PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Mode coupling points to functionally important residues in myosin II

Onur Varol,[1,2] Deniz Yuret,[1] Burak Erman,[1] and Alkan Kabakçıoğlu[1]*

[1] Colleges of Engineering and Sciences, Koç University, Sarıyer 34450, İstanbul, Turkey
[2] School of Informatics and Computing, Indiana University, Bloomington, Indiana

ABSTRACT

Relevance of mode coupling to energy/information transfer during protein function, particularly in the context of allosteric interactions is widely accepted. However, existing evidence in favor of this hypothesis comes essentially from model systems. We here report a novel formal analysis of the near-native dynamics of myosin II, which allows us to explore the impact of the interaction between possibly non-Gaussian vibrational modes on fluctutational dynamics. We show that an information-theoretic measure based on mode coupling *alone* yields a ranking of residues with a statistically significant bias favoring the functionally critical locations identified by experiments on myosin II.

Proteins 2014; 82:1777–1786.

Key words: mode coupling; molecular dynamics; anharmonicity; allostery; myosin II.

INTRODUCTION

Fluctuation based analysis of protein dynamics has long proven to be an invaluable tool for investigating the interplay between protein dynamics and function.[1,2] Despite the past success enjoyed by the bead-and-spring-type linear models (such as, elastic, Gaussian, and anisotropic network models), it is well known that, both experimental evidence and simulations reveal strong departure from purely Gaussian (harmonic) behavior at physiological temperatures.[3] Deviations from harmonicity are most pronounced in slow, collective modes which are significant, for example, in the context of vibrational absorption spectrometry,[4] dimensional reduction,[5,6] and the role of hydration effects.[7]

On the other hand, the decomposition of MD fluctuation data into independent, possibly anharmonic modes is only the first step in an infinite cascade of corrections that bridge the gap between the dynamics of actual proteins and Gaussian models. The contribution of high-order corrections signifies the degree to which the experimental/computational free energy landscape fails to conform to a representation composed of independent modes (harmonic or anharmonic). In other words, they are "mode-coupling" corrections which yield valuable information on means of energy transfer and associated correlated activity within the protein.[8–13] Characterization of the conformational population sampled by near-native dynamics is believed to be the key to understanding the functioning of allosteric proteins,[14,15] if not all.[16] The interactions between vibrational modes play an essential role in shaping this population.

Recently, we introduced a systematic mathematical analysis of the fluctuational data (e.g., obtained from full-atomistic simulations) that naturally distinguishes the anharmonic and mode-coupling contributions to the free energy.[17] Here, we combine this analytical formulation with computer simulations of the near-native dynamics of myosin II and demonstrate that the mode coupling *alone* highlights functionally critical sites of this allosteric protein. The relevance of coupling between vibrational modes in the context of allosteric transitions in myosin II was also pointed out in an earlier work.[18]

The paper is organized as follows: The section "Modal expansion and beyond" describes the theoretical framework used for isolating the contribution of mode coupling from other anharmonic effects in the MD fluctuation data; the section "Mode-coupling based ranking of residues" discusses how the formulation above can be used to

*Correspondence to: Alkan Kabakç ioğlu; Colleges of Engineering and Sciences, Sci-118, Koç University, Sar iyer, 34450 İstanbul, Turkey. E-mail: akabakcioglu@ku.edu.tr
Received 21 October 2013; Revised 7 January 2014; Accepted 28 January 2014
Published online 7 February 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24531

select out residues that are highlighted by mode coupling; the section "Myosin II" introduces the motor protein myosin II which we use here as a test case; the section "MD simulations and the eigenmodes" gives the details of the molecular dynamics (MD) simulations performed on myosin II; the section "Results" reports our findings.

MATERIALS AND METHODS

Modal expansion and beyond

Our raw data is the time series for the space coordinates of the α-carbons obtained from a full-atomistic molecular dynamics (MD) simulation whose details are given in the section "MD simulations and the eigenmodes". Using the MD trajectory, we derive a multivariate probability distribution function $p(\Delta\mathbf{R})$, where $\Delta R_i, \Delta R_{i+N}$, and ΔR_{i+2N} with $i=1,\ldots,N$ are the deviations from the mean position along the coordinate axes x, y, and z, respectively, of the ith C^α atom in a protein with N amino acids. The covariance matrix $\Gamma=\langle\Delta\mathbf{R}\Delta\mathbf{R}^T\rangle$ is then used to transform the coordinate system by means of a scaling and a rotation into the modal space: $\Delta\mathbf{r}=\Gamma^{-1/2}\Delta\mathbf{R}$. For a purely harmonic system, the resulting distribution function is given by

$$f(\Delta\mathbf{r})=\prod_i \frac{\exp\left[-\Delta r_i^2/2\right]}{\sqrt{2\pi}} \tag{1}$$

while deviations from Eq. (1) due to anharmonicity and mode coupling are observed in proteins, as mentioned before.

Hermite expansion

Building on an earlier proposal,[19] we recently developed an analytical formalism that naturally extends Eq.(1) into the regime where harmonicity breaks down.[13,17] In this framework, $f(\Delta\mathbf{r})$ is expressed as an infinite sum:

$$f(\Delta\mathbf{r})=\frac{1}{\sqrt{(2\pi)^{3N}}}e^{-\sum_i \Delta r_i^2/2}\left[1+\sum_i\sum_{\nu=3}^{\infty} c_\nu^i H_\nu(\Delta r_i) + \sum_{i\neq j}\sum_{\nu=3}^{\infty}\sum_{p=1}^{\nu-1} c_{p,\nu-p}^{ij} H_p(\Delta r_i)H_{\nu-p}(\Delta r_j) + \sum_{i\neq j\neq k}\cdots\right] \tag{2}$$

where H_i is the Hermite polynomial of rank i. The choice of the Hermite basis ensures that the expansion coefficients are given by $c_\nu^i=\langle H_\nu(\Delta r_i)\rangle/\nu!$ and $c_{p,\nu-p}^{ij}=\binom{\nu}{p}\langle H_p(\Delta r_i)H_{\nu-p}(\Delta r_j)\rangle/\nu!$, where $\langle\cdot\rangle$ denotes the time average evaluated over the MD data. Kabakçioğlu *et al.* [17] describes how the symmetry properties of Hermite tensor polynomials can be exploited to reduce the computational complexity associated with estimating these coefficients from the MD trajectory.

The leading term in Eq. (2), which is identical to Eq. (1), corresponds to a purely harmonic dynamics and is referred as f_0 here. This zeroth-order form is the basis for many protein fluctuation models.[20–22] Remaining terms within the square brackets in Eq. (2) reflect all possible corrections due to non-Hookian modes, as well as pairwise, threesome, and high-order mode–mode interactions. We wish to focus on the impact of mode coupling in our study, therefore, our first goal is to distinguish the contributions that yield anharmonic (and still independent) modes from those that are due to the interactions among such modes. Here, we will refer to the former as *marginal anharmonicity*, since this contribution is uniquely determined by the deviations of the marginal distributions $f(\Delta r_i)=\int\prod_{j\neq i} dr_j f(\Delta\mathbf{r})$ from Gaussian.

To this end, let f_1 refer to the best possible description of the data under the assumption of marginal anharmonicity:

$$f_1(\Delta\mathbf{r})=\frac{1}{\sqrt{(2\pi)^{3N}}}e^{-\sum_k \Delta r_k^2/2}\prod_i\left[1+\sum_{\nu=3}^{\nu_{\max}} c_\nu^i H_\nu(\Delta r_i)\right] \tag{3}$$

where $\nu_{\max}$ is a cutoff degree imposed by practical considerations (see Myosin II). This approximation to the conformational distribution function obtained from near-native dynamics yields exact single-mode (marginal) histograms in the limit $\nu_{\max}\to\infty$. Note that f_1 is fully specified by the coefficients $\{c_\nu^i\}$. Nevertheless, marginal anharmonicity is reflected at all orders in $f(\Delta\mathbf{r})$ (i.e., $c_{p,\nu-p}^{ij}$ and high-order coefficients are typically nonzero). At first sight, these high-order contributions may be confused with mode coupling since they are in the form of a product involving multiple vibrational modes. However, it is transparent from Eq. (3) that the information on mode-mode interactions is contained in *everything but* f_1.

Mode-coupling based ranking of residues

It is tempting at this point to attempt to identify pairs of modes which interact strongly and/or have the most impact on protein function. Numerous studies in this spirit can be found in the literature (see, e.g., Refs. 18,23). However, interpreting such data usually requires an understanding of the functional dynamics and does not immediately relate to experiments. Furthermore, a pairwise interaction picture is incomplete in the current context, because the corrections to the fluctuational free energy are not additive in mode pairs. In other words, high-order contributions exist.

Instead, we here focus directly on the critical residues of the protein which are highlighted by mode coupling at all orders. This kind of information is not only easier to compare with available experimental data (such as site-directed mutation scans), but, as it turns out, it is also computationally cheaper to access. As is evident from Eq. (2), estimating mode-coupling corrections per mode pair involves calculating second- and high-order coefficients $c_{\nu\eta\ldots}^{ij\ldots}$ associated with the individual mode pair, repeated for $\binom{N}{2}$ pairs; a CPU demanding task. The cumulative effect of mode coupling, however, is already available in the difference between f and f_1. This information can be projected onto the protein's sequence axis by the procedure outlined in the section "Identifying per residue impact of mode coupling". The outcome is a score profile for each amino acid in the protein, reflecting the degree to which their near-native fluctuations are modulated by mode coupling.

Identifying per residue impact of mode coupling

To identify the residues highlighted by marginal anharmonicity and mode coupling, separately, we back-project the distributions f_0 and f_1 onto the space of C^α atomic coordinates:

$$p_{0,1}(\Delta\mathbf{R}) = f_{0,1}(\Delta\mathbf{r}(\Delta\mathbf{R}))/\sqrt{\det\Gamma}.$$

p_0 and p_1 are approximations, at two different levels (Gaussian and marginally anharmonic), to the original distribution $p(\Delta\mathbf{R})$ obtained from the MD trajectory.

Next, we consider the marginal distributions $p(\Delta\mathbf{R_i}) = \int \prod_{j\neq i} \mathbf{dR_j}\, p(\Delta\mathbf{R})$ for individual coordinates C_i^α and measure the Kullback–Leibler (KL) divergence,[24] d_{KL}, between p and p_1, as well as between p_1 and p_0 for a given residue i. The former distance yields quantitative information on the extent to which mode coupling governs fluctutations of the given coordinate C_i^α, while the latter yields a similar measure as regards to marginal anharmonicity. For distributions p and q of a continuous random variable x, KL-divergence is defined to be the integral

$$d_{\mathrm{KL}}(p||q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)}\, \mathrm{d}x. \tag{4}$$

The integration steps involved in the KL divergence estimation require the discrete probability distributions obtained from the MD data to be smoothened out into continuous functions. To this end, we use kernel density estimation (KDE)[25,26] which yields a continuous probability distribution $\hat{p}(x)$ from a set of samples $\{x_i\}$ as

$$\hat{p}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \tag{5}$$

where K is the kernel function (chosen to be Gaussian) and h is the bandwidth parameter (determined by the method in Ref. 27).

The total impact for a given residue is taken to be the sum of the KL-divergence values for its three spatial coordinates:

$$S_i^{\mathrm{mc}} = \sum_{\alpha=x,y,z} d_{\mathrm{KL}}\left[p(\Delta\mathbf{R}_{\mathbf{i},\alpha})||p_1(\Delta\mathbf{R}_{\mathbf{i},\alpha})\right], \tag{6}$$

$$S_i^{\mathrm{ma}} = \sum_{\alpha=x,y,z} d_{\mathrm{KL}}\left[p_1(\Delta\mathbf{R}_{\mathbf{i},\alpha})||p_0(\Delta\mathbf{R}_{\mathbf{i},\alpha})\right]. \tag{7}$$

Above, "mc" and "ma" stand for "mode coupling" and "marginal anharmonicity", respectively. We note that, S_i^{ma} is a measure of the information added by anharmonicity (under the condition of separability) on top of the best fitting harmonic (f_0) reference distribution, while S_i^{mc} measures the information added by mode coupling on top of *its* reference distribution which is the best separable function that matches the data. With this perspective, the chosen measure is naturally an asymmetric one under the exchange of its argument functions. Finally, for comparison, we also consider the mean residue displacements (akin to experimental B-factors) measured by the variation:

$$\sigma_i^2 \equiv \int \Delta\mathbf{R_i}^2 p(\Delta\mathbf{R_i})\, d\Delta\mathbf{R_i} \tag{8}$$

as the fluctuation based score for a residue. Below, we apply this analysis to the MD data from myosin II, a molecular motor protein, and compare the performance of the above ranking schemes in distinguishing functionally significant locations on the protein. Note that, once the MD data are available, estimation of the residue scores above is a mechanical process, without any tuning paramaters.

Myosin II

Dictyostelium discoideum myosin II is an allosteric protein which has been extensively studied both experimentally and computationally. It is an actin-binding molecular motor protein crucial for various biological processes, such as, cell movement, muscle contraction in higher organisms, membrane transport and several signaling pathways. Among the 35 known classes of myosin, 13 appear in Human.[28] The motor domain of myosin II shown in Figure 1 goes through conformational changes at each stage of its four-stroke catalytic cycle which converts the chemical energy derived from ATP hydrolysis

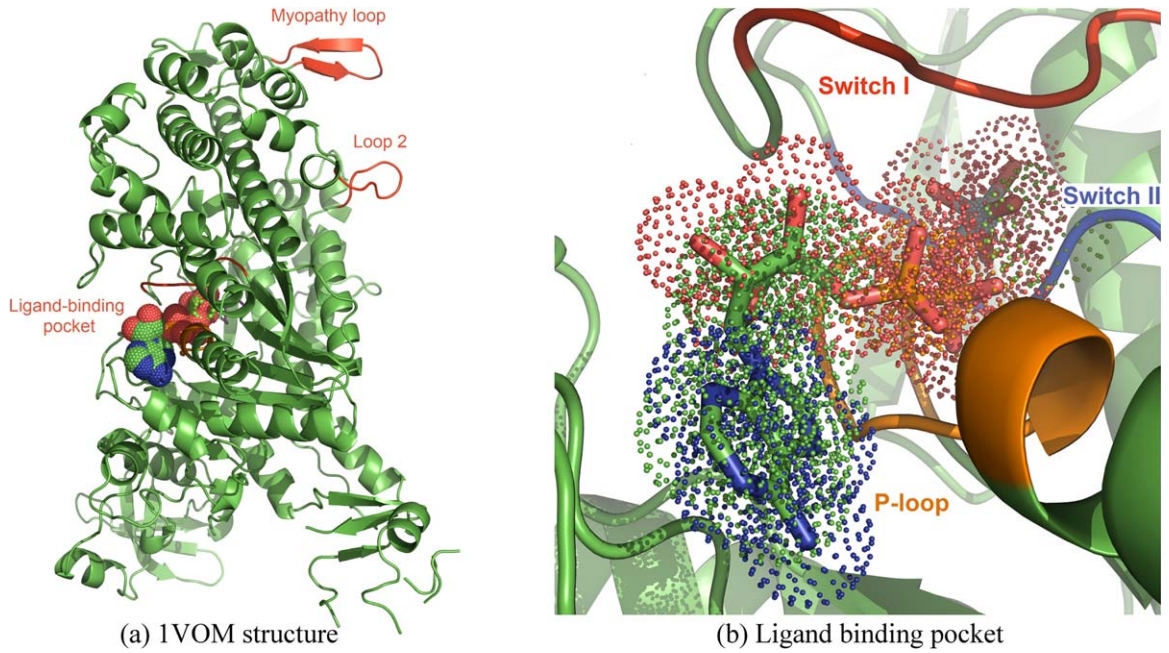

Figure 1

Functional sites of myosin II (PDB:1VOM). The full structure of the motor domain is shown on the left (**a**). Actin- and ADP-binding loops are indicated in red. On the right, the ligand-binding pocket is shown in greater detail (**b**). The ADP molecule in the middle is surrounded by functional elements Switch I (red), Switch II (blue), and P-loop (orange). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

into mechanical work. The results presented below are obtained from MD simulations of the structure PDB:1VOM,[29] where an ADP is bound on the protein.

There exist several, well-studied functionally relevant locations on the structure. Switch II and P-loop shown in Figure 1(a) are known to control the MgADP release mechanism.[29] Structural changes during the characteristic cleft closure motion in the motor domain is believed to be related to activity in Switch I region, which opens the binding pocket and modifies the relative placement of the P-loop and Switch II regions.[30] The actin-binding pocket of the structure is composed of the myopathy loop and loop-2, shown in Figure 1(b). The interaction between loop-2 and the negatively charged parts of the actin is also documented.[31]

Below, we perform the proposed fluctuational analysis on the motor domain of myosin II and identify the residues whose fluctuations are most significantly modified by mode coupling. We find that there is a statistically significant correlation between these and the functional regions mentioned above, as well as a subset of critical residues of the protein determined by experimental methods (such as point mutations).

MD simulations and the eigenmodes

The structure is composed of 730 residues and the ligand whose atomic coordinates (the initial configuration of the MD simulation) were extracted from the PDB database. The MD simulations were carried out using NAMD 2.7 software package[32] with CHARMM27 force field[33] in explicit solvent (water) at 310 K. Langevin dynamics was used to control the temperature and the pressure in an NPT ensemble. A water box with a 15 Å cushion and periodic boundary conditions were applied. The integration time step in the simulation was selected as 1 fs for both nonbonded and electrostatic forces and no rigid bonds were used. The trajectory was captured every 50 fs within several windows of ~2 ns duration, for a total run of 10 ns.

Note that, a much longer simulation time would be required to observe the functional dynamics of the protein. The purpose of our simulation, however, is merely to monitor the fluctuations and to gather sufficient data on the non-Gaussian nature of the conformational distribution. The procedure may be crudely likened to recording a short bike ride and then analyzing the small displacements of various components in order to identify the elements that are critical for mechanical energy transfer (except, of course, thermal fluctuations are much more significant in the current system).

The MD output was analyzed in several time windows. Let us consider the steps of the analysis on the first 2 ns of the simulation data after equilibration, where snapshots taken 0.5 ps apart amount to $T = 4000$ data points: We first construct the vector $\Delta\mathbf{R}^{(i)}$ of the C^{α} positions in

each snapshot with $i=1,\ldots,T$. We next calculate the covariance matrix Γ given in Modal expansion and beyond, and identify the modal coordinates $\Delta\mathbf{R}^{(i)}$ for the 2184 fluctuation modes (out of $730\times3=2190$ degrees of freedom, excluding six associated with the center of mass translation/rotation). At this point, the zeroth-order approximant f_0 to $f(\Delta\mathbf{r}^{(i)})$ (and to $f(\Delta\mathbf{R}^{(i)})$, through the inverse transform) is already available. Next, we find the best marginally anharmonic description of the data, f_1, given by Eq. (3). This is done by estimating c_ν^i in Eq. (3) as averages over the MD snapshots, up to a sufficiently high cutoff degree $\nu_{max}=32$ which is obtained empirically (see Fig. 2). Finally, we calculate the residue scores using Eqs. (6–8).

Each time window considered was subjected to the same analysis. We here present results for the first 1, 2, 5, and the full 10 ns of the simulation. Figure 3 shows that the eigenvectors corresponding to the 1 and 10 ns time frames, ordered with respect to the amplitude of the corresponding eigenvalue, are in visible agreement for the slow modes (lower left corner of the figure). This observation is in line with earlier work which argues that the slow modes retain their identity across different time scales, even if the eigenvalue spectrum may change.[34,35] Early identifyability of these most relevant modes points to the internal consistency of our approach and supports our observation that, a mere 10 ns simulation is sufficient to extract meaningful information about the biologically critical correlations in the dynamics which are imprinted into the protein's complex structure.

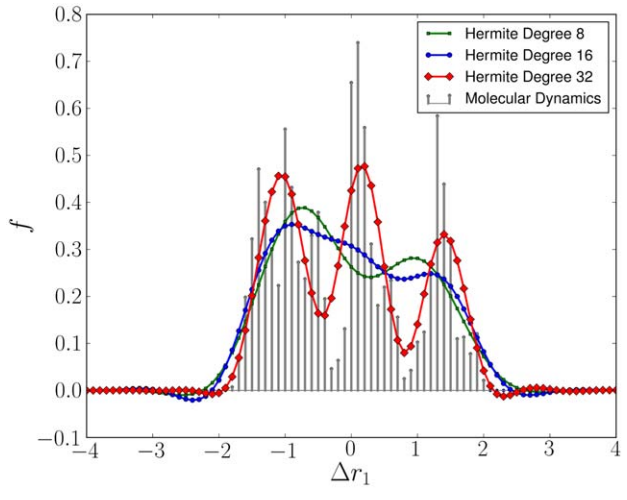

Figure 2

A comparison of the slowest mode's amplitude distribution for different choices for the maximum Hermite degree considered in Eq. (3). The cutoff degree $\nu_{max}=32$ was determined according to the criterion that the marginal distributions for all modes are captured with an accuracy same as above or better. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

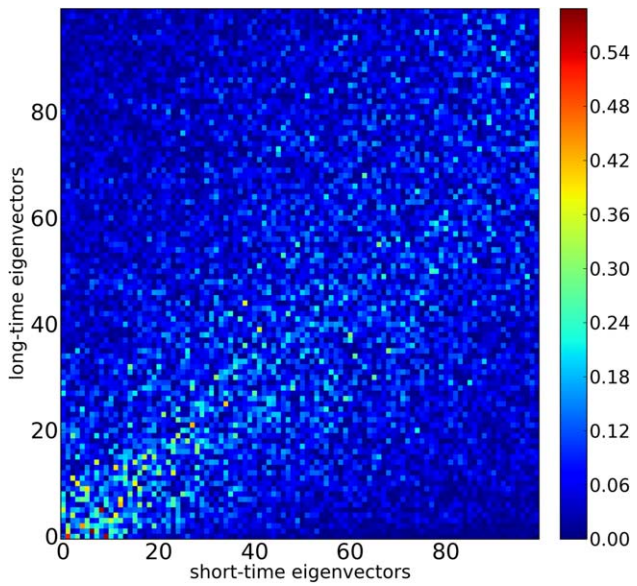

Figure 3

The overlap between the first 100 eigenvectors corresponding to the slowest fluctuational modes, ordered according to their eigenvalues, obtained from the first 1 (horizontal axis) and 10 ns (vertical axis) time frames. The accumulation along the lower diagonal indicates that the modal subspace spanned by slow eigenvectors retains its identity to a significant degree, with some amount of mixing between nearby modes. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

RESULTS

Figure 4 is a side-by-side comparison of residue rankings obtained from the first 1 ns, 2 ns, 5 ns, and 10 ns simulations of the ligand-bound motor domain. The initial configuration for both simulations was the structure PDB:1VOM and the analysis was performed on the data collected after equilibration. In each column, we consider three scoring schemes based on: (1) mode coupling, S_i^{mc}; (2) marginal anharmonicity, S_i^{ma}; (3) mean residue fluctuation, $\sigma_i^2=\langle\Delta R_i^2\rangle$. In Figure 4, the scores of the residues are shown for each time frame and evaluation criterion, alongside the functional regions indicated by different colored columns. We find that the high-scoring residues are marked by distinct peaks in mode-coupling based ranking (more so than that based on fluctuation amplitudes) while marginal anharmonicity is noisy and displays less selectivity among residues. Upon comparing the magnitudes on the vertical axes in the first two columns of Figure 4, the relative weight of mode coupling in an amino acid's fluctutational behavior is found to be larger by an order of magnitude than that of marginal anharmonicity. A similar observation was made on Crambin earlier.[17] Therefore, between the two non-Gaussian contributions, mode coupling appears as the dominant factor in shaping the configurational landscape. Consistently, only the residue-specific information

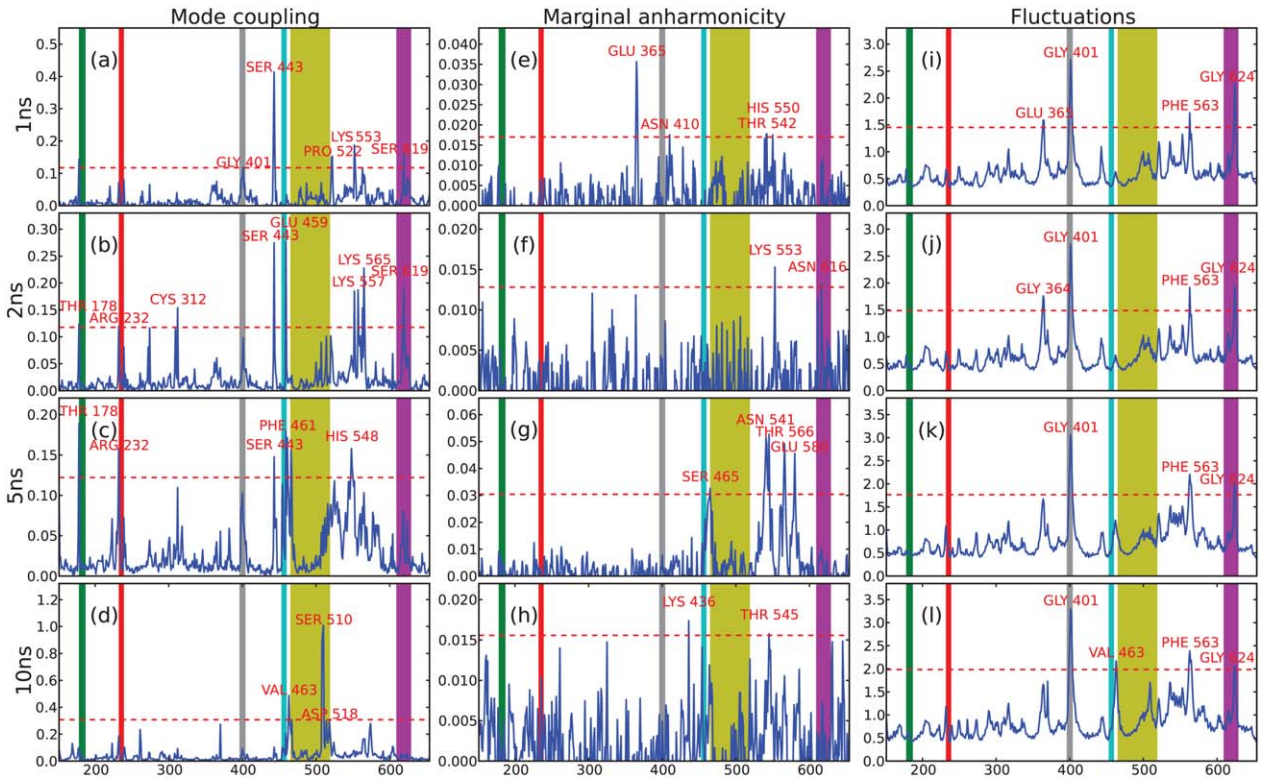

Figure 4

The residue scores obtained using Eqs. (6)(7) in the first 1, 2, 5, and 10 ns of the simulation, with respect to mode coupling (first column), marginal anharmonicity (second column), and residue fluctuations (third column). Colored bars in each figure indicate functionally relevant regions, P-loop (green), Switch I (red), myopathy loop (gray), Switch II (cyan), relay helix (yellow), and loop-2 (purple), reported in the literature and described in Myosin II. 3σ threshold is shown by the dashed line. The residue IDs of prominent peaks outside the 3σ margin are given in red.

gathered from mode-coupling corrections yields a rank profile in significant agreement with site-specific mutation data available for myosin II, as demonstrated in Comparison with point mutation data.

A closer inspection of mode-coupling based scores in different time frames (first column in Fig. 4) reveals an interesting progression. We observe an increasing level of activity in the ligand binding pocket during the first half of the simulation, but not later. On the other hand, the contribution to the second half of the simulation is mostly from the ends of the relay helix which is known to mechanically couple the ligand- and actin-binding pockets. Such shifts in activity in different time frames may be reminiscent of the complex communication patterns established through energy transfer between different vibrational modes in the system.[9,11]

The scoring based on residue fluctuation amplitudes alone (last row in Fig. 4) shows little difference between different time frames, as one might expect.

Comparing first and last columns of Fig. 4, we observe that there is some overlap between regions accentuated by mode coupling versus by mean fluctuation amplitudes. This is expected, not only because our analysis derives from fluctuation data, but also because some functional sites reside on flexible loop regions. We find that the myopathy loop and Loop-2, which are essential for actin binding, yield a strong signal in both fluctuation and mode-coupling based rankings. In contrast, loop regions in the ligand binding pocket (P-loop, switch I/II) do not fluctuate as much (presumably due to the presence of the ligand), yet, they are still highlighted by mode coupling.

Comparison with point mutation data

Score profiles in Figure 4(a–d) further single out few locations which are not in the immediate vicinity of the color coded functional regions. For example, Ser-443, such a site selected by mode coupling, coincides with the bent at the distal end of the long helix between residues 411–440, known to promote the myopathy loop to bind actin.[36] These may correspond to further residues that are critical for protein function, for example relaying information between the ligand- and actin-binding regions. For an unbiased evaluation of all such instances, a higher resolution target set is desirable. To this end, we performed a thorough literature survey for residue-specific experimental data on myosin II. Table I is a comprehensive list of

Table I
Amino Acids that are Experimentally Verified to be Critical for Myosin II Function

N483, F487, I499, F506, L508, I687, F692, F745	Ref. [38]	D403, V405	Ref. [39]
N464, C470,N472,Y473, N475, F481, E746	Ref. [40]	Y494, W501	Refs. [38,41]
D590, P591, L592, Q593	Ref. [42]	S181	Ref. [43]
E467, E586,G624, G740	Ref. [41]	S236	Ref. [44]
N233, S237, R238	Ref. [45]	E459	Refs. [40,46]
I499, F692, R738	Ref. [47]	F482	Refs. [43,48]
D454, G457, F458	Ref. [46]	G680	Refs. [41,43,48,49]
E531, P536, R562	Ref. [50]	G691	Refs. [41,49]

This list is used in the text as a target set for evaluating the relevance of various physical processes; namely, marginal anharmonicity, mode coupling, and fluctuation amplitudes, to protein's function.

amino acids that we could gather for myosin II, which have been experimentally verified (mostly through point mutations) to be critical for its function. It is possible that, some of these are important structural elements, say, required for folding, and not necessarily critical in the sense of residing on a functional site or maintaining allosteric communication. Nevertheless, this list comprises a solid target set, free from theoretical considerations or interpretations of structural data.

We next compare various ranking methods (including random) against this target set. Figure 5 shows for each case, the percentage of top residues that need to be considered (vertical axis) to capture a certain fraction of the target set (horizontal axis). For a random ordering of the residues, this curve is expected to lie on the diagonal, with fluctuations typically less than 10% for the current data. A perfect ordering which places the target residues on top of the list would yield another straight line following the upper edge of the forbidden gray region at the bottom.

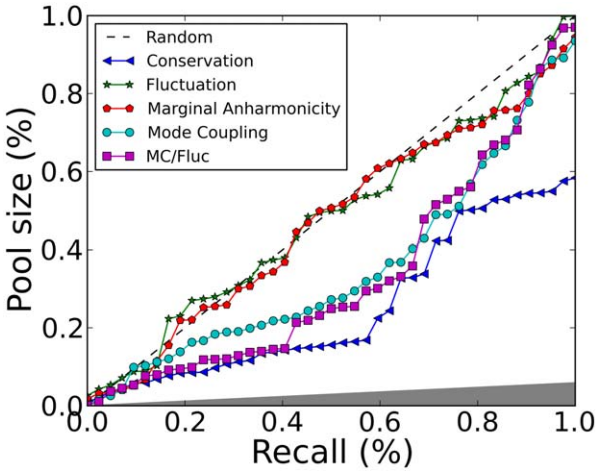

Figure 5
The performance of each ranking scheme considered in the paper is shown together with that of a random (dashed) and a perfect (upper edge of the forbidden black region) rank assignment. Each data point corresponds to the fraction of target residues captured out of a total of 43 (horizontal axis) by the top ranking residues covering a given percentage (vertical axis) of 730 residues in total. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Upon inspecting Figure 5, we observe that fluctuations and marginal anharmonicity display no preference for the target set. The conservation scores show the best correlation with the experimentally determined targets. This is hardly surprising, since experimental studies are in fact guided by conservation scores. The key result of our study is the green curve representing the mode-coupling based ordering. The clear deviation from the diagonal in the downward direction demonstrates that coupling between vibrational modes is an important physical mechanism for protein function and that this information can be cast into a predictive tool by means of the computational/analytical framework described above.

We also noticed that the scoring function $f_i \equiv S_i^{\mathrm{mc}}/\sigma_i^2$ which is also shown in Figure 5 (the curve labeled as "MC/Fluc") is consistently better than S_i^{mc} in terms of highlighting the target set listed in Table I. We checked that the general trend and the relative performances of considered scoring criteria are robust under different methods one might choose while harnessing the information from several time windows (such as, considering, for each amino acid, the maximum of a score among those calculated from successive time intervals, instead of a single score obtained from the full simulation). Note that, a gap builds in the tail of Figure 5 between conservation and mode-coupling rankings. This discrepancy may reflect that some residues in the target list are relevant for the folding process (therefore have high conservation scores), but not the functional dynamics where we expect mode coupling to play a role. Further investigations on myosin II and other proteins are in progress to confirm this hypothesis, as well as to verify the generality of the present approach.

Figure 5 presents a direct validation of the fact that, coupling between vibrational modes, as formulated here, is a significant physical mechanism underlying myosin IIs functional dynamics. This is the main message we wish to convey in this paper. Since the existing data on myosin II is far from being exhaustive, it is likely that further experiments will identify more essential residues for this protein. The method introduced here may also help in these future endeavors as a new guide for target selection (for example, Ser-443 appears to be a promising candidate). The distribution of top amino acids with

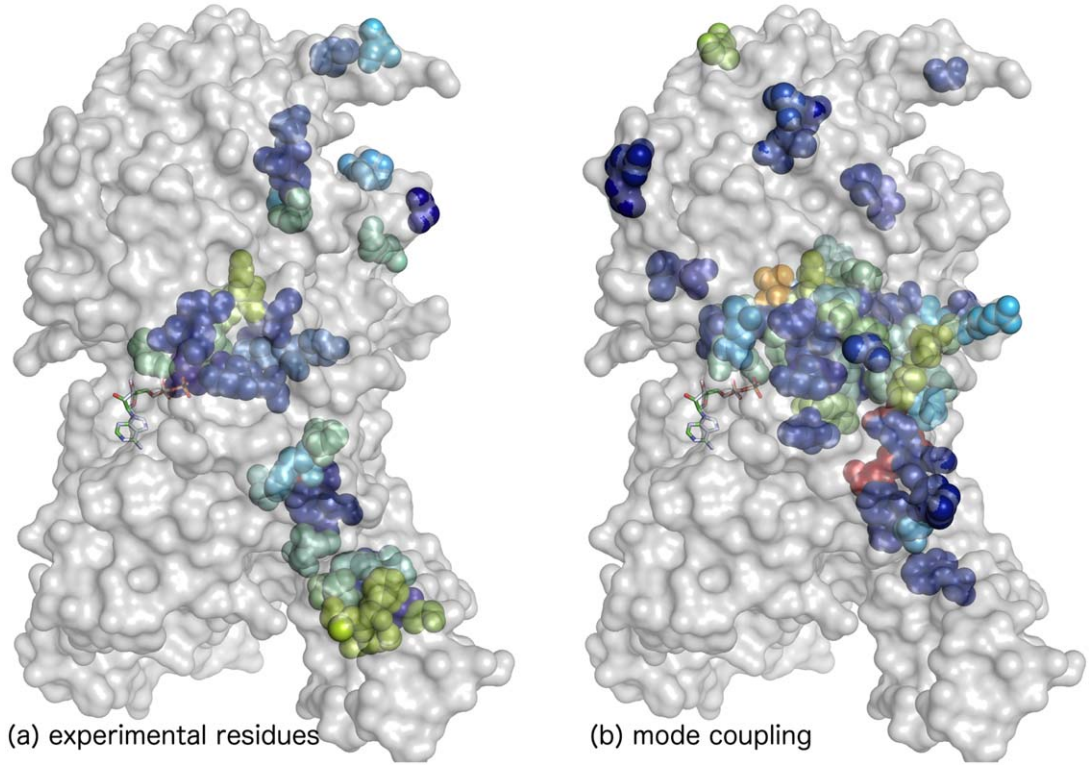

Figure 6

A side-by-side comparison of the locations of residues in (**a**) the experimentally determined target set and (**b**) top 10% with respect to mode-coupling scores. Hot colors indicate higher scores. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

high mode-coupling scores shown in Figure 6 is further encouraging in this respect. One observes that the top scoring residues (from the mode-coupling perspective) are not randomly distributed across the structure. Rather, they accumulate around the core of the motor domain, in visible agreement with the distribution of the residues in the target set in Table I.

Relevant modes and mode pairs

A key aspect of the above analysis is that the coupling between the modes is taken into account at all orders and is not restricted to pairwise interactions. Nevertheless, it is instructive to investigate the strongly coupled mode pairs and their influence in the real (residue) domain. To this end, we calculated the second-order coefficients $c^{ij}_{p,\nu-p}$ in Eq. (2) for the slowest 25 modes and determined the impact of the interactions between all $\binom{25}{2}$ mode pairs on the configurational free energy of the system, as described in Ref. 17. Repeating the analysis for the first and second half of the MD run, we obtained the coupling strength matrix given in Figure 7(a). In agreement with Figure 4(a–d), we observe that the modes that have the strongest pairwise coupling also change in time. In particular, 3–4, 1–4, and 1–6 are the prominent pairs in the first half of the simulation, while mode pairs 1–2 and 1–10 interact strongly in the second half. Consistently, upon analyzing fluctuations of Ser-443 (singled out in the first half) during the time evolution of each mode, we find that this residue is mostly driven by the modes 1, 3, 4, and 6 [Fig. 7(b)].

We finally computed, for each residue, the difference between fluctuation amplitudes due to the modes obtained from the first half and from the second half [Fig. 7(c)]. We found that the ligand-binding pocket residues [green/red/cyan columns in Fig.7(c)] mostly carry a positive signal, meaning they are driven more by the first mode set, while the opposite is true for the relay helix (yellow column). Although these observations support our earlier findings, we stress that, mode coupling has a significant "many-body" aspect and a pairwise analysis as above only partially reflects the impact of mode coupling on the protein's complex dynamics.

DISCUSSION

It is generally accepted that the coupling between different vibrational modes is an important physical mechanism driving correlated functional dynamics in proteins, particularly in the context of allosteric communication.

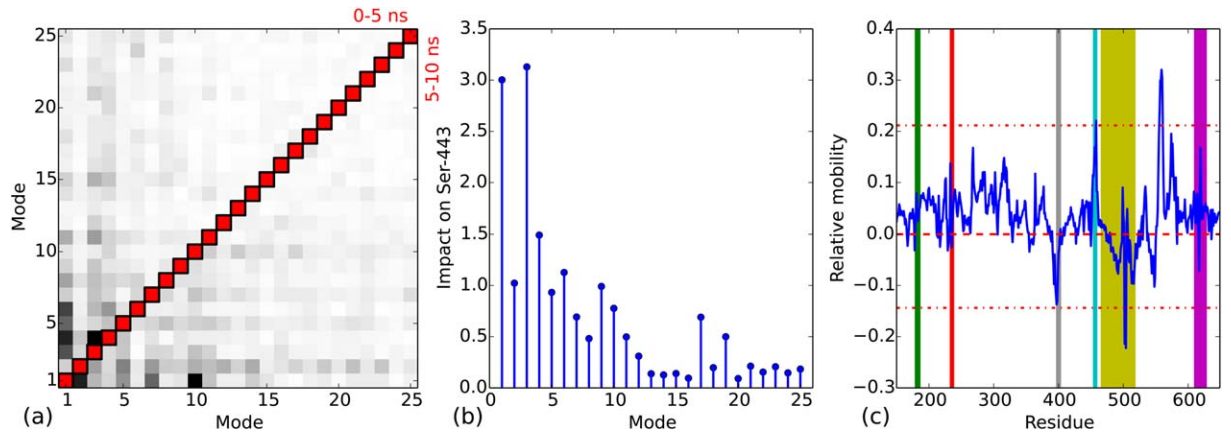

Figure 7

(**a**) Pairwise mode-coupling strengths for the slowest 25 modes. Above/below the red diagonal corresponds to the first/second half of the MD run. (**b**) Impact of each mode on the mobility of Ser-443. (**c**) The relative mobility of each residue due to the dominant modes in the two halves of the simulation. Positive values reflect higher contribution from the first half. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

However, this wisdom generated little input for experimentalists so far. We here propose an analytical/computational framework where the "noninteracting" limit is composed of already anharmonic modes (consistent with the observed slow modes in proteins). Mode coupling is then defined as everything that falls outside the best possible description of the configurational distribution (obtained from simulations) as a function factorizable in such modes. The information content of the mode-coupling contribution obtained by this operational definition can be utilized to highlight certain locations on the protein. Despite the fact that the MD simulations are much shorter than the time required to observe functionally relevant dynamics, we show here that these locations correlate with critical residues/regions obtained from experiments on the motor domain of myosin II.

Our work simultaneously confirms the relevance of mode coupling to function and proposes a new computational tool for predicting functionally critical locations on proteins. However, considering the multitude of factors that contribute to the evolutionary design of these complex machines, it is difficult to imagine our approach (or any other nonhybrid, *ab initio* method) to singlehandedly yield predictions sufficiently accurate for use in drug design or similar technologies. Assuming our tests currently in progress on several other proteins yield favorable results, a more apt use of the present computational framework would be to employ it as a module in a multifaceted prediction algorithm that seeks consensus between complementary approaches. Several such tools are publicly available.[37]

ACKNOWLEDGMENTS

We thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) for the grant MFAG-113F092 provided for the application of the present framework on a number of allosteric proteins. Part of the computing resources used during this work were provided by the National Center for High Performance Computing of Turkey (UYBHM) under grant number 4001752012.

REFERENCES

1. Bahar I, Rader AJ. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 2005;15:586–592.
2. Berendsen HJC, Hayward S. Collective protein dynamics in relation to function. Curr Opin Struct Biol 2000;10:165–169.
3. Tilton RF, Jr, Dewan JC, Petsko GA. Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease *a* at nine different temperatures from 98 to 320 K. Biochemistry 1992;31:2469–2481.
4. Roitberg A, Gerber RB, Elber R, Ratner MA. Anharmonic wave functions of proteins: Quantum self-consistent field calculations of BPTI. Science 1995;268:2.
5. Hayward S, Kitao A, Go N. Harmonic and anharmonic aspects in the dynamics of BPTI: A normal mode analysis and principal component analysis. Prot Sci Pub Prot Soc 1994;3:936–943.
6. Kitao A, Go N. Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol 1999;9:164–169.
7. Nakagawa H, Joti Y, Kitao A, Kataoka M. Hydration affects both harmonic and anharmonic nature of protein dynamics. Biophys J 2008;95:2916–2923.
8. Moritsugu K, Miyashita O, Kidera A. Vibrational energy transfer in a protein molecule. Phys Rev Lett 2000;85:3970–3973.
9. Xie A, van der Meer L, Hoff W, Austin RH. Long-lived amide I vibrational modes in myoglobin. Phys Rev Lett 2000;84:5435.
10. Leitner DM. Energy flow in proteins. Annu Rev Phys Chem 2008; 59:233–259.
11. Piazza F, Sanejouand Y-H. Long-range energy transfer in proteins. Phys Biol 2009;6:046014.
12. Piazza F, Sanejouand Y-H. Breather-mediated energy transfer in proteins. Discrete Cont Dyn Syst Ser S 2010;4:1247–1266.
13. Gur M, Erman B. Quasi-harmonic analysis of mode coupling in fluctuating native proteins. Phys Biol 2010;7:046006.

14. Monod J, Changeux J-P, Jacob F. Allosteric proteins and cellular control systems. J Mol Biol 1963;6:306–329.
15. Kern D, Zuiderweg ERP. The role of dynamics in allosteric regulation. Curr Opin Struct Biol 2003;13:748–757.
16. Tsai C-J, Sol AD, Nussinov R. Allostery: Absence of a change in shape does not imply that allostery is not at play. J Mol Biol 2008; 378:1–11.
17. Kabakç ioğlu A, Yuret D, Gür M, Erman B. IOP science—Anharmonicity, mode-coupling and entropy in a fluctuating native protein. Phys Biol 2010;7:046005.
18. Zheng W, Thirumalai D. Coupling between normal modes drives protein conformational dynamics: Illustrations using allosteric transitions in myosin II. Biophys J 2009; 96:2128–2137.
19. Yoon DY, Flory PJ. Moments and distribution functions for polymer chains of finite length. II. Polymethylene chains. J Chem Phys 1974; 61:5366.
20. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2:173–181.
21. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
22. Yogurtcu ON, Gur M, Erman B. Statistical thermodynamics of residue fluctuations in native proteins. J Chem Phys 2009;130:095103.
23. Zheng W, Brooks B. Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. J Mol Biol 2005;346:745–759.
24. Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951;22:79–86.
25. Rosenblatt M. Remarks on some nonparametric estimates of a density function. Ann Math Stat 1956;27:832–837.
26. Parzen E. On estimation of a probability density function and mode. Ann Math Stat 1962;33:1065–1076.
27. Silverman BW. Density estimation for statistics and data analysis (Chapman & Hall/CRC monographs on statistics and applied probability), 1st ed. Chapman and Hall/CRC; 1986.
28. Sweeney HL, Houdusse A. Structural and functional insights into the myosin motor mechanism. Annu Rev Biophys 2010;39:539–557.
29. Smith CA, Rayment I. X-ray structure of the magnesium(II)·ADP·vanadate complex of the dictyostelium discoideum myosin motor domain to 1.9 Å resolution. Biochemistry 1996;35:5404–5417.
30. Reubold TF, Eschenburg S, Becker A, Kull FJ, Manstein DJ, et al. A structural model for actin-induced nucleotide release in myosin. Nat Struct Biol 2003;10:826–830.
31. Spudich JA. How molecular motors work. Nature 1994;372:515–518.
32. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–1802.
33. Brooks BR, Brooks CL, III, Mackerell AD, Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The biomolecular simulation program. J Comput Chem 2009; 30:1545–1614.
34. Ma J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure 2005;13:373–380.
35. Pontiggia F, Colombo G, Micheletti C, Orland H. Anharmonicity and self-similarity of the free energy landscape of protein g. Phys Rev Lett 2007;98:048102.
36. Malnasi-Csizmadia A, Dickens JL, Zeng W, Bagshaw CR. Switch movements and the myosin crossbridge stroke. J Muscle Res Cell Motility 2005;26:31–37.
37. Prymula K, Jadczyk T, Roterman I. Catalytic residues in hydrolases: Analysis of methods designed for ligand-binding site prediction. J Comput Aided Mol Des 2011;25:117–133.
38. Tsiavaliaris G, Fujita-Becker S, Batra R, Levitsky DI, Kull FJ, Geeves MA, Manstein DJ. Mutations in the relay loop region result in dominant-negative inhibition of myosin II function in dictyostelium. EMBO Rep 2002;3:1099–1105.
39. Onishi H, Morales MF. A closer look at energy transduction in muscle. Proc Natl Acad Sci 2007;104:12714–12719.
40. Ruppel KM, Spudich JA. Structure–function studies of the myosin motor domain: Importance of the 50-kDa cleft. Mol Biol Cell 1996; 7:1123–1136.
41. Patterson B, Spudich JA. Cold-sensitive mutations of dictyostelium myosin heavy chain highlight functional domains of the myosin motor. Genetics 1996;143:801–810.
42. Sasaki N, Ohkura R, Sutoh K. Insertion or deletion of a single residue in the strut sequence of dictyostelium myosin II abolishes strong binding to actin. J Biol Chem 2000;275:38705–38709.
43. Tang S, Liao J-C, Dunn AR, Altman RB, Spudich JA, Schmidt JP. Predicting allosteric communication in myosin¡ i¿ via¡/i¿ a pathway of conserved residues. J Mol Biol 2007;373:1361–1373.
44. Frye JJ, Klenchin VA, Bagshaw CR, Rayment I. Insights into the importance of hydrogen bonding in the γ-phosphate binding pocket of myosin: Structural and functional studies of serine 236. Biochemistry 2010;49:4897–4907.
45. Shimada T, Sasaki N, Ohkura R, Sutoh K. Alanine scanning mutagenesis of the switch I region in the ATPase site of dictyostelium discoideum myosin II. Biochemistry 1997;36:14037–14043.
46. Sasaki N, Shimada T, Sutoh K. Mutational analysis of the switch II loop of dictyostelium myosin II. J Biol Chem 1998; 273:20334–20340.
47. Sasaki N, Ohkura R, Sutoh K. Dictyostelium myosin II mutations that uncouple the converter swing and ATP hydrolysis cycle. Biochemistry 2003;42:90–95.
48. Ito K, Uyeda TQP, Suzuki Y, Sutoh K, Yamamoto K. Requirement of domain-domain interaction for conformational change and functional ATP hydrolysis in myosin. J Biol Chem 2003;278:31049–31057.
49. Patterson B, Ruppel KM, Wu Y, Spudich JA. Cold-sensitive mutants G680V and G691C of dictyostelium myosin II confer dramatically different biochemical defects. J Biol Chem 1997;272:27612–27617.
50. Giese KC, Spudich JA. Phenotypically selected mutations in myosin's actin binding domain demonstrate intermolecular contacts important for motor function. Biochemistry 1997;36:8465–8473.