Mode-coupling points to functionally important residues in Myosin II

O. Varol, D. Yuret, B. Erman, A. Kabakçıoğlu

Colleges of Engineering[†] and Sciences[*], Koç University, Sarıyer, 34450, İstanbul, Turkey

E-mail: @ku.edu.tr

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.

1. 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 fluctution 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 higher-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, 9, 10, 11, 12, 13]. Characterization of the confirmational 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 (for example 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: Section 2 describes the theoretical framework used for isolating the contribution of mode-coupling from other anharmonic effects in the MD fluctuation data; Section 3 discusses how the formulation above can be used to select out residues that are highlighted by mode-coupling; Section 4 introduces the motor protein myosin II which we use here as a test case; Section 5 gives the details of the molecular dynamics (MD) simulations performed on myosin II; Section 6 reports and Section 7 discusses our results.

2. 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 Section 5. Using the MD trajectory, we derive a multivariate probability distribution function $p(\Delta\mathbf{R})$, where ΔR_i, ΔR_{i+N}, and ΔR_{i+2N} with $i = 1, ..., N$ are the deviations from the mean position along the coordinate axes x, y, and z, respectively, of the i-th C^α atom in a protein with N amino acids. The covariance matrix $\Gamma = \left\langle \Delta\mathbf{R}\Delta\mathbf{R}^T \right\rangle$ 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

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

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

2.1. 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 [17, 13]. In this framework, $f(\mathbf{\Delta r})$ is expressed as an infinite sum:

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

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. Ref.[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, 21, 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 higher-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 prefer to the former as *marginal anharmonicity*, since this contribution is uniquely determined by the deviations of the marginal distributions $f(\Delta r_i) = \int \prod_{j\neq i} dr_j f(\mathbf{\Delta r})$ from Gaussian.

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

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

where ν_{max} is a cut-off degree imposed by practical considerations (see Section 4). 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 $f(\mathbf{\Delta r})$ (i.e., $c_{p,\nu-p}^{ij}$ and higher-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.

3. 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., [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, higher-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 higher-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 $f \backslash f_1$. This information can be projected onto the protein's sequence axis by the procedure outlined in Section 3.1. 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.

3.1. Identifying per residue impact of mode-coupling

In order 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:

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

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

Next, we consider the marginal distributions $p(\mathbf{\Delta R_i}) = \int \prod_{j\neq i} \mathbf{dR_j}\ p(\mathbf{\Delta R})$ 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:

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

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

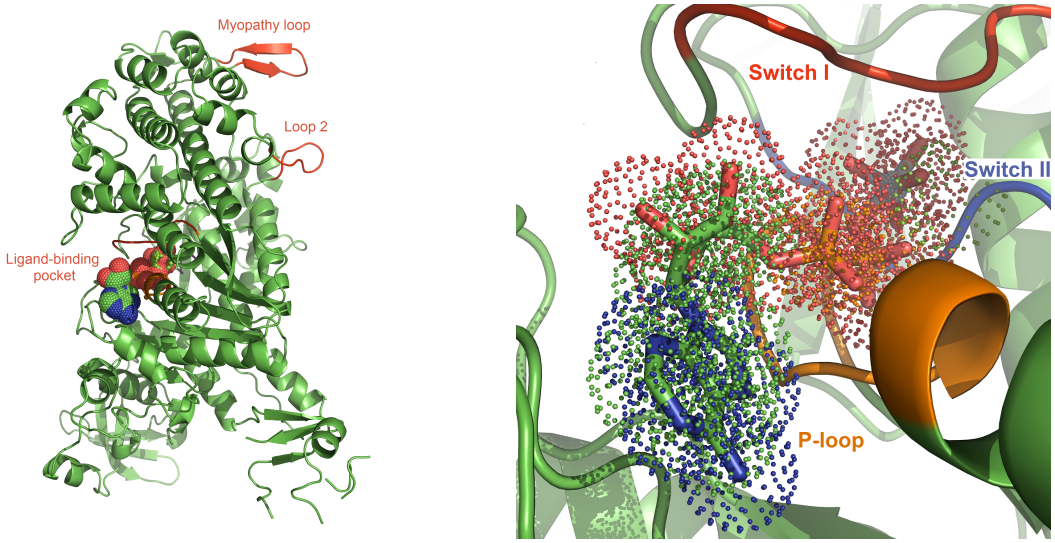
Above, "*mc*" and "*ma*" stand for "mode coupling" and "marginal anharmonicity", respectively. As a reference, we also consider the mean residue displacements (akin to experimental B-factors) measured by the variation

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

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 is available, estimation of the residue scores above is a mechanical process, without any tuning paramaters.

4. 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 Fig.6 goes through conformational changes at each stage of its four-stroke catalytic cycle which converts the chemical energy derived from ATP hydrolysis into mechanical work. The results presented below are obtained from MD simulations of the structure PDB:1VOM [29], where an ATP is bound on the protein.



(a) 1VOM structure (b) Ligand binding pocket

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). (1a)

There exist several, well-studied functionally relevant locations on the structure. Switch II and P-loop shown in Fig.1b 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 Fig.1a. 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).

5. 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$\mathring{A}$ cushion and periodic boundary conditions were applied. The integration time step in the simulation was selected as 1 fs for both non-bonded and electrostatic forces and no rigid bonds were used. The trajectory was captured every 50 fs within several windows of $\sim$ 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 elements in order to identify the components that are critical in transfer mechanical energy (except, thermal fluctuations are significantly more influential in the current system.)

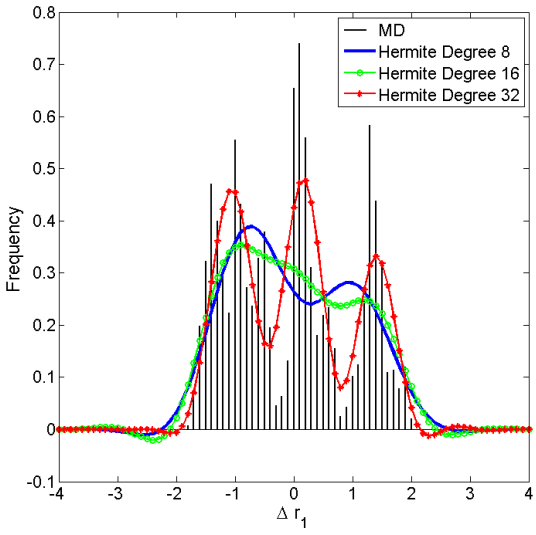


Figure 2: A comparison of the slowest mode's amplitude distribution for different choices for the maximum Hermite degree considered in Eq.(3). The cut-off degree $\nu_{max} = 32$ was determined according to the criterion that the marginal distributions for all modes are captured with an accuracy same as above or better.

The MD output was analysed 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 $N = 4000$ data points: We first construct the vector $\Delta\mathbf{R}^{(i)}$ of the C^α positions in each snapshot, with $i = 1, \ldots, N$. We next calculate the covariance matrix Γ given in Section 2, and identify the modal coordinates $\Delta\mathbf{R}^{(i)}$ for the 2184 fluctuation modes (out of $730 \times 3 = 2190$ 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 cut-off degree $\nu_{max} = 32$ which is obtained empirically (see Fig.2). Finally, we calculate the residue scores using Eqs.(6-8).

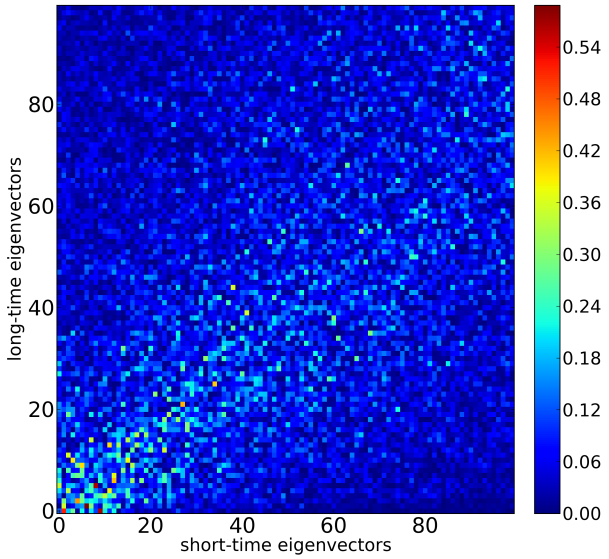


Figure 3: The overlap between the first 100 eigenvectors corresponding to the slowest fluctuational modes, ordered according to their eigenvalues, obtained from the first 1 ns (horizontal axis) and 10 ns (vertical axis) time frames. The accumulation along the lower diagonal indicates that the modal subspace spanned by slow eigenvectors retains its identity to a significant degree, with some amount of mixing between nearby modes.

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. Fig.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.

6. Results

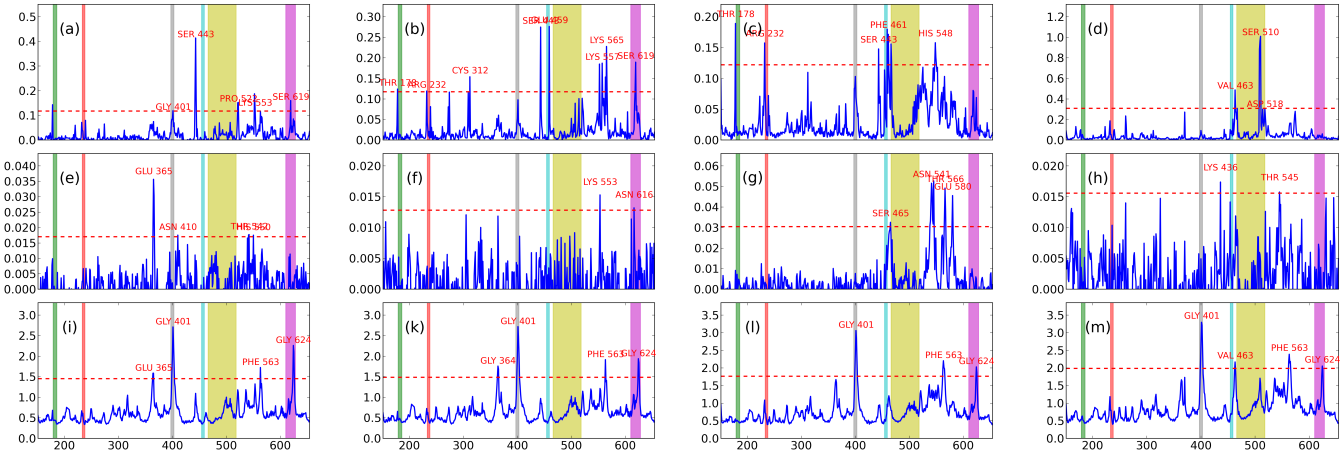


Figure 4: The residue scores obtained using Eqs.(6,7) in the first 1 ns, 2 ns, 5 ns, and 10 ns of the simulation, *w.r.t.* mode-coupling (first row), marginal anharmonicity (second row), and residue fluctuations (third row). Colored columns indicate functionally relevant regions, P-loop (green), Switch I (red), Myopathy loop (grey), Switch II (cyan), relay helix (yellow), and Loop-2 (purple), reported in the literature and described in Section 4. 3σ threshold is shown by the dashed line. The residue IDs of prominant peaks outside the 3σ margin are given in red.

Fig.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 Fig.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 rows of Fig.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 gathered from mode-coupling corrections yields a rank profile in significant agreement with site-specific mutation data available for myosin II, as demonstrated in Section 6.1.

A closer inspection of mode-coupling based scores in different time frames (first row 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 [**?**, 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 rows of Fig.4, we observe that there is some overlap between regions accentuated by mean fluctuation amplitudes *vs.* by mode coupling. 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 is 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.

6.1. Comparison with point mutation data

Score profiles in Fig.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 1 is a comprehensive list of 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 use a standard recall analysis to compare various ranking methods (including random) against this target set. Fig.5 shows for each case, the percentage of top residues that need to be considered (vertical axis) in order to capture a certain

N483, F487, I499, F506, L508, I687, F692, F745	Ref. [37]	D403, V405	Ref. [38]
N464, C470,N472,Y473, N475, F481, E746	Ref. [39]	Y494, W501	Ref. [37, 40]
D590, P591, L592, Q593	Ref. [41]	S181	Ref. [42]
E467, E586,G624, G740	Ref. [40]	S236	Ref. [43]
N233, S237, R238	Ref. [44]	E459	Ref. [45, 39]
I499, F692, R738	Ref. [46]	F482	Ref. [42, 47]
D454, G457, F458	Ref. [45]	G680	Ref. [42, 40, 48, 47]
E531, P536, R562	Ref. [49]	G691	Ref. [40, 48]

Table 1: Amino acids that are experimentally verified to be critical for myosin II function. 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.

fraction of the target set (horizontal axis). For a random ordering of the residues, this curve is supposed 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 a curve following the upper edge of the forbidden gray region at the bottom.

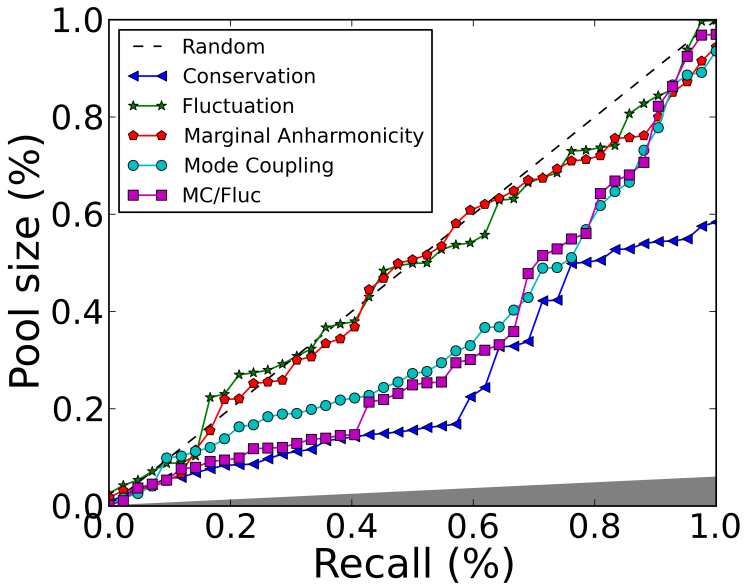


Figure 5: The performance of each ranking scheme considered in the paper is shown together with that of a random (dashed) and a perfect (upper edge of the forbidden black region) rank assignment. Each data point corresponds to the fraction of target residues captured out of a total of 43 (horizontal axis) by the top ranking residues covering a given percentage (vertical axis) of 730 residues in total.

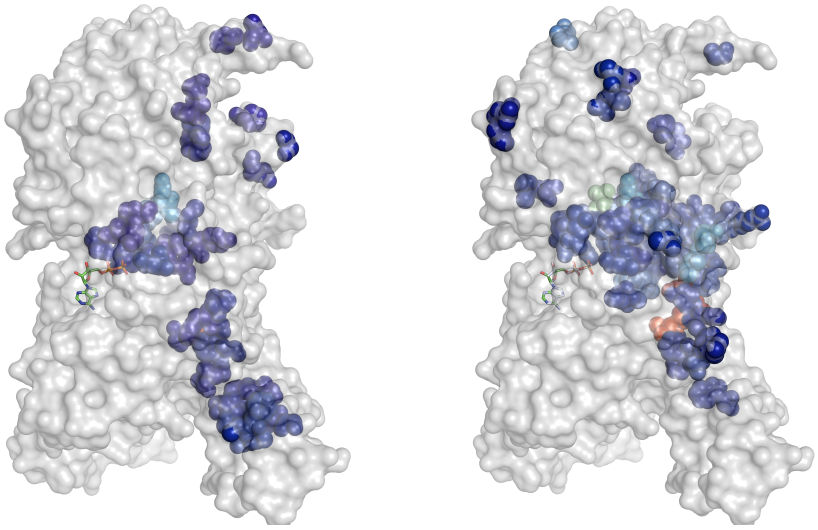
Upon inspecting Fig. 5, we observe that fluctuations and marginal anharmonicity display no preference for the target set. The conservation scores show the best correlation with the experimentally determined targets. This is hardly surprising, since experimental studies are in fact guided by conservation 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 $f_i \equiv S_i^{mc}/\sigma_i^2$ which is also shown in Fig.5 (the curve labeled as "MC/Fluc") is consistently better than S_i^{mc} in terms of highlighting the target set listed in Table 1. We checked that the general trend and the relative performances of different scoring criteria are robust under different methods one might choose while harnessing the information from different 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 Fig.5 between conservation and mode-coupling rankings. This discrepancy may be due to some residues in the target list which are relevant for the folding process (therefore also 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.

Figs.5 presents a direct validation of the fact that, coupling between vibrational modes, as formulated here, is a significant physical mechanism underlying myosin II's 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 endevours as a new guide for target selection (for example, Ser-443 appears to be a promising candidate). The distribution of top amino acids with high mode-coupling scores shown in Fig. 6 is further encouraging in this respect. One observes that, the top scoring residues 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 1.

7. 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. 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



(a) experimental residues (b) mode coupling

Figure 6: A side-by-side comparison of the locations of residues in (a) top 10% w.r.t. mode-coupling scores, (b) the experimentally determined target set. Hot colors indicate higher scores.

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 these complex machines, it is difficult to imagine our approach (or any other non-hybrid, *ab initio* method) to singlehandedly yield sufficiently accurate predictions. 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 [50].

Acknowledgments

We thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) for the grant MFAG-113F092 awarded to support 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] Ivet Bahar and AJ Rader. Coarse-grained normal mode analysis in structural biology. *Current opinion in structural biology*, 15(5):586–592, 2005.

[2] Herman JC Berendsen and Steven Hayward. Collective protein dynamics in relation to function. *Current opinion in structural biology*, 10(2):165–169, 2000.

[3] Robert F Tilton Jr, John C Dewan, and Gregory A Petsko. Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-a at nine different temperatures from 98 to 320k. *Biochemistry*, 31(9):2469–2481, 1992.

[4] Adrian Roitberg, R Benny Gerber, Ron Elber, and Mark A Ratner. Anharmonic wave functions of proteins: Quantum self-consistent field calculations of bpti. *Science*, 268:2, 1995.

[5] S. Hayward, A. Kitao, and N. Go. Harmonic and anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis. *Protein science : a publication of the Protein Society*, 3(6):936–943, June 1994.

[6] Akio Kitao and Nobuhiro Go. Investigating protein dynamics in collective coordinate space. *Current opinion in structural biology*, 9(2):164–169, 1999.

[7] H Nakagawa, Y Joti, A Kitao, and M Kataoka. Hydration affects both harmonic and anharmonic nature of protein dynamics. *Biophysical journal*, 95(6):2916–2923, 2008.

[8] Kei Moritsugu, Osamu Miyashita, and Akinori Kidera. Vibrational Energy Transfer in a Protein Molecule. *Physical Review Letters*, 85(18):3970–3973, October 2000.

[9] Aihua Xie, Lex van der Meer, Wouter Hoff, and Robert H Austin. Long-lived amide i vibrational modes in myoglobin. *Physical review letters*, 84(23):5435, 2000.

[10] David M Leitner. Energy flow in proteins. *Annu. Rev. Phys. Chem.*, 59:233–259, 2008.

[11] Francesco Piazza and Yves-Henri Sanejouand. Long-range energy transfer in proteins. *Physical Biology*, 6(4):046014+, November 2009.

[12] Francesco Piazza, Yves-Henri Sanejouand, et al. Breather-mediated energy transfer in proteins. *Discrete and Continuous Dynamical Systems, Series S*, 4:1247–1266, 2010.

[13] Mert Gur and Burak Erman. Quasi-harmonic analysis of mode coupling in fluctuating native proteins. *Physical Biology*, 7(4):046006, 2010.

[14] Jacques Monod, Jean-Pierre Changeux, and Francois Jacob. Allosteric proteins and cellular control systems. *Journal of molecular biology*, 6(4):306–329, 1963.

[15] Dorothee Kern and Erik RP Zuiderweg. The role of dynamics in allosteric regulation. *Current opinion in structural biology*, 13(6):748–757, 2003.

[16] Chung-Jung Tsai, Antonio Del Sol, and Ruth Nussinov. Allostery: absence of a change in shape does not imply that allostery is not at play. *Journal of molecular biology*, 378(1):1–11, 2008.

[17] A. Kabakçıoğlu, D. Yuret, M. Gür, and B. Erman. IOPscience - Anharmonicity, mode-coupling and entropy in a fluctuating native protein. *Physical Biology*, 7(4):046005+, December 2010.

[18] Wenjun Zheng and D Thirumalai. Coupling between normal modes drives protein conformational dynamics: illustrations using allosteric transitions in myosin ii. *Biophysical journal*, 96(6):2128–2137, 2009.

[19] DY Yoon and PJ Flory. Moments and distribution functions for polymer chains of finite length. ii. polymethylene chains. *The Journal of Chemical Physics*, 61:5366, 1974.

[20] I. Bahar, A.R. Atilgan, and B. Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, 2(3):173–181, 1997.

[21] AR Atilgan, SR Durell, RL Jernigan, MC Demirel, O. Keskin, and I. Bahar. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.*, 80(1):505–515, 2001.

[22] O.N. Yogurtcu, M. Gur, and B. Erman. Statistical thermodynamics of residue fluctuations in native proteins. *J. Chem. Phys.*, 130(9), 2009.

[23] Wenjun Zheng and Bernard Brooks. Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. *Journal of molecular biology*, 346(3):745–759, 2005.

[24] S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[25] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pages 832–837, 1956.

[26] E. Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

[27] Bernard. *Density Estimation for Statistics and Data Analysis (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)*. Chapman and Hall/CRC, 1 edition, April 1986.

[28] H. Lee Sweeney and Anne Houdusse. Structural and functional insights into the Myosin motor mechanism. *Annual review of biophysics*, 39:539–557, June 2010.

[29] Clyde A. Smith and Ivan Rayment. X-ray Structure of the Magnesium(II)ADPVanadate Complex of the Dictyostelium discoideum Myosin Motor Domain to 1.9 Å Resolution†,‡. *Biochemistry*, 35(17):5404–5417, January 1996.

[30] T.F. Reubold, S. Eschenburg, A. Becker, F.J. Kull, D.J. Manstein, et al. A structural model for actin-induced nucleotide release in myosin. *Nature structural biology*, 10(10):826–830, 2003.

[31] J. A. Spudich. How molecular motors work. *Nature*, 372(6506):515–518, December 1994.

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

[33] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The Biomolecular Simulation Program. *JOURNAL OF COMPUTATIONAL CHEMISTRY*, 30(10, Sp. Iss. SI):1545–1614, JUL 30 2009.

[34] Jianpeng Ma. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, 13(3):373–380, 2005.

[35] Francesco Pontiggia, Giorgio Colombo, Cristian Micheletti, and Henri Orland. Anharmonicity and self-similarity of the free energy landscape of protein g. *Physical review letters*, 98(4):048102, 2007.

[36] Andras Malnasi-Csizmadia, JANE L Dickens, Wei Zeng, and Clive R Bagshaw. Switch movements and the myosin crossbridge stroke. *Journal of Muscle Research & Cell Motility*, 26(1):31–37, 2005.

[37] Georgios Tsiavaliaris, Setsuko Fujita-Becker, Renu Batra, Dmitrii I. Levitsky, F. Jon Kull, Michael A. Geeves, and Dietmar J. Manstein. Mutations in the relay loop region result in dominant-negative inhibition of myosin II function in Dictyostelium. *EMBO reports*, 3(11):1099–1105, November 2002.

[38] Hirofumi Onishi and Manuel F. Morales. A closer look at energy transduction in muscle. *Proceedings of the National Academy of Sciences*, 104(31):12714–12719, July 2007.

[39] K. M. Ruppel and J. A. Spudich. Structure-function studies of the myosin motor domain: importance of the 50-kDa cleft. *Molecular biology of the cell*, 7(7):1123–1136, July 1996.

[40] B. Patterson and J. A. Spudich. Cold-sensitive mutations of Dictyostelium myosin heavy chain highlight functional domains of the myosin motor. *Genetics*, 143(2):801–810, June 1996.

[41] Naoya Sasaki, Reiko Ohkura, and Kazuo Sutoh. Insertion or deletion of a single residue in the strut sequence of dictyostelium myosin ii abolishes strong binding to actin. *Journal of Biological Chemistry*, 275(49):38705–38709, 2000.

[42] Susan Tang, Jung-Chi Liao, Alexander R Dunn, Russ B Altman, James A Spudich, and Jeanette P

Schmidt. Predicting allosteric communication in myosin¡ i¿ via¡/i¿ a pathway of conserved residues. *Journal of molecular biology*, 373(5):1361–1373, 2007.

[43] Jeremiah J. Frye, Vadim A. Klenchin, Clive R. Bagshaw, and Ivan Rayment. Insights into the Importance of Hydrogen Bonding in the γ-Phosphate Binding Pocket of Myosin: Structural and Functional Studies of Serine 236,. *Biochemistry*, 49(23):4897–4907, May 2010.

[44] T. Shimada, N. Sasaki, R. Ohkura, and K. Sutoh. Alanine scanning mutagenesis of the switch I region in the ATPase site of Dictyostelium discoideum myosin II. *Biochemistry*, 36(46):14037–14043, November 1997.

[45] N. Sasaki, T. Shimada, and K. Sutoh. Mutational analysis of the switch II loop of Dictyostelium myosin II. *The Journal of biological chemistry*, 273(32):20334–20340, August 1998.

[46] Naoya Sasaki, Reiko Ohkura, and Kazuo Sutoh. Dictyostelium myosin ii mutations that uncouple the converter swing and atp hydrolysis cycle. *Biochemistry*, 42(1):90–95, 2003.

[47] Kohji Ito, Taro Q. P. Uyeda, Yoshikazu Suzuki, Kazuo Sutoh, and Keiichi Yamamoto. Requirement of Domain-Domain Interaction for Conformational Change and Functional ATP Hydrolysis in Myosin. *Journal of Biological Chemistry*, 278(33):31049–31057, August 2003.

[48] B. Patterson, K. M. Ruppel, Y. Wu, and J. A. Spudich. Cold-sensitive mutants G680V and G691C of Dictyostelium myosin II confer dramatically different biochemical defects. *The Journal of biological chemistry*, 272(44):27612–27617, October 1997.

[49] Kim C. Giese and James A. Spudich. Phenotypically Selected Mutations in Myosin's Actin Binding Domain Demonstrate Intermolecular Contacts Important for Motor Function†. *Biochemistry*, 36(28):8465–8473, July 1997.

[50] Katarzyna Prymula, Tomasz Jadczyk, and Irena Roterman. Catalytic residues in hydrolases: analysis of methods designed for ligand-binding site prediction. *Journal of computer-aided molecular design*, 25(2):117–133, 2011.