucleotide binding.

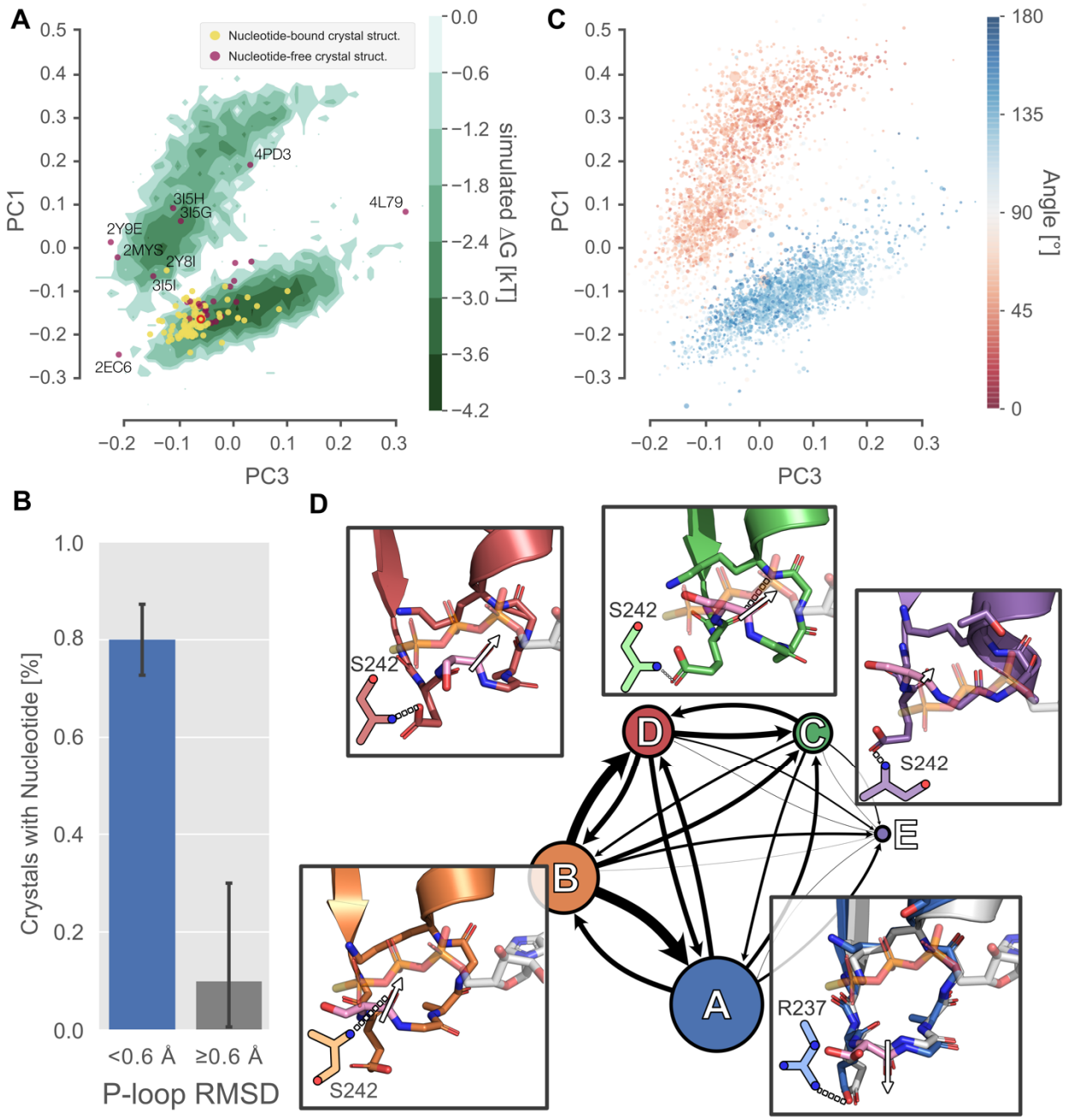


Figure 3: Excited P-loop states are less compatible with nucleotide than the states preferred in crystal structures. (A) The P-loop conformational space projected onto two principal components (PCs) reveals two distinct free energy basins (green level sets). Yellow and purple points represent crystal structures with and without ligand, respectively. Structures farther than 0.6 Å from the β-cardiac myosin structure (red empty circle) are labeled with their PDB ID. (B) Proximity to the β-cardiac myosin reference conformation is associated with the presence of a nucleotide in crystal structures ($p < 1.3 \cdot 10^{-5}$ by Fisher's exact test), suggesting that the ligand stabilizes the A state. Error bars represent the 95% confidence interval of 1000 bootstrap realizations. (C) The re-orientation of the S180 backbone carbonyl accounts for the

split between upper and lower basins. Points represent P-loop conformations from each state in the β-cardiac whole-motor MSM projected onto the same PCs as in panel A. Points are sized by their probability from the MSM and colored by the angle between the backbone carbonyl bond vectors of S180 and K184. (D) Center, each of the five states of the P-loop MSM are indicated as nodes in a network, sized by their equilibrium probability and connected by arrows with line width proportional to the transition probabilities between them. Surrounding the model, insets show example configurations of the P-loop in sticks colored to match the state they represent. State A is associated with a conformation of the S180 (pink sticks) carbonyl bond vector (white arrow) directed away from the nucleotide binding pocket, whereas states B-D are associated with the opposite orientation of the S180 backbone carbonyl bond vector. The A state conformation is the conformation found in most crystal structures. For reference, PDB 1MMA is shown in grey sticks and the crystallographic position of ATP is shown in semi-opaque grey sticks. For all states, important interactions with the Switch-I loop are shown as two-dimensional sketches for visual clarity. An interaction between R237 and E179 is specific to state A, whereas various interactions with S242 are indicative of other states (Figure S4).

The balance between nucleotide-favorable and nucleotide-unfavorable P-loop states predicts duty ratio.

We reasoned that motors with a higher probability of adopting nucleotide-favorable P-loop conformations in isolation are likely to have an increased affinity for nucleotide and, therefore, spend more time in nucleotide-bound states of the mechanochemical cycle. Our reasoning is that motors that prefer nucleotide-favorable P-loop conformations in isolation pay a lower energetic cost to adopting these same nucleotide-favorable conformations when they form a complex with nucleotide. Supporting this logic, it has been observed that, absent load, a large free energy difference between ADP-bound and nucleotide-free states is associated with a low duty ratio (46, 47). Thus, we hypothesized that a preference for the nucleotide-favorable A state should correlate with low duty ratio.

To test if differences in the probability of excited states encodes information about duty ratio, we simulated an additional seven myosin isoforms of differing duty ratio for a total of ~2 ms of aggregate simulation in all-atom, explicit solvent detail. Specifically, we simulated four human low duty ratio myosin motor domains (from myosin-II genes MYH13, MYH7, MYH10, and myosin-I gene MYO1B) and four human high duty ratio myosin motor domains (from genes MYO5A, MYO6, MYO7A, and MYO10), for between 125 and 325 μs each (see *Methods*). These motors were selected because extensive kinetic characterization (4, 6, 8, 11, 48–51) has revealed very diverse kinetic tuning, providing a robust test of our hypotheses. Because no crystal structure of the human sequence was available for any of these proteins except MYH7, homology models were built in each case and used as starting points for simulations (see *Methods* and Table S1). To allow for direct comparisons between motors, we used the same PCA and state definitions as described above for MYH7.

As expected, high duty ratio motors have a stronger *in silico* preference for nucleotide-favoring P-loop states than low duty ratio motors (Figure 4A). Figure 4A shows an example of this effect

on the P-loop conformational distributions of high duty ratio motor MYO6 and low duty ratio motor MYH7. The low duty ratio motor explores both upper and lower basins (Figure 4A, left) while the high duty ratio motor strongly prefers the lower basin (Figure 4A, right). Provocatively, when motors are crystallized without ligand, only motors with low unloaded duty ratios have been crystallized with P-loops outside the nucleotide-favorable conformation (Figure 4A, red and blue points). Of 29 unliganded crystal structures, 8/20 (40%) of low duty ratio motors' P-loops crystallized outside the A state, whereas 0/9 (0%) high duty ratio motors' P-loops crystallized outside state A ($p < 0.034$ by Fisher's exact test, see *Methods*).

Given this trend, we reasoned that the relative free energies of the nucleotide-favorable state and the nucleotide-disfavoring excited states would provide a useful predictor of a motor's duty ratio. We assigned every whole-motor MSM state to one of the five P-loop states and used these assignments to compute the free energies of each of the five states for each of the eight motors (see *Methods*). We then took the difference in free energy between states A and B, which are the two best sampled states and therefore give statistically robust results. Numerical values and references for these experimental values can be found in Table S3.

As expected, we find a strong correlation between motors' duty ratios and their preferences for the nucleotide-favorable A state over the nucleotide-unfavorable B state (Figure 4B). Specifically, high duty ratio motors have a strong preference for the A state (negative free energy difference) while low duty ratio motors spend more time in state B (positive free energy difference). Decreased stability of the nucleotide-favorable conformation in these low duty ratio motors could explain this observation.

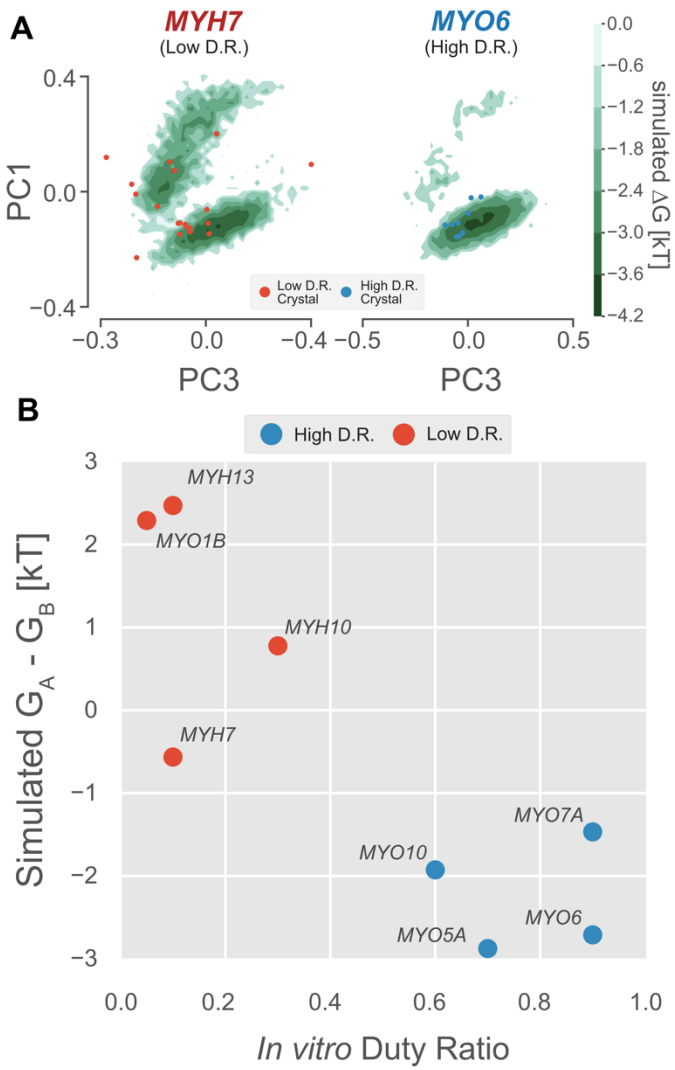


Figure 4: The free energy landscape of the P-loop encodes duty ratio. (A) Free energy landscapes in the PC1/PC3 plane demonstrate that the upper basin is well sampled by an example low duty ratio motor (MYH7, left) and poorly sampled by an example high duty ratio motor (MYO6, right). Ligand-free crystal P-loop conformations from high and low duty ratio motors are shown as blue and red points, respectively. (B) Experimental duty ratio (*x*-axis) is correlated with the simulated free energy difference between nucleotide favorable and nucleotide-unfavorable states (*y*-axis, more negative values mean higher probability of the nucleotide-favorable A state).

Simulations predict ADP release rates better than loop 1 length does by capturing sequence-specific effects.

Because ADP release allows a motor to adopt nucleotide-incompatible P-loop conformations, we reasoned that the rate at which a motor can transition to these conformations *in silico* might correlate with *in vitro* ADP release kinetics. While we expect a correlation, we acknowledge that the absolute rates will almost certainly differ, since the rates themselves

likely differ in the presence and absence of nucleotide. To test for a correlation, we first focus on data sets that examine several motors under the same experimental conditions. Identical conditions are important because *in vitro* biochemical rates depend strongly on experimental conditions such as salt and temperature (49, 52, 53). We focus on low duty ratio motors, since their frequent transitions to nucleotide-unfavorable states make it possible to estimate their transition rates with confidence. In contrast, in high duty ratio motors, transitions between these states are sufficiently rare that their rates cannot be estimated with confidence.

An especially useful dataset for comparing relative ADP release rates was created by Sweeney *et al* (10), which carefully dissected the effect of variation in loop 1 length and sequence on ADP release rates using the same experimental conditions. These authors established a positive relationship between loop 1 length and ADP release rate using engineered constructs of chicken gizzard myosin-II (shown in Figure 5A, henceforth *Gg* MYH11). A notable exception, however, was the myosin with wild-type loop 1, which had an ADP release rate more than three times faster than predicted by the length-based model (Figure 5B). This deviation from a purely length-driven ADP release rate led these authors to hypothesize that there must also be sequence-specific effects of loop 1 on ADP release rate. They then identified an alanine mutant that ablated the sequence-specific effects of the wild-type loop (henceforth *Gg* MYH11-ala).

To assess the capacity of *in silico* P-loop kinetics to capture the experimentally measured ADP release rates in the constructs investigated by Sweeny *et al*, we simulated and analyzed four Gg MYH11 constructs. These constructs are a subset of the variants considered by Sweeny *et al*. We selected the wild-type loop (*Gg* MYH11-wt) because it was the primary outlier in their length-only model. We selected the alanine mutant (*Gg* MYH11-ala) because it, with just 5 mutations, shifted the wild type loop in line with the length-only model proposed by Sweeny *et al*. Then, we selected the extreme points that were well fit by the loop length-only model: the loop 1 deletion (*Gg* MYH11-Δloop1) and the construct using the loop 1 from *Xenopus* non-muscle myosin (*Gg* MYH11-xeno). We simulated these four constructs for 6-16 μs each beginning from a homology model (see *Methods* and Table S1) and built whole-motor MSMs which, as before, were used to compute five-state P-loop MSMs. Each P-loop MSM contains a parameter P(A→B) which captures the probability that a conformation in state A transitions to state B within a fixed period of time (known as the lag time of the model). We then compared P(A→B) to ADP release rates measured *in vitro* for these four constructs.

As expected, there is a strong positive relationship (Pearson's R=0.99) between the P(A→B) fit by our MSMs and *in vitro* ADP release rate (Figure 5C). This is stronger than the equivalent correlation for the length-based model (Pearson's R=0.72). Importantly, the rank order of the four isoforms is correct, whereas using a loop 1 length-only model dramatically underestimates the ADP release rate for the wild-type motor. Together, the fact that the sequence change is small (only five residues differ between wild type and the alanine mutant) and the change is distant (~25Å) from the P-loop indicate that our model is exquisitely sensitive to sequence, even at sites distant from the active site.

P-loop kinetics *in silico* correlate with ADP release rates across conditions.

To further assess the generalizability of our model, we considered several additional datasets that relax constraints placed on data sets in the previous section. First, we relaxed the constraint that motors differ by just one structural element (loop 1). Specifically, we considered several skeletal myosin isoforms, including MYH7 and MYH13 that Johnson *et al* (35) studied under the same conditions (Figure 5D and E, yellow points). These motor domains are an interesting case because, at 80% sequence identity, their sequences differ much more than Sweeney *et al*'s constructs, and these differences are distributed throughout the protein. Crucially, and despite having roughly the same loop 1 length, their ADP release rates differ by about an order of magnitude (59 s^{-1} vs 400 s^{-1}). Owing to the fact that Johnson *et al*'s data were collected under different experimental conditions than Sweeny *et al's* data (5 mM $MgCl_2$ at 25 °C vs 1 mM $MgCl_2$ at 20 °C with different light chains), we only expect a general trend to hold, since motors' properties are very sensitive to magnesium, temperature, and light chain identity (49, 53, 54). Second, we assessed the trend in two human non-muscle motor domains, MYO1B and MYH10 with measurements carried out under different conditions. Notably, because they both release ADP very slowly, they test our model's capacity to evaluate very slow ADP release rates.

Consistent with our expectations, and despite the diverse experimental conditions, we still observe a reasonable correlation between P(A→B) and ADP release across all data sets (Figure 5E, Pearson's R = 0.75). This dramatically improves on the length-based model (Pearson's R = 0.14). Importantly, under the matched experimental conditions for MYH7 and MYH13 we still find the correct order of ADP release rates (Figure 3C, yellow points), suggesting that this method generalizes well to the larger phylogenic distances between myosin isoforms. Furthermore, MYO1B and MYH10 are correctly identified as very slow releasers of ADP, although the point estimates appear to be quite noisy. MYH10 is known to be exquisitely sensitive to light chains (54), so it is not surprising that it is one of the greatest outliers given that we did not include these in our simulations.

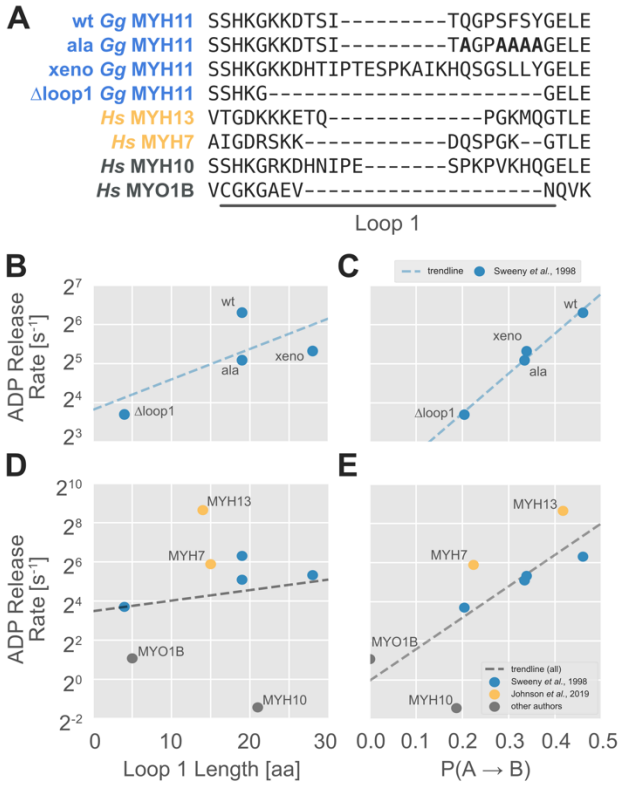


Figure 5: The probability of transitioning from nucleotide-favorable to nucleotide- unfavorable P-loop conformations (P(A→B)) predicts ADP release rates for motors with low duty ratios. (A) Loop 1 sequences and lengths considered in this work. Residues mutated to alanine in the wild-type chicken gizzard MYH11 (wt *Gg* MYH11) are bolded in the appropriate row. (B) For the Sweeney dataset, there is a moderate relationship between loop 1 length and ADP release rate (Pearson's R = 0.75) but, (C) there is a much stronger correlation between P(A→B) and ADP release rate (Pearson's R = 0.99). (D) Across all datasets, the relationship between loop 1 length and ADP release rate is weak (Pearson's R = 0.14), and (E) there is a much stronger correlation between P(A→B) and ADP release rate (Pearson's R = 0.75).

Conclusions

In this work, we used computer simulations of isolated myosin motor domains to predict the *in vitro* ADP release rate and duty ratio of unloaded myosin motors. To do this, we identified systematic shifts in the distribution of conformations that a motor explores that correlate with changes in biochemistry, rather than by directly simulating the biochemical processes themselves, which would have been prohibitively expensive. While binding partners (actin and nucleotide, for instance) and structural elements outside the motor domain almost certainly affect the distribution of conformations, our results demonstrate that it is nevertheless possible to extract reasonable estimates for at least some unloaded biochemical properties from only the isolated motor domain's conformational distribution. The ability of the isolated motor domain's fluctuations to predict these parameters likely stems from a link between the isolated and bound conformational distributions. In other words, because the motor domain active site must adopt certain key conformations during its functional interactions with binding

partners (i.e., nucleotide and actin), it is nearly guaranteed to at least transiently sample those conformations even in the absence of those binding partners. Importantly, our simulations only require a reasonable homology model as a starting point, so our methods should be applicable to a broad range of motor variants, including mutations implicated in disease.

Given the high degree of structural conservation of the myosin motor domain, it was not previously possible to directly predict the duty ratio or kinetics for a given myosin isoform from the sequence or structure of a motor domain alone. Our studies demonstrate that the duty ratio and the rate of ADP release are not captured by a single structural element, but rather by the distribution of conformations that the motor explores in solution. Throughout our simulations, we observed that the distribution of P-loop conformations is sensitive to relevant sequence changes, both large and small, throughout the myosin motor domain. Presumably, these changes are allosterically propagated through the myosin motor domain through complex networks of coupled motions. Thus, capturing the difference between the wild-type and alanine-substituted chicken gizzard myosins (Figure 5C), for instance, required the model to capture the allosteric perturbation induced by a change of a few dozen atoms in a molecule of ~12,500 atoms at a distance of ~25 Å (Figure 1A). Meanwhile, classifying the duty ratio of diverse myosin motors requires the P-loop to integrate signals from across the molecule into a single overall conformational preference. This underscores a key advantage of physics-based simulations, which is the ability to represent these allosteric networks by modeling in detail the complex, nonlinear couplings throughout the molecule.

One tantalizing interpretation of the excited states of the P-loop we observe *in silico* is that they may be related to the biochemically-observed "open" and "closed" states that nucleotide-free myosin motors populate *in vitro* (55). In our simulations, we see that the P-loop fluctuates between conformations that are nucleotide-compatible and conformations that probably are not. In biochemical experiments, at least some myosin isoforms in the nucleotide-free actin-bound state fluctuate between a state that binds nucleotide and a state that does not. It has also been shown that the equilibrium between these two biochemical states (K_α), correlates with duty ratio and the transition rate from the nucleotide binding incompetent state to the nucleotide binding competent state ($k_{+\alpha}$) correlates with the ADP release rate (46). Similarly, we showed that the equilibrium between nucleotide-favorable and nucleotide-disfavorable conformations predicted duty ratio, while the rate of transition predicted ADP release rate. A simple explanation for these similarities is that there may be a correspondence between these biochemical states and the structural states that we observe in our MSMs *in silico*.

Finally, our results highlight the general capacity of computational modeling to link sequence and function. One immediate application of our work here is to estimate *in silico* the biochemical parameters of new or difficult-to-study myosins. In the near term, constructing such models could help us learn more about the atomic basis for healthy functional diversity in myosin motors, and how small changes can give rise to malfunction and disease. Indeed, in the coming years it may prove possible to use these models as a tool for studying patient-specific mutations by understanding the atomic basis for diseases caused by dysfunction of myosin motors or to aid in developing therapeutics. Finally, because we find no reason to believe our

approach's applicability is limited to myosin motors, we expect the techniques we have presented here to be of use for any protein where the physics that maps sequence to biochemistry is not straightforward.

Materials & Methods

Preparation of homology models

For simulations, the initial structure of each myosin motor domain was prepared by first obtaining the full-length protein's sequence from PubMed Protein, trimming the sequence down to include only the motor domain using crystal structure 4PA0 of MYH7 as a guide, and submitting that sequence to SWISS-MODEL for homology modeling (56). Templates were chosen with a preference for those that were high-resolution, high sequence similarity, and in the rigor state. A complete list of sequences, templates, and motor domains can be found in table S1.

Preparation of example myosin conformation

In Figure 1A, the position of ATP is based on ligand-bound crystal structure 1MMA (36). The actin binding region was defined by all atoms within 10 Å of the actin filament after alignment to 6BNP chain K (57).

Sequence alignments

All sequence alignments were performed with MUSCLE 3.8.1551 (58) using default parameters. Phylogenetic trees were inferred with the neighbor joining method using these alignments. Distances between sequences were *k*-mer distances (35).

Molecular dynamics simulations

For each protein simulated, the following procedure was used to prepare the simulation. The protein structure was solvated in a dodecahedron box of TIP3P water (59) that extended 1 nm beyond the protein in every dimension. Thereafter, sodium and chloride ions were added to produce a neutral system at 0.1 M NaCl. The proteins were minimized and briefly equilibrated with all atoms restrained in place. Molecular dynamics were performed using GROMACS (60, 61) using the AMBER03 force field (62). Details of minimization, equilibration, and trajectories can be found in Supplementary Materials.

Production simulations were performed on a mixture of Folding@home (63) and an in-house supercomputing cluster (see Supplementary Materials).

Markov state models

Fine-grain, whole-motor domain Markov state models were constructed first by defining microstates using the *k*-hybrid clustering algorithm with five rounds of *k*-medoids refinement using the Euclidean distance between residue sidechain solvent accessible surface area (scSASA) as a distance metric. This approach first appeared in Porter, *et al*, (31) and was

chosen because it scales well for extremely large datasets compared to traditional RMSD clustering. The reasons for this are discussed in ref. (64) but, briefly, although scSASA calculations are initially expensive, they realize substantial performance gains in clustering because each frame's scSASA need only be computed once. ach frame can be computed independently, allowing for massive parallelization. It also reduces the size of the input data size, since only a single floating point number represents an entire residue, and allows the use of a cheaper distance metric (Euclidean distance rather than RMSD).

Markov state models were then fit for each variant by applying a $1/n$ pseudocount to each element of the transition counts matrix and row-normalizing, as recommended in Zimmerman, *et al* (65). Lag times were chosen by the implied timescales test and by examining the equilibrium probability distribution for unrealistically overpopulated states (suggesting insufficient sampling of a particular transition or internal energy barriers). Important hyperparameters are listed in Table S2.

Fitting coarse-grained P-loop MSMs used the same procedure, but assignments based on P-loop state were used, rather than assignments to whole-motor SASA states. $P(A \rightarrow B)$ is a parameter of these MSMs. In all cases for coarse-grained P-loop MSMs, a lag time of 37.5 ns was used.

Clustering and Markov state model routines are implemented in enspara, git revision f874ba. Solvent accessibility, atomic distance, and RMSD calculations were performed with MDTraj (66).

We made extensive use of jug (67) and GNU Parallel (68) for task-level parallelization and management of dependencies between tasks.

Construction of the P-loop free energy surface

Pairwise interatomic distances in the P-loop were computed using MDTraj (66), selecting all possible pairs of a backbone amide nitrogen and a backbone carbonyl oxygen atom in the GESGAG portion of the Walker A motif (*i.e.*, the conserved P-loop sequence) that makes up the P-loop.

Principle components analysis (PCA) was performed on the 36-dimensional pairwise atomic distance vectors for each MSM microstate using the PCA implementation in sklearn (69). No whitening was employed and the full SVD was calculated.

The surface was then estimated by constructing a weighted two-dimensional histogram in the PC1/PC3 plane with 50 bins between the minimum and the maximum data in each direction. The resulting array of probabilities was then converted into free energies of units kT by taking the natural logarithm of each value. It was then convoluted with a gaussian of variance 0.3 per grid cell using scipy's gaussian_filter method (70). The resulting array was then level-set into six level sets.

Selection of myosin motor domain PDB crystal structures

We selected crystal structures to map on to the P-loop free energy landscape by querying the PDB (37) for all structures with sequence identities to the motor domain of *Hs MYH7* greater than 10%, resolution <= 5.0 Å and a BLAST E-value less than 10^{-10}. We then selected the largest chain in each crystal structure, used muscle (58) to align that chain's sequence to the motor domain of *Hs MYH7*, and used the resulting alignment to identify the P-loop. P-loop distances were computed and projected into the low-dimensional space as described above. Sequence bookkeeping and I/O relied heavily on scikit-bio (github.com/biocore/scikit-bio).

Crystal structures were classified as bound to a nucleotide or nucleotide analogue if they contained a residue with the name ADP, ATP, ANP, MNQ, MNT, ONP, PNQ, DAE, DAQ, NMQ, AGS, AD9, AOV, or FLC.

Hierarchical clustering of the P-loop

The five coarse-grained MSM microstates for MYH7 were learned using agglomerative clustering on the four-dimensional P-loop features learned by PCA for the free energy surface. Ward linkage and a Euclidean distance metric were used. Briefly, the states are recursively combined in a way that minimizes the within-cluster variance in a until the specified number of clusters is reached. The number of clusters were increased until no obvious internal free energy barriers were seen in the four PC dimensions. Agglomerative clustering was implemented by sklearn 0.21.2 (69).

Assignment of new conformations to P-loop states

P-loop state assignments for conformations of motors other than *Hs* MYH7 were made using a *k*-nearest neighbors (69) approach. In this approach, a query conformation is assigned to a cluster based on the assignments of nearest *k* points in the labeled dataset (*i.e.* MYH7). In other words, the nearest *k* points to the query point "vote" on the assignment of the query point to a cluster. In our case, *k* was 5, but we did not appreciate any differences for values of *k* from 3 to 15.

Implementation of *k*-nearest neighbors was from sklearn 0.21.2. A ball tree was used to speed the search for neighbors (71).

Estimation of equilibrium probability of P-loop states

For each motor, the probability of a P-loop state was calculated by summing the equilibrium probabilities of all states in the whole-motor MSM assigned to that P-loop state.

Duty Ratios of Crystal Structures

While many myosin motors' duty ratios have been well characterized, some constructs' unloaded duty ratio have not been measured. For these motors, it was therefore necessary to infer duty ratios from phylogeny. For our analysis of duty ratio and P-loop crystal position, we considered only the 29 ligand-free structures, namely: 4DBP, 2MYS, 3I5H, 2Y0R, 2BKH, 6I7D, 1DFK, 1OE9, 3I5I, 2OS8, 4P7H, 5V7X, 4ZLK, 1MNE, 1FMV, 2AKA, 3MYL, 2EC6, 4L79, 3L9I, 2BKI, 2Y9E, 1KK7, 1W8J, 2X51, 4PA0, 4PD3, 3I5G, and 1SR6. Based upon previous biochemical experiments (6, 7, 48, 72, 73), myosin-Is and IIs were assumed to have low duty ratios. Myosin-VIs were assumed to have high duty ratio. Myosin-Va and Vb from all organisms were assumed to have high duty ratios and Myosin-Vc was assumed to have a low duty ratio. *Plasmodium falciparum* MyoA (6I7D) has been shown to have a high duty ratio (74).

Myosin class was inferred as follows. Where a roman numeral was given in the PDB description (e.g. Myosin-II) this classification was used. Otherwise, if "muscle" or "striated" was appeared in the PDB polymerDescription field, the myosin was classified as a myosin-II. Finally, in the absence of other indicators, myosins from *Doryteuthis pealeii*, *Placopecten magellanicus*, and *Argopecten irradians* were classified as Myosin-IIs, and myosins from *Plasmodium falciparum* were classified as Myosin-XIVs.

Visualization

Proteins structures were visualized and rendered with PyMOL. Data plots were constructed with matplotlib (75). Free energy surface colormaps were constructed with the cubehelix color system (76).

Code and model availability

MSMs and starting conformations for each of the myosin constructs studied in this have been uploaded to the Open Science Framework as project ID 54G7P, along with the parameters for the PCA used in Figures 2 and 3. This OSF project also includes a CSV that lists the P-loop definition, P-loop RMSD from the reference state, and assignment to P-loop state A-E for each crystal structure. This data repository will be made public upon the publication of this manuscript.

Acknowledgements

We are extremely grateful to the citizen scientists of Folding@home for their generous donation of computing resources. We are also grateful to Prof. Eric Galburt, Prof. John Edwards, and Dr. Joshua Alinger for their insight and helpful comments about this work. We are also grateful to the Center for High Performance Computing at the Mallinkrodt Institute for Radiology for computer time.

This work was funded by National Institutes of Health grants R01GM12400701 (G.R.B.), R01HL141086 (M.J.G.), T32GM02700 (A.M.), and F30HL146052 (J.R.P.), National Science Foundation CAREER Award MCB-1552471 (GRB), G.R.B. holds a Career Award at the Scientific

Interface from the Burroughs Wellcome Fund and a Packard Fellowship for Science and Engineering from the David & Lucile Packard Foundation (G.R.B.). M.I.Z. was supported in part by a Monsanto Graduate Fellowship and a Center for Biological Systems Engineering Fellowship.

References

1. M. Krendel, M. S. Mooseker, Myosins: tails (and heads) of functional diversity. *Physiol. Bethesda Md* **20**, 239–251 (2005).

2. M. J. Greenberg, G. Arpağ, E. Tüzel, E. M. Ostap, A Perspective on the Role of Myosins as Mechanosensors. *Biophys. J.* **110**, 2568–2576 (2016).

3. E. M. De La Cruz, E. M. Ostap, Relating biochemistry and function in the myosin superfamily. *Curr. Opin. Cell Biol.* **16**, 61–67 (2004).

4. A. Nagy, *et al.*, Kinetic characterization of nonmuscle myosin IIb at the single molecule level. *J. Biol. Chem.* **288**, 709–722 (2013).

5. K. Ito, *et al.*, Kinetic mechanism of the fastest motor protein, Chara myosin. *J. Biol. Chem.* **282**, 19534–19545 (2007).

6. E. M. De La Cruz, A. L. Wells, S. S. Rosenfeld, E. M. Ostap, H. L. Sweeney, The kinetic mechanism of myosin V. *Proc. Natl. Acad. Sci.* **96**, 13726–13731 (1999).

7. Y. Takagi, *et al.*, Human myosin Vc is a low duty ratio, nonprocessive molecular motor. *J. Biol. Chem.* **283**, 8527–8537 (2008).

8. M. J. Bloemink, J. C. Deacon, D. I. Resnicow, L. A. Leinwand, M. A. Geeves, The superfast human extraocular myosin is kinetically distinct from the fast skeletal IIa, IIb, and IId isoforms. *J. Biol. Chem.* **288**, 27469–27479 (2013).

9. C. A. Johnson, *et al.*, The ATPase cycle of human muscle myosin II Isoforms: adaptation of a single mechanochemical cycle for different physiological roles. *J. Biol. Chem.*, jbc.RA119.009825 (2019).

10. H. L. Sweeney, *et al.*, Kinetic Tuning of Myosin via a Flexible Loop Adjacent to the Nucleotide Binding Pocket. *J. Biol. Chem.* **273**, 6262–6270 (1998).

11. J. C. Deacon, M. J. Bloemink, H. Rezavandi, M. A. Geeves, L. A. Leinwand, Identification of functional differences between recombinant human α and β cardiac myosin motors. *Cell. Mol. Life Sci. CMLS* **69**, 2261–2277 (2012).

12. Z. Ujfalusi, *et al.*, Dilated cardiomyopathy myosin mutants have reduced force-generating capacity. *J. Biol. Chem.* **293**, 9017–9029 (2018).

13. F. Tesson, *et al.*, Genotype-phenotype analysis in four families with mutations in beta-myosin heavy chain gene responsible for familial hypertrophic cardiomyopathy. *Hum. Mutat.* **12**, 385–392 (1998).

14. V. J. Planelles-Herrero, J. J. Hartman, J. Robert-Paganin, F. I. Malik, A. Houdusse, Mechanistic and structural basis for activation of cardiac myosin force production by omecamtiv mecarbil. *Nat. Commun.* **8**, 190 (2017).

15. S. Fischer, B. Windshügel, D. Horak, K. C. Holmes, J. C. Smith, Structural mechanism of the recovery stroke in the myosin molecular motor. *Proc. Natl. Acad. Sci.* **102**, 6873–6878 (2005).

16. I. Rayment, *et al.*, Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science* **261**, 50–58 (1993).

17. P.-D. Coureux, H. L. Sweeney, A. Houdusse, Three myosin V structures delineate essential features of chemo-mechanical transduction. *EMBO J.* **23**, 4527–4537 (2004).

18. H. Shuman, *et al.*, A vertebrate myosin-I structure reveals unique insights into myosin mechanochemical tuning. *Proc. Natl. Acad. Sci.* **111**, 2116–2121 (2014).

19. M. J. Greenberg, T. Lin, H. Shuman, E. M. Ostap, Mechanochemical tuning of myosin-I by the N-terminal region. *Proc. Natl. Acad. Sci.* **112**, E3337–E3344 (2015).

20. A. Mentes, *et al.*, High-resolution cryo-EM structures of actin-bound myosin states reveal the mechanism of myosin force sensing. *Proc. Natl. Acad. Sci.* **110**, 201718316 (2018).

21. J. S. Allingham, R. Smith, I. Rayment, The structural basis of blebbistatin inhibition and specificity for myosin II. *Nat. Struct. Mol. Biol.* **12**, 378–379 (2005).

22. D. A. Winkelmann, E. Forgacs, M. T. Miller, A. M. Stock, Structural basis for drug-induced allosteric changes to human β-cardiac myosin motor activity. *Nat. Commun.* **6**, 7974 (2015).

23. S. Hashem, M. Tiberti, A. Fornili, Allosteric modulation of cardiac myosin dynamics by omecamtiv mecarbil. *PLoS Comput. Biol.* **13**, e1005826 (2017).

24. J. D. Powers, *et al.*, Cardiac myosin activation with 2-deoxy-ATP via increased electrostatic interactions with actin. *Proc. Natl. Acad. Sci.* **116**, 11502–11507 (2019).

25. F. Blanc, *et al.*, An intermediate along the recovery stroke of myosin VI revealed by X-ray crystallography and molecular dynamics. *Proc. Natl. Acad. Sci.* **115**, 6213–6218 (2018).

26. K. Chinthalapudi, S. M. Heissler, M. Preller, J. R. Sellers, D. J. Manstein, Mechanistic insights into the active site and allosteric communication pathways in human nonmuscle myosin-2C. *eLife* **6**, 1017 (2017).

27. K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).

28. C. R. Knoverek, G. K. Amarasinghe, G. R. Bowman, Advanced Methods for Accessing Protein Shape-Shifting Present New Therapeutic Opportunities. *Trends Biochem. Sci.* **44**, 351–364 (2018).

29. G. R. Bowman, P. L. Geissler, Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci.* **109**, 11681–11686 (2012).

30. K. M. Hart, C. M. W. Ho, S. Dutta, M. L. Gross, G. R. Bowman, Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat. Commun.* **7**, 12965 (2016).

31. J. R. Porter, *et al.*, Cooperative Changes in Solvent Exposure Identify Cryptic Pockets, Switches, and Allosteric Coupling. *Biophys. J.* (2019).

32. M. Saraste, P. R. Sibbald, A. Wittinghofer, The P-loop–a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434 (1990).

33. G. R. Bowman, V. S. Pande, F. Noé, *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* (Springer Science & Business Media, 2013).

34. J. D. Chodera, F. Noé, Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144 (2014).

35. R. C. Edgar, Local homology recognition and distance measures in linear time using compressed amino acid alphabets. *Nucleic Acids Res.* **32**, 380–385 (2004).

36. A. M. Gulick, C. B. Bauer, J. B. Thoden, I. Rayment, X-ray structures of the MgADP, MgATPgammaS, and MgAMPPNP complexes of the Dictyostelium discoideum myosin motor domain. *Biochemistry* **36**, 11619–11628 (1997).

37. H. M. Berman, *et al.*, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

38. G. R. Bowman, E. R. Bolin, K. M. Hart, B. C. Maguire, S. Marqusee, Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proc. Natl. Acad. Sci.* **112**, 2734–2739 (2015).

39. M. I. Zimmerman, *et al.*, Prediction of New Stabilizing Mutations Based on Mechanistic Insights from Markov State Models. *ACS Cent. Sci.*, acscentsci.7b00465 (2017).

40. S. Singh, G. R. Bowman, Quantifying Allosteric Communication via Both Concerted Structural Changes and Conformational Disorder with CARDS. *J. Chem. Theory Comput.*, acs.jctc.6b01181 (2017).

41. Z. Su, *et al.*, Electron Cryo-microscopy Structure of Ebola Virus Nucleoprotein Reveals a Mechanism for Nucleocapsid-like Assembly. *Cell* **172**, 966–978.e12 (2018).

42. X. Sun, S. Singh, K. J. Blumer, G. R. Bowman, Simulation of spontaneous G protein activation reveals a new intermediate driving GDP unbinding. *eLife* **7**, 19 (2018).

43. J. Shlens, A Tutorial on Principal Component Analysis (2014).

44. B. Patterson, K. M. Ruppel, Y. Wu, J. A. Spudich, Cold-sensitive mutants G680V and G691C of Dictyostelium myosin II confer dramatically different biochemical defects. *J. Biol. Chem.* **272**, 27612–27617 (1997).

45. F. Kinose, S. X. Wang, U. S. Kidambi, C. L. Moncman, D. A. Winkelmann, Glycine 699 is pivotal for the motor activity of skeletal muscle myosin. *J. Cell Biol.* **134**, 895–909 (1996).

46. M. J. Bloemink, M. A. Geeves, Shaking the myosin family tree: Biochemical kinetics defines four types of myosin motor. *Semin. Cell Dev. Biol.* **22**, 961–967 (2011).

47. M. Nyitrai, M. A. Geeves, Adenosine diphosphate and strain sensitivity in myosin motors. *Phil Trans R Soc B* **359**, 1867–1877 (2004).

48. E. M. De La Cruz, E. M. Ostap, H. L. Sweeney, Kinetic mechanism and regulation of myosin VI. *J. Biol. Chem.* **276**, 32373–32381 (2001).

49. J. H. Lewis, M. J. Greenberg, J. M. Laakso, H. Shuman, E. M. Ostap, Calcium Regulation of Myosin-I Tension Sensing. *Biophys. J.* **102**, 2799–2807 (2012).

50. S. Watanabe, R. Ikebe, M. Ikebe, Drosophila myosin VIIA is a high duty ratio motor with a unique kinetic mechanism. *J. Biol. Chem.* **281**, 7151–7160 (2006).

51. K. Homma, M. Ikebe, Myosin X is a high duty ratio motor. *J. Biol. Chem.* **280**, 29381–29391 (2005).

52. E. M. De La Cruz, E. M. Ostap, Kinetic and equilibrium analysis of the myosin ATPase. *Methods Enzymol.* **455**, 157–192 (2009).

53. I. Chizhov, F. K. Hartmann, N. Hundt, G. Tsiavaliaris, Global Fit Analysis of Myosin-5b Motility Reveals Thermodynamics of Mg2+-Sensitive Acto-Myosin-ADP States. *PloS One* **8**, e64797 (2013).

54. S. M. Heissler, J. R. Sellers, Myosin light chains: Teaching old dogs new tricks. *Bioarchitecture* **4**, 169–188 (2014).

55. M. A. Geeves, C. Perreault-Micale, L. M. Coluccio, Kinetic analyses of a truncated mammalian myosin I suggest a novel isomerization event preceding nucleotide binding. *J. Biol. Chem.* **275**, 21624–21630 (2000).

56. A. Waterhouse, *et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

57. P. S. Gurel, *et al.*, Cryo-EM structures reveal specialization at the myosin VI-actin interface and a mechanism of force sensitivity. *eLife* **6**, 213 (2017).

58. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

59. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

60. M. J. Abraham, *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

61. H. J. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).

62. Y. Duan, *et al.*, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).

63. M. Shirts, V. S. Pande, COMPUTING: Screen Savers of the World Unite! *Science* **290**, 1903–1904 (2000).

64. J. R. Porter, M. I. Zimmerman, G. R. Bowman, Enspara: Modeling molecular ensembles with scalable data structures and parallel computing. *J. Chem. Phys.* **150**, 044108 (2019).

65. M. I. Zimmerman, J. R. Porter, X. Sun, R. R. Silva, G. R. Bowman, Choice of adaptive sampling strategy impacts state discovery, transition probabilities, and the apparent mechanism of conformational changes. *J. Chem. Theory Comput.* **14**, acs.jctc.8b00500–5475 (2018).

66. R. T. McGibbon, *et al.*, MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–1532 (2015).

67. L. P. Coelho, Jug: Software for Parallel Reproducible Computation in Python. *J. Open Res. Softw.* **5**, 022109 (2017).

68. O. Tange, GNU Parallel - The Command-Line Power Tool. *Login USENIX Mag.*, 42–47 (2011).

69. F. Pedregosa, *et al.*, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

70. T. E. Oliphant, Python for scientific computing. *Comput. Sci. Eng.* **9** (2007).

71. S. Omohundro, *Five balltree construction algorithms* (1989).

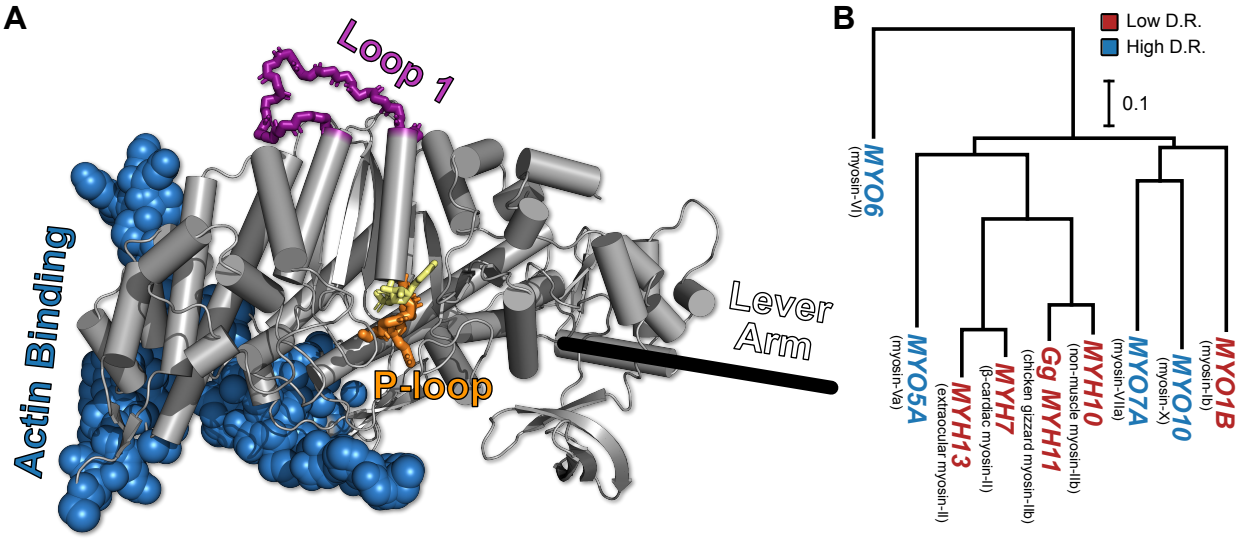
72. M. J. Greenberg, E. M. Ostap, Regulation and control of myosin-I by the motor and light chain-binding domains. *Trends Cell Biol.* **23**, 81–89 (2013).

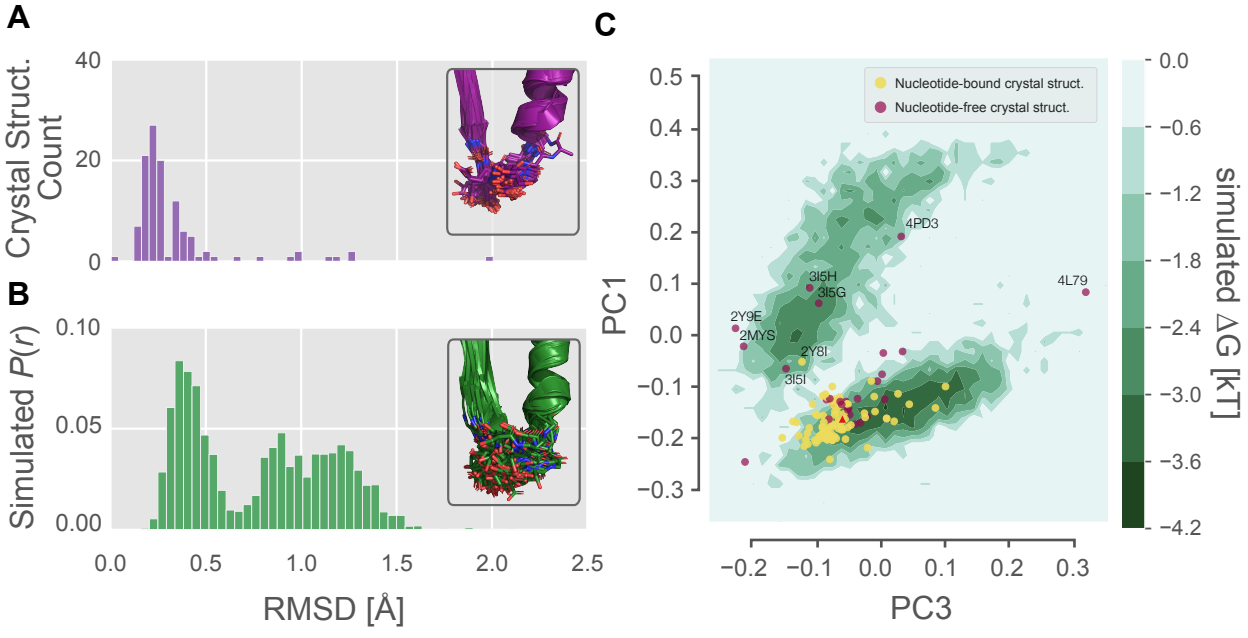
73. Y. E. Goldman, E. M. Ostap, "4.9 Myosin Motors: Kinetics of Myosin" in *Comprehensive Biophysics*, (Elsevier, 2012), pp. 151–169.

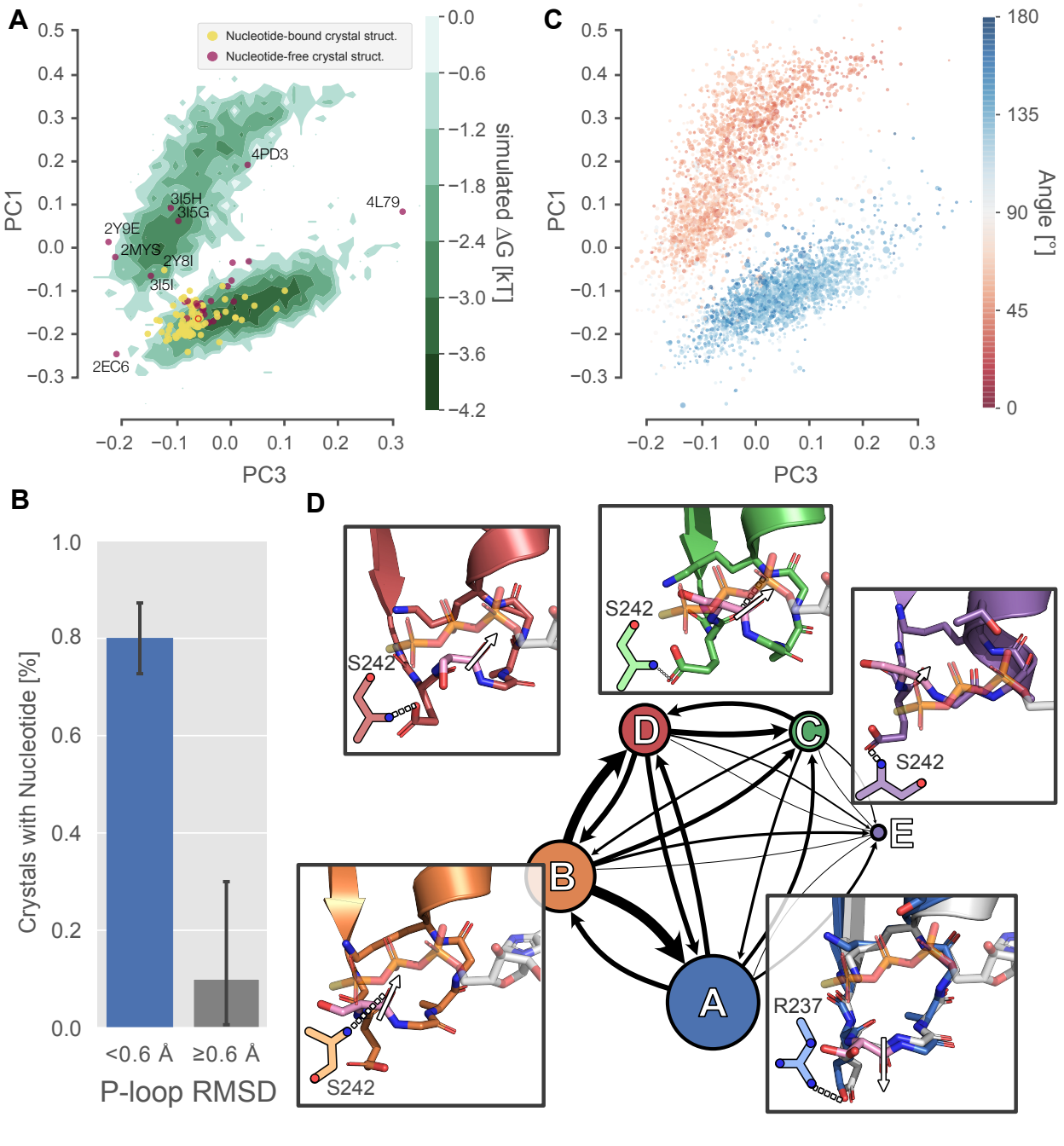
74. J. Robert-Paganin, *et al.*, Plasmodium myosin A drives parasite invasion by an atypical force generating mechanism. *Nat. Commun.* **10**, 3286 (2019).

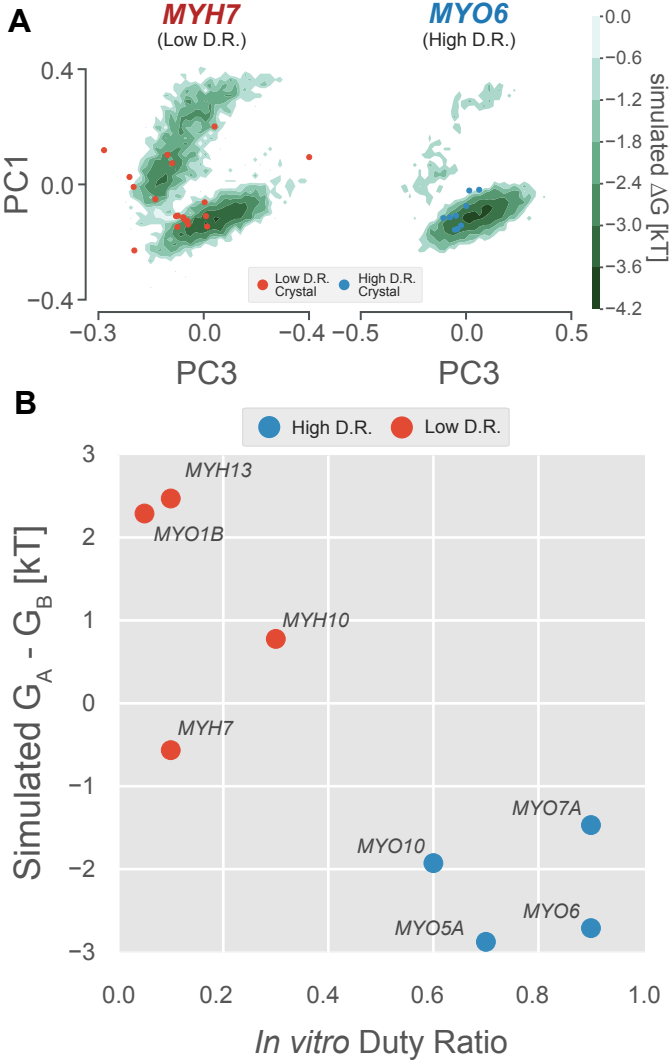
75. J. D. Hunter, Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

76. D. A. Green, A colour scheme for the display of astronomical intensity images. *arXiv.org* (2011).


A
Loop 1
Actin Binding
P-loop
Lever Arm
B
Low D.R.
High D.R.
0.1
MYO6
(myosin-VI)
MYO5A
(myosin-Va)
MYH13
(extraocular myosin-II)
MYH7
(β-cardiac myosin-II)
Gg MYH11
(chicken gizzard myosin-IIb)
MYH10
(non-muscle myosin-IIb)
MYO7A
(myosin-VIIa)
MYO10
(myosin-X)
MYO1B
(myosin-Ib)


A
Crystal Struct. Count
40
20
0
B
Simulated P(r)
0.10
0.05
0.00
0.0
0.5
1.0
1.5
2.0
2.5
RMSD [Å]
C
Nucleotide-bound crystal struct.
Nucleotide-free crystal struct.
4PD3
4L79
3I5H
3I5G
2Y9E
2MYS
2Y8I
3I5I
PC1
PC3
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
simulated ΔG [kT]
0.0
−0.6
−1.2
−1.8
−2.4
−3.0
−3.6
−4.2


A
Nucleotide-bound crystal struct.
Nucleotide-free crystal struct.
4PD3
4L79
3I5H
3I5G
2Y9E
2MYS
2Y8I
3I5I
2EC6
PC1
PC3
simulated ΔG [kT]
C
Angle [°]
B
Crystals with Nucleotide [%]
<0.6 Å
≥0.6 Å
P-loop RMSD
D
S242
R237
A
B
C
D
E

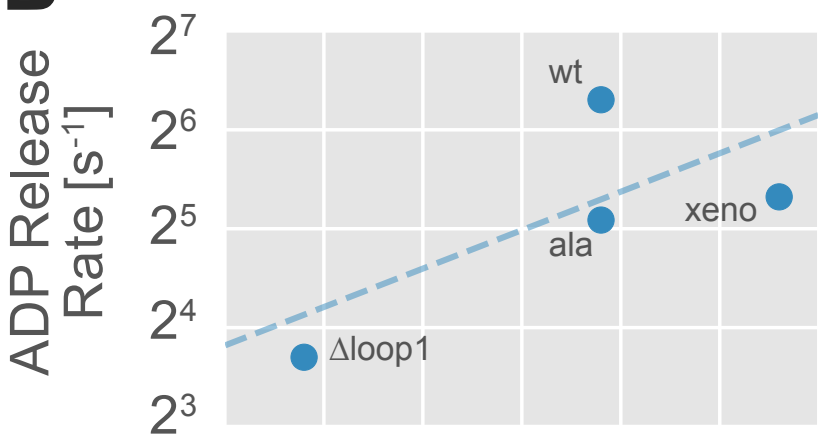

A
MYH7
(Low D.R.)
MYO6
(High D.R.)
PC1
PC3
PC3
Low D.R. Crystal
High D.R. Crystal
simulated ΔG [kT]
B
High D.R.
Low D.R.
MYH13
MYO1B
MYH10
MYH7
MYO7A
MYO10
MYO6
MYO5A
Simulated G_A - G_B [kT]
In vitro Duty Ratio

A

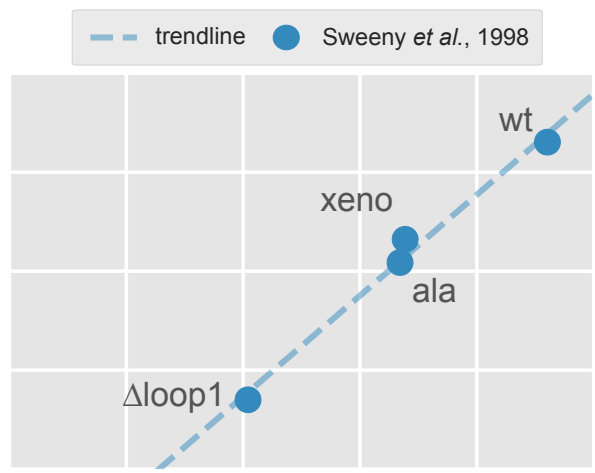
wt *Gg* MYH11	SSHKGKKDTSI---------TQGPSFSYGELE
ala *Gg* MYH11	SSHKGKKDTSI---------T**AGPAAAA**GELE
xeno *Gg* MYH11	SSHKGKKDHTIPTESPKAIKHQSGSLLYGELE
Δloop1 *Gg* MYH11	SSHKG----------GELE
Hs MYH13	VTGDKKKETQ----------PGKMQGTLE
Hs MYH7	AIGDRSKK----------DQSPGK--GTLE
Hs MYH10	SSHKGRKDHNIPE-------SPKPVKHQGELE
Hs MYO1B	VCGKGAEV----------NQVK

Loop 1

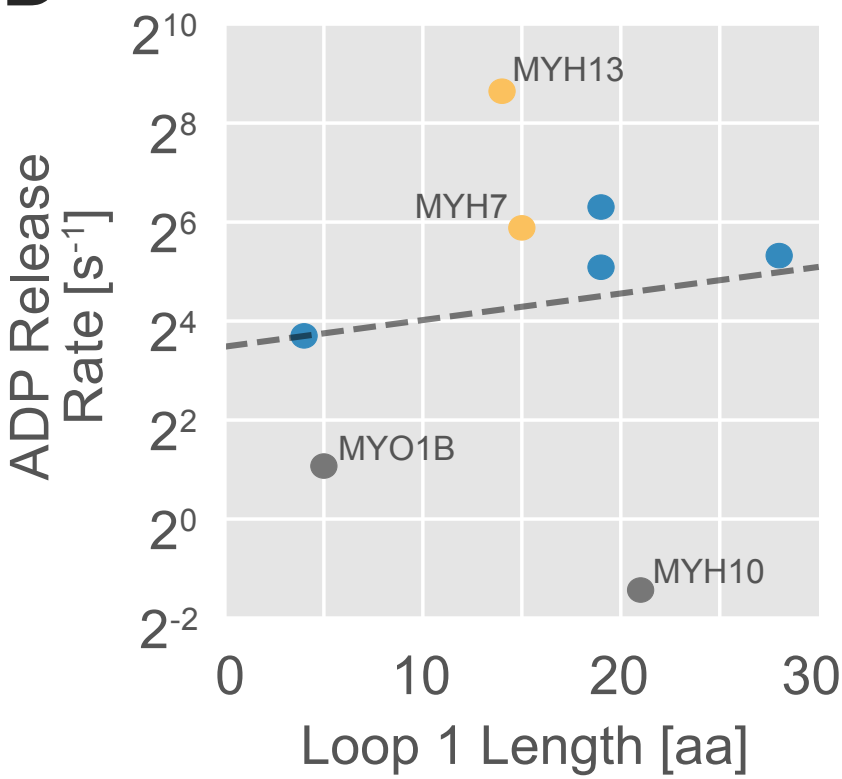
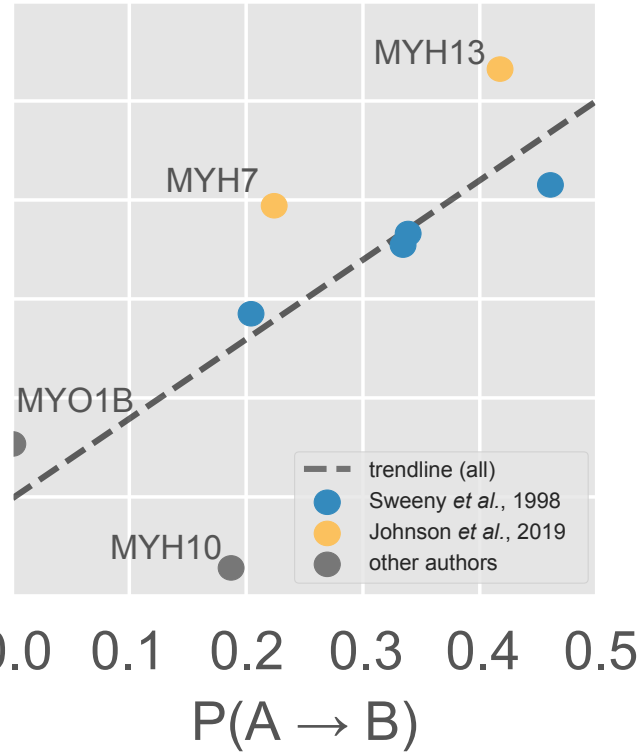
B



C



D



E