Functional Motions in Biomolecules: Insights from Computational Studies at Multiple Scales

For “Computational Structural Biology”, Eds. T. Schwede, M. Peitsch

1Adam W. van Wynsberghe, 1Liang Ma, 2Xi Chen and 1Qiang Cui

1Department of Chemistry and Theoretical Chemistry Institute,
University of Wisconsin, Madison, 1101 University Ave, Madison, WI 53706 and
2Department of Civil Engineering and Engineering Mechanics,
Columbia University, New York City, NY 10027

(Dated: June 27, 2007)

Abstract

Motions at both the domain and local scales are important to the function of biomolecules. Computational techniques for probing these functional motions are discussed. These include atomistic simulations that characterize the energetics of local motions, various normal-mode based methods that capture the directionality of domain scale motions as well as effective coarse-grained methods that are necessary for probing motions at very large length and time scales. The values and limitations of these techniques are illustrated by selected applications that analyzed the role of local motions in enzyme catalysis, mechanochemical coupling in signaling proteins and biomolecular motors, and gating of the mechanosensitive channel. A number of outstanding and emerging questions regarding functional motions in biomolecular systems are briefly discussed.
I. INTRODUCTION

Mounting evidence from numerous experimental\textsuperscript{1,2} and computational studies\textsuperscript{3} has demonstrated that biomolecules have motions that span a wide range of time and spatial scales. Some of those motions reflect the importance of maintaining a “minimal” level of flexibility for function. For example, a recent insightful analysis\textsuperscript{4} examined the magnitude of atomic fluctuations in proteins using data from both molecular dynamics simulations and crystallographic Debye-Waller factors. Based on the Lindemann criterion, atomic fluctuations indicate that the surface of proteins is liquid like while the core is solid like. This result makes intuitive sense in that the solid core is important for stability while the fluidic surface is essential for the structural changes required by basic functions such as ligand binding. As the temperature approaches the so-called “glass-transition” temperature ($\sim$ 180 K for many proteins), the Lindemann criterion suggests that the entire protein becomes solid like; at the same temperature, most proteins lose their ability to function.

In addition to such “generic” thermal fluctuations, it is generally agreed that there are also “functional motions”, which have specific characters (in direction, magnitude and time-scale) that make these motions essential to the unique function of a particular biomolecule. These range from structural transitions at the domain scale, which are implicated in the function of many “biomolecular machines”\textsuperscript{5} and multi-subunit enzymes,\textsuperscript{6} to relatively localized vibrations that have been proposed to facilitate chemical reactions.\textsuperscript{7} In this regard, we note that a rather broad notion of “motion” is adopted here, which includes both equilibrium fluctuations in a single state \textit{and} structural transitions between two (or more) distinct functional states of a system.

Despite their biological importance, functional motions are difficult to identify and characterize at a quantitative level. The multiple length and time scales spanned by these motions pose tremendous challenges to experimental measurements and their interpretation. A significant body of studies has demonstrated that careful computational studies can nicely complement experimental work for better characterizing and understanding the working mechanism of functional motions. In the following, we first briefly review several computational methods that are particularly useful for studying motions in biomolecules at multiple scales; then, we discuss a few examples from our labs to illustrate the value and limitation of these techniques as well as the mechanistic insights derived from compu-
tional analyses regarding the nature and functional implication of specific motions in the corresponding systems. Finally, a number of outstanding and emerging questions regarding functional motions in biomolecular systems are briefly discussed.

II. COMPUTATIONAL METHODS

In principle, the most robust computational approach for studying motions in biomolecules is atomistic molecular dynamics (MD). Ever since the first molecular dynamics simulations of proteins thirty years ago, striking progresses have been made in both theoretical/computational algorithms and computational hardwares. As a result, sophisticated molecular dynamics simulations have become an indispensable tool in the analysis of structural, energetic and dynamical properties of biomolecules. Nevertheless, for many processes, such as domain motions, atomistic molecular simulations are still too expensive for obtaining statistically meaningful results. Even for relatively local structural transitions, it is challenging to quantify the underlying thermodynamics and kinetics using straightforward molecular dynamics simulations. In those cases, alternative computational approaches have to be used. In the following, we briefly summarize a few computational approaches that are useful for characterizing motions at different scales and evaluating the functional significance of these motions.

A. Local Motions: Advanced Atomistic Molecular Dynamics

Local motions such as sidechain flips, loop displacements and break-formation of salt-bridge interactions play important roles in many systems. For example, isomerization of a Histidine from a buried configuration to a solvent exposed orientation is implicated in its proton shuttling function in carbonic anhydrase; the closure of a “lid” composed of a 11-residue loop sets up the active site of triosephosphate isomerase to avoid side reactions; the formation of a critical salt-bridge between two loop motifs in myosin helps to align water molecules properly in the nucleotide binding sites for the subsequent hydrolysis of ATP. Due to the presence of free energy barriers higher than \( k_B T \), characterizing the corresponding thermodynamics and kinetics (barriers) for local structural changes is not always straightforward. In principle, these quantities can be estimated from the relevant
potential of mean force (PMF, \( W(\xi) \))\(^{14} \) profile (Fig.1a) using umbrella samplings\(^{15} \) to obtain

\[
W(\xi) = -k_B T \ln P(\xi) + C
\]  

(1)

where \( \xi \) is the chosen reaction coordinate and \( P(\xi) \) is the probability distribution along \( \xi \), \( C \) is a normalization constant.

In practice, however, even localized structural changes may implicate variations in a handful geometrical parameters and it can be difficult to identify the most important one(s) as the principal “reaction coordinate(s)”\(^{16} \); computing PMF along an inappropriate reaction coordinate may lead to significant error in the computed energetics, especially barriers.\(^{16} \) The situation can be even more complex if there is significant involvement of the solvent degrees of freedom; this might be more prevalent than one may naively assume, and even the isomerization of an alanine dipeptide, for example, has been shown to implicate significant solvent participation.\(^{17} \) Another example along this line is the dimeric hemoglobin in scallop, where a change in the number of interfacial water molecules is coupled to the rotation of a Phe residue at the dimer interface and key to the allosteric communication between the two subunits.\(^{18} \)

In other words, a major challenge for quantifying local motions (including chemical reactions) is the identification of variables whose changes best describe the kinetic bottleneck of the process. In some applications, experience and intuition can be very instructive (see the following discussion on CheY). Nonetheless, a less ad hoc approach is highly desirable. In this context, the transition path sampling (TPS) technique proposed by Chandler and co-workers\(^{19} \) provides a theoretically sound framework for studying reactive processes (either chemical or structural) in complex systems like biomolecules. Unlike the minimum energy path analysis, which is powerful for gas-phase processes but significantly less appropriate for processes in the condensed phase, TPS collects real-time “reactive trajectories” and therefore samples the true kinetic bottleneck and includes entropic effects (Fig.1b). As described in details in Ref.\(^{20} \), TPS employs a Monte Carlo procedure to sample the trajectory space with emphasis on reactive trajectories that lead to the structural transitions of interest. This is possible because most local structural transitions are thermally activated, meaning that the rate is low due to significant (free) energy barriers but the barrier crossing process itself, once activated, is fast (often in the picosecond regime).\(^{21} \)

Briefly, a TPS simulation starts with a single reactive trajectory, which can be obtained in
a number of ways such as by forcing the relevant structural transition to occur via artificial
restraints and then gradually reducing the strength of the restraints in a series of “annealing”
simulations.\textsuperscript{22} Then new trajectories derived by slightly perturbing the existing trajectory
(e.g., by modifying the velocity of certain atoms in a frame) will be generated and accepted
based on a Metropolis criterion to ensure detailed balance (i.e., the proper canonical weights
of trajectories). This is carried out iteratively until a significant number of uncorrelated
reactive trajectories have been collected; the precise number depends on the system and the
goal of an application.

Clearly, the TPS approach is computationally intense and typically involves at least
collecting thousands of short trajectories on the order of 10-100 ps. Therefore, TPS is
ideally suited for studying relatively local structural transitions in biomolecules where the
process can be too complex to characterize with a few “obvious” choice of variables but the
intrinsic transition time-scale is still well in the subnanosecond regime. It is important to
recognize, however, that the standard TPS protocol is still a local sampling technique and
therefore the results are likely dependent on the initial “reactive trajectory” (see example
below).\textsuperscript{23} Other challenges associated with TPS include analyzing the results,\textsuperscript{24} dealing with
processes of diffusive nature\textsuperscript{25} and/or involving intermediate(s); given the limited space here,
we refer the readers to recent discussions in the literature.

\textbf{B. Domain Motions: Normal Mode Analysis}

Large-scale structural transitions at the domain scale are involved in many “biomolecular
machines” such as molecular motors\textsuperscript{5} and allosteric multi-subunit enzymes.\textsuperscript{2} They are diffi-
cult to study using regular atomistic simulations because they occur at time scales typically
at or longer than \textit{ms}. Various “unconventional” molecular dynamics techniques have been
proposed accordingly, which either applies specific biasing potentials to artificially speed
up the structural transitions\textsuperscript{26,27} or aims at identifying the approximate transition path(s)
between two functional states.\textsuperscript{28}

One interesting alternative that has found great popularity in recent years is the normal
mode analysis (NMA).\textsuperscript{29} In NMA, one approximates the motion of the system as harmonic
vibrations around a local minimum on the potential energy surface. Following the diag-
onalization of the force constant (second-order derivative, or the hessian, \textbf{H}) matrix in
mass-weighted coordinates,

\[ \mathbf{H} \mathbf{l}_i = \omega_i^2 \mathbf{l}_i, \quad i = 1, \cdots, 3N - 6 \quad (2) \]

the equations of motion can then be simplified as a set of uncoupled harmonic oscillators of frequencies \( \{\omega_i\} \); here \( N \) is the total number of atoms. Through the eigenvectors, \( \mathbf{l}_i \), the time evolution of the Cartesian coordinates \( (q_j(t)) \) can be expressed analytically at all time,

\[ q_j^\alpha(t) = m_j^{-1/2} \sum_{i=1}^{3N-6} [A_i \cos(\omega_i t + \phi_i)] I_{ji}^\alpha, \quad j = 1, \cdots, N; \alpha = x, y, z \quad (3) \]

where \( A_i, \phi_i \) are the amplitude and phase factor for the \( i \)-th mode; this allows the calculation of many thermally averaged results such as atomic fluctuations at a given temperature \( T \).\(^{14}\)

Although clearly approximate, a significant body of research has demonstrated that NMA is uniquely useful for characterizing collective motions in biomolecules (Fig.1c).\(^{30-33}\) In particular, large-scale structural transitions between different functional states have been found to correlate very well with the low-frequency normal modes; in many cases, in fact, a large fraction of the structural transitions can be expressed as the linear combination of a very small number of low-frequency normal modes. This leads to the idea that the flexibility required for the functional transitions is an inherent feature of the system encoded by the structure.

The fact that low-frequency modes are most relevant for characterizing domain-scale motions suggests that further approximations can be made to NMA such that the efficiency of the computation can be improved. One idea is to divide the system into a set of “blocks” (e.g., one amino acid per block), and then ignore the internal motion of the blocks when solving the NMA problem;\(^{34}\) this significantly reduces the size of the eigenvalue problem.\(^{35}\) Such a “block normal mode” approach has been shown to give very reliable results for low-frequency eigenvalues/eigenvectors and therefore can be used to explore structural flexibilities of very large biomolecular complexes (such as protein-DNA complexes and the ribosome) with atomistic interactions.\(^{36,37}\)

A further approximation is to simplify the interaction potential into that of a set of elastic springs, which leads to the “elastic network model (ENM)”.\(^{38}\)

\[ U_{\text{elastic}} = \gamma \sum_{i \neq j} \Theta(r_{ij}^0 - r_{\text{cut}})(r_{ij} - r_{ij}^0)^2 \quad (4) \]
where $\Theta$ is a Heaviside step function, $r_{\text{cut}}$ a parameter that determines the range of interactions and $\gamma$ a scaling factor; $r_{ij}^0$ is the distance between atoms $i$ and $j$ in the current structure. In addition to its computational efficiency, a nice feature of ENM is that the potential ensures that the current structure is the energy minimum and therefore no energy minimization is needed. A large body of studies have shown that despite its simplicity, ENM produces reliable results for the low-frequency eigenvectors for compact structures\textsuperscript{32,33,39} although the results tend to deteriorate as the frequency increases.\textsuperscript{40} In a recent analysis,\textsuperscript{41} for example, results from several variations of ENM and the block normal mode (BNM) using an atomistic potential were compared to the anisotropic displacement parameters (ADPs) from high-resolution x-ray structures. It was found that most methods produce favorable agreement with the experiment ADPs, although there are notable differences between the eigenvectors from ENM and BNM calculations except for the first few modes. For very large systems, reliable ENM results can be obtained with a significantly smaller number of interaction sites than the number of atoms. This leads to the impressive application of ENM to systems with low-resolution structural information (such as electron microscopy maps).\textsuperscript{42,43}

It is important to emphasize that by “characterizing” domain motions with normal modes, we mean that the directions of large-scale flexibility correlate well with a small number of low-frequency normal modes. The time-scale and magnitude of motions along these directions, however, are beyond the capability of the NMA approaches discussed above (see Sect.IV). Moreover, care has to be exercised when interpreting “correlated motions” in biomolecules using only a small number of modes.\textsuperscript{44}

C. Coarse-grained Models Beyond the Harmonic Limit

Another active area of research involves developing effective coarse-grained models such that anharmonic motions (thus beyond normal modes) at long-time scale and large spatial scale can be effectively simulated and analyzed. The most popular approach is to reduce the resolution of the representation by grouping several atoms into a single bead and then parameterize effective interactions between the beads.\textsuperscript{45–47} Different approaches have been proposed for parameterizing effective interactions based on atomistic simulations, and it remains a challenging task to develop potentials that are accurate, flexible and transferrable at the coarse-grained level.\textsuperscript{48} Nevertheless, impressive results have been obtained for a number
of systems, perhaps most notably for lipid systems by Marrink and co-workers.\textsuperscript{49}

An alternative direction that we have been exploring recently is to adopt a continuum mechanics framework with the finite element (FEM) representation. Although continuum mechanics have been used in the past to model the mechanical behavior of biomolecules,\textsuperscript{50,51} they are usually based on highly idealized geometries and materials properties. The FEM analysis,\textsuperscript{52} on the other hand, is widely used in the engineering field for solving mechanical and transport problems and can be applied to systems with complex geometries and boundary conditions.

In our recent study, we have established a proof-of-concept continuum mechanics model for the mechanosensitive channel of the large conductance (MscL).\textsuperscript{53} Inspired by its crystal structure,\textsuperscript{54} this model treats transmembrane helices of MscL as elastic rods and the lipid membrane as an elastic sheet of finite thickness; in the more recent model,\textsuperscript{55} periplasmic loops are also included as elastic springs. In the FEM framework, these continuum components (indicated as domain $\Omega$, are represented by a set of “elements” in the shape of, for example, tetrahedra (Fig.1d). The size of the elements can be determined adaptively, small for regions of interest and large for far-away areas, which makes the simulation framework ideal for very large systems such as a protein complex embedded in a large sheet of membrane. Similar to atoms in particle-based simulations, each element is associated with the materials’ properties (e.g., Young’s moduli) and parameters that describe inter-element interactions. These important parameters can be derived from calculations using all-atom force fields. The interaction between the continuum components and the surrounding solvent can be treated at the Poisson-Boltzmann level.\textsuperscript{56} Currently, we are developing, in a systematic manner, a semi-quantitative framework that treats irregular shapes of continuum components and employs more sophisticated description of materials properties. The solution of an unknown variable (function), $\phi$, is then approximated by a series of shape functions, $s_i$, and a set of unknown parameters $a_i$ (e.g., nodal displacements), as, $\phi \approx \hat{\phi} = \sum_i a_i s_i$. The values of $a_i$ are then determined from equations established from, for example, a variational principle,

$$\int_{\Omega} G(\hat{\phi}) d\Omega + \int_{\Gamma} g(\hat{\phi}) d\Gamma = 0$$

(5)

where $\Gamma$ is the domain boundary and $g, G$ are the relevant energy/work functionals.

Since most interactions in a FEM model are local in nature, the cost of the simulation is modest. Therefore, once parameterized, the continuum mechanics model is ideally suited
for studying the structural response of the biomolecule to various external mechanical perturbations of different form and scale. In the simplest application, this involves applying mechanical loads as the boundary condition and evolve the structure of the continuum components (i.e., positions of the FEM nodes) in a quasi-static fashion. Even at this level, interesting insights can be obtained. For example, qualitatively different responses of MscL were observed when the membrane was subject to in-plane tension vs. out-of-plane bending (see below);\textsuperscript{53} this is very difficult to achieve with any other widely available simulation techniques. At a more sophisticated level, the real-time dynamics of the continuum system at the finite temperature can be monitored by propagating Langvin dynamics; the potential of such studies on biomolecular systems, however, remain to be fully developed and explored.

III. APPLICATIONS

In this section, we discuss a few applications from our own studies to illustrate how the techniques outlined above are used to provide useful insights into functional motions in biomolecules at multiple scales.

A. Functional “Dynamics” of Ribonuclease A

Whether there are specific motions (or loosely referred to as “dynamics” in the relevant literature) in enzymes that facilitate the catalytic step has been a topic of intense interest. An intriguing recent example is Ribonuclease A, for which the motion-catalysis relationship has been analyzed in details by Loria and co-workers using NMR relaxation measurements.\textsuperscript{57,58} The most interesting finding concerns the effect of mutating Asp 121 near the active site (Fig.2a) to an alanine. More than 95% of the activity is lost upon mutation, although neither substrate affinity nor the electrostatic properties of the active site (e.g., as reflected by the catalytic His residues) was significantly perturbed.\textsuperscript{58–60} There was, however, interesting changes in the $\mu s \rightarrow ms$ motions upon mutation. In the WT enzyme, the motions of different motifs (e.g., loop 1, 4 and His 119) are very close in time scale to each other and to the observed catalytic rate; in the D121A mutant, by contrast, the time-scale for the motions of different structural motifs become substantially different and product release (the rate limiting step in the wild type) actually became faster. These observations led Loria and
coworkers to suggest that the synchronicity of global protein motions plays an important role in determining the rates of catalytically important steps, and the loss of catalysis in the D121A mutant is from the disruption of these global dynamics.58

**Atomic scale hypothesis for D121A effects** Although the hypothesis that the global millisecond dynamics of RNase A are “coordinated” and “timed” to help catalysis occur is intriguing, it is nonetheless difficult to imagine a detailed atomic level mechanism. We present an alternative hypothesis for the decrease in catalytic rate that involves changes in the free energy landscape of the *apo enzyme* and then attempt to verify this hypothesis using PMF simulations.

The catalytically active His 119’s side chain has been observed to exist in two configurations in both crystallographic and NMR experiments.58,61,62 These two configurations are defined by the *trans* (≈180°) or *gauche*+ (≈−60°) rotamers of the sidechain χ1 dihedral angle and are known as the “A” and “B” sites, respectively (Fig.2b). The enzyme is only active when His119 is in the A site and A/B site conformational exchange is very unlikely if substrate is bound. If one assumes that substrate can bind to RNase A when it is in the B conformation, this complex would be unreactive. The kinetic scheme for such a situation, where the enzyme can transition between an active and inactive form with both having the ability to bind substrate is given by

$$E_I \cdot S \xrightleftharpoons[k_{-4}]{k_4} E_I + S \xrightleftharpoons[k_{-3}]{k_3} E_A + S \xrightleftharpoons[k_{-1}]{k_1} E_A \cdot S \xrightleftharpoons[k_2]{k_3} E_A + P$$

(6)

Analyzing this scheme with the steady state approximation and the assumption that the inactive and active forms bind substrate with similar affinities results in a modified Michaelis-Menten equation where the apparent catalytic rate, $k_2$ is multiplied by the fraction of active enzyme, $f_A$,

$$v = \frac{f_A k_2 [E_I][S]}{K_M + [S]}$$

(7)

If the assumptions of the above analysis are correct, one can observe a change in the *apparent* catalytic rate, $f_A k_2$, simply by changing the relative populations of the A and B sites. Therefore, our hypothesis for the decrease in catalytic activity upon the D121A mutation is that the A site His 119 conformer is destabilized relative to the B site conformer, which leads to a smaller fraction of the *apo* enzyme being in a catalytically active state at any one time.
**Simulation studies** To test our hypothesis, umbrella sampling is used to calculate potentials of mean force (PMFs) along the His 119 $\chi_1$ dihedral, which defines the A or B site position, for both the D121A mutant and the wild type enzyme. As a check of convergence, two independent sets of PMF simulations are carried out for each system, starting from either a *trans* (A site) or a *gauche* $^+$ (B site) His119 $\chi_1$ dihedral. Both PMF simulations consist of approximately 2-3 ns of production sampling and the results are referred to as the $A \rightarrow B$ PMF and the $B \rightarrow A$ PMF, respectively.

As shown by the $B \rightarrow A$ PMF in Fig.2c, the A site is destabilized by approximately 1 kcal/mol more with respect to the B site in the mutant than in the wild type. This suggests that the B site is more populated in the mutant enzyme than in the wild type. The $A \rightarrow B$ simulations show this same trend, although they produce different relative stabilities of the two sites. For the WT enzyme, for example, whereas the $B \rightarrow A$ PMF predicts that the B site is more stable, the $A \rightarrow B$ PMF predicts that the A site is more stable. In both cases, however, the $\Delta \Delta G_{B-A}$ corresponding to $(\Delta G_{B-A}^{D121A} - \Delta G_{B-A}^{WT})$ is negative, suggesting that the D121A mutation destabilizes the A site with respect to the B site, supporting our hypothesis.

Careful analysis of the simulation trajectories suggests that this hysteresis arises because different His 119 $\chi_2$ angles are sampled in the $B \rightarrow A$ and $A \rightarrow B$ PMF simulations, especially in regions near site B. To test this, a “hybrid” A/B PMF calculation is set up, which uses the A site and transition barrier windows from the $A \rightarrow B$ simulations; for windows near the B site, the starting configurations are generated from an A site equilibrium snapshot, which are then constrained during equilibration to give consistent $\chi_2$’s. As shown in Fig.2d, the A site ($-210 \leq \chi_1 \leq -165$) and transition barrier ($-150 \leq \chi_1 \leq -105$) regions of this hybrid PMF are exactly (within a small constant) the same as in the $A \rightarrow B$ PMF, but with the new B site windows ($-90 \leq \chi_1 \leq -45$), the resulting PMF shows a stabilization of the B site, much like the $B \rightarrow A$ simulations. In fact, the shape of the hybrid PMF in the A site region matches nearly perfectly with the $B \rightarrow A$ PMF.

In short, although there remain important hysteresis problems for the computed $\chi_1$ PMFs due to the lack of sufficient sampling in $\chi_2$ (which illustrates the subtlety of such simulations for even a *local* event!), the hypothesis that the D121A mutant’s loss of activity came from a shift in population of the His119 from the A site to the B site is feasible. In both the $A \rightarrow B$ and $B \rightarrow A$ simulations, a *destabilization* of the A site relative to the B site upon mutation
was observed. That is, although $\Delta G_{B-A}^{D121A}$ and $\Delta G_{B-A}^{WT}$ change signs between the $A \rightarrow B$ and $B \rightarrow A$ simulations, the $\Delta \Delta G_{B-A}$ stays negative for both cases. The more rigorous way to characterize the relevant energetics associated with the His119 isomerization is to compute a two-dimensional PMF along both the $\chi_2$ and $\chi_1$ coordinates; this is underway.

B. Activation of a signaling protein: CheY

Signaling proteins are activated to perform their biological function through a localized event such as phosphorylation or ligand (ion) binding. Understanding how such local modifications lead to striking transitions in the structure and therefore activity of signaling proteins is evidently of great value from both fundamental and biomedical perspectives. Recent NMR studies of small signaling proteins in two-component systems suggested that the structural motifs to be activated have a small but non-negligible population in the active conformation prior to phosphorylation; the role of phosphorylation is to shift this population to become the dominant one rather than inducing new conformations. Such a “population shift” framework, which has features of the Monod-Wyman-Changeux (MWC) model for allosteric, emphasizes the dynamical nature of signaling proteins (and allosteric systems in general) and provides a rather different picture from the “push and pull” type of description as characterized by the stereochemical model for hemoglobin. To fully understand the activation mechanism, however, it is important to characterize the energetics of the relevant motion and reveal how the energetics are modulated by the activation event (i.e., how “population shift” is induced).

“Y-T” coupling vs. population shift CheY is a 129 residue prototypical response regulator in a two-component signal transduction system. It is activated through phosphorylation and the most important conformational change in CheY upon activation is the rotation of the Tyr106 sidechain from a solvent exposed orientation to a fully buried state (Fig.3a). The distance between Tyr 106 and the phosphorylation site (Asp 57) is more than 9.5 Å, which makes CheY a prototypical single-domain protein that exhibits allosteric behavior. The highly conserved Thr 87 spatially separates Asp 57 and Tyr 106, thus the conventional description for CheY activation is the “Y-T coupling” model: phosphorylation of Asp57 displaces Thr87 due to a hydrogen-bonding interaction, which in turn allows the rotation of Tyr106. Since partial activity has been observed for the wild type CheY
and the T87A mutant in the absence and presence of phosphorylation, respectively, the “Y-T coupling” model has been questioned. In particular, since the β4-α4 loop (Ala 88 to Lys 91) also undergoes a major displacement upon activation (root-mean-square-deviation for the backbone and all non-hydrogen atoms is 1.9 and 3.6 Å, respectively), it has been speculated that this loop in fact gates the rotation of Tyr106 and the role of phosphorylation and Thr87 is to select specific loop configuration, which is reminiscent of the “population shift” model.

**Simulation studies** To gain further mechanistic insights into the activation of CheY as an example of monomeric protein allostery, extensive molecular dynamics and PMF simulations were used to explore the coupling between various conformational transitions (e.g., the β4 – α4 loop transition, Tyr 106 rotation, Thr 87 displacement) and phosphorylation in both the wild type CheY and the T87A mutant. In particular, using transition path sampling (TPS), it was shown through ~ 160 natural reactive trajectories (Fig.3b) that the isomerization of Tyr 106 does not require the displacement of Thr 87 and that the hydrogen bond between Thr 87 and Asp 57 phosphate, an essential element of the “Y-T” scheme, is not formed. Recognizing the local nature of TPS simulations, extensive 2-dimensional PMF simulations were also carried out to explore the energetic coupling between key degrees of freedom; each 2-dimensional projection in Fig.3c is generated using between 100-200 ns of simulations. The results showed that the isomerization of Tyr 106 and formation of the Thr 87-phosphate hydrogen bond have similar barriers and are thermodynamically coupled; i.e., kinetically, either event can occur first and facilitate the other. The PMF results also showed that the β4 – α4 loop transition has substantially higher barriers and therefore is unlikely to gate the Tyr 106 rotation; rather, the rotation of Tyr 106 stabilizes the active configuration of this loop, which is consistent with a statistical analysis of all CheY structures in the PDB. Thus, the CheY simulations show that a structural transition at the response site (Tyr 106 isomerization) can occur prior to the so-called activation event (Thr 87-phosphate hydrogen-bond formation). This suggests that the Tyr orientations are in equilibrium and that the active conformation is stabilized by Thr 87-phosphate hydrogen bond formations; kinetically, either event can occur first. In the NMR study of the closely related NtrC, motion associated with the equivalent Tyr was observed to persist in both the unphosphorylated and phosphorylated forms, which led the interpretation that Tyr rotation is “uncoupled” from phosphorylation. Combined TPS/PMF analyses of CheY support
that the rotation of Tyr may occur in the absence of phosphorylation, but it is coupled thermodynamically with phosphorylation.

C. Functional motions in molecular motors at multiple scales

Molecular motors are fascinating systems that convert the chemical free energy in the form of ATP binding/hydrolysis into mechanical work with high efficiency.\textsuperscript{81,82} These systems are rich in motions/reactions that span multiple scales ranging from Angstrom-level changes in ATP-H\textsubscript{2}O during hydrolysis, through local structural rearrangements in the nucleotide binding site, to domain-scale conformational transitions associated with displacement of the motor. Understanding the “mechanoochemical” coupling in motors clearly requires characterizing these motions individually and revealing how they are coupled.\textsuperscript{5,83}

Mechanochemical coupling in myosin  The specific system we focus on is the conventional myosin (referred to as myosin below), which is involved in muscle contraction.\textsuperscript{84} It is one of the few motor systems for which high-resolution x-ray structures (for the motor domain) have been solved for multiple functional states.\textsuperscript{85} The two x-ray structures of interest here\textsuperscript{86,87} are believed to correspond to the post-rigor and pre-powerstroke states in the kinetic scheme,\textsuperscript{84} and the transition from the former to the latter is referred to as the “recovery stroke”. In these two states, the motor domain is detached from the actin and ATP hydrolysis is believed to occur only in the pre-powerstroke state. Comparison of the two x-ray structures reveals structural transitions at different scales, and the most notable ones are (Fig.4a): (i). the C-terminal converter subdomain undergoes a ~60 degrees rotation, which corresponds to a RMSD of more than 20 Å; (ii). the active-site undergoes an open/close transition with a small RMSD of 3 Å; (iii). the relay helix, which connects the active site and the converter, undergoes a significant kink. The fundamental challenge is to understand how these motions are coupled and their relationships to the nucleotide state (ATP-H\textsubscript{2}O vs. ADP-P\textsubscript{i}) in the active site.

Simulation studies  To meet this challenge, a multitude of computational methods have been combined synergistically in our study.\textsuperscript{13,88–91} The general strategy is to characterize the energetics of local events such as ATP hydrolysis\textsuperscript{13} and open/close transition of the active site\textsuperscript{89} in different x-ray structures; the results provide important information about how these local motions/reactions are coupled to structural changes elsewhere in the motor
domain. In addition, approximate transition path calculations, normal mode analyses and an informatics based approach are combined to identify residues/interactions that play an important role in the recovery stroke. Due to the limited space, we restrict ourselves to discussions on the structural transitions. Regarding ATP hydrolysis, it suffices to say that QM/MM calculations of the hydrolysis energetics with different active-site structures clearly showed that the activity relies on the complete closure of the active site, which in turn is coupled to the converter rotation through the relay helix (see below); as a result, the hydrolysis of ATP in the active site is tightly coupled to the converter rotation, despite the separation of more than 40 Å.

First, normal mode calculations found that with either of the two x-ray structures, a small number of low-frequency normal modes sums up to a large fraction of the Cartesian displacements corresponding to the recovery stroke. This is shown more quantitatively with two commonly used descriptors, the involvement coefficients \( I_k \) and the cumulative involvement coefficients \( CI_n \),

\[
I_k = \frac{X_1 - X_2}{|X_1 - X_2|} \cdot L_k
\]

\[
CI_n = \sum_{k=1}^{n} I_k^2
\]

where \( X_1 - X_2 \) is the displacement vector between two conformations \( (X_1, X_2) \) and \( L_k \) is the \( k \)-th eigenvector. As shown in Fig.4b, using less than 20 lowest-frequency modes, more than 50% of the displacement can be accounted for, indicating that the motor domain has inherent flexibility in the specific direction of the recovery stroke. Comparatively, with the same number of modes, the \( CI_n \) for the pre-powerstroke state is notably higher than that for the post-rigor state (Fig.4b), which can be interpreted to suggest that the former is more flexible in the direction of the functional transition. This in fact is consistent with the FRET study of Spudich et al., who found that the orientation of the converter (lever arm) is relatively rigid in the post-rigor state but spans a broader range of angles in the pre-powerstroke state.

To characterize the open/close transition of the active site, PMF calculations were carried out with both x-ray structures. The differential RMSD with respect to the open and closed configurations of the Switch I and II motifs is used as the reaction coordinate. As shown in Fig.4c, the results are strikingly different in the two x-ray structures. In the post-rigor state, the PMF profile is very flat, suggesting that the open and closed configurations have similar
energetics and the transition between them is a low barrier process. In the pre-powerstroke structure, by contrast, the PMF is strongly titled toward the closed configuration while the open configuration is at least 8 kcal/mol higher in free energy. Therefore, the PMFs quantitatively showed that rotation of the converter causes structural changes that propagate to the neighborhood of the active site such that the relative stability of the open/close configurations is strongly perturbed. In fact, data from the PMF simulations can also be used to construct the \((\phi, \psi)\)-free energy profile of residues near the active site. The results (not shown here)\(^{89}\) clearly indicate that the motion of these residues becomes substantially restricted in the post-rigor state. As mentioned above, since the ATP hydrolysis activity is very sensitive to the active site configuration (including the position of water molecules), this tight coordination between converter orientation and active site stability ensures that the later is also tightly coupled to ATP hydrolysis (i.e., “mechanochemical coupling”\(^{1}\)).

Finally, to further explore residues/interactions that dictate the coupling between converter rotation and active site closure, the approximate transition path for the recovery stroke was studied using targeted molecular dynamics simulations,\(^{26}\) as an alternative to minimum energy path analysis.\(^{93}\) The main goal is to observe the formation of transient interactions that are not present in either end-states and therefore difficult to identify using the static x-ray structures. Analysis of the results\(^{90}\) indicates that different types of interactions (polar vs. hydrophobic) along the relay helix play an important role during the recovery stroke. Around halfway in the relay helix, the hydrophobic cluster provides stabilization to the kink of the relay helix, while at the joint between the relay helix and the relay loop region, strong polar interactions facilitate co-operative changes in the relay helix, the SH1 helix and the converter domain. In addition to those local interactions, hinge residues in the low-frequency modes with large \(I_k\) values were also analyzed; the idea is that disruption of these hinges may perturb the flexibility of the system in important directions thus the hinge residues should be of functional significance.\(^{88}\) Among all the hinges identified, a small but significant fraction is highly conserved (> 80% across all species), which supports their functional importance. More interestingly, among the fifty-two residues identified as “strongly coupled (co-evolved)” by the Statistical Coupling Analysis (SCA),\(^{94}\) most are either a hinge residue or involved in an important interaction in the TMD simulations. This is a significant finding because the SCA algorithm works with sequence information only, thus the identified residues are not guaranteed to involve in allostery and might instead
play a role in, for example, co-operative folding. This observation highlights the value of combining an informatics based approach with physically motivated analyses for identifying key residues that dictate functional motions.

D. Mechanical response of a mechanosensitive channel

As the last example, we illustrate how continuum mechanics models, even with a simple parameterization at this stage, can offer unique insights regarding functional motions triggered by external mechanical perturbations.

**Gating transition of MscL** The specific system is the mechanosensitive channel of the large conductance (MscL) in *E. Coli.*, which acts as the “safety valve” for the bacterium by opening up when osmotic pressure is above a certain threshold. MscL is one of the first examples that illustrated that mechanical sensing can occur without the involvement of the cytoskeleton. It is now commonly accepted that the sensing process occurs through the mechanical deformation of the lipid membrane and its interaction with the embedded protein although a complete understanding of the gating mechanism is not yet available. For example, although protein-lipid mismatch has been shown to be important in the gating process, additional force is required to fully open the channel. Moreover, the cytoplasmic S3 helix-bundle was thought to play an important role in the gating process in the first version of structural models. More recently, however, it was argued that the structural changes in the S3 bundle should be substantially smaller. Since the gating process occurs on the millisecond time scale, it is difficult to simulate the transition using atomistic molecular dynamics. For example, even with a steered molecular dynamics approach, the pore radius reached only 9.4 Å after more than 10 ns, which is significantly smaller than the experimentally estimated radius (∼19 Å) for the fully opened state. Approximate open-close transition trajectory can be obtained with targeted molecular dynamics simulations which, however, requires the detailed knowledge of both the closed and open states. In addition, in these biased all-atom simulations the pulling force on the protein is large in magnitude and artificial since the simulations were short and lipid bilayer membrane was completely ignored in the TMD simulations.

**Simulation studies** Motivated by the x-ray structures of MscL from *Tb.* and the homology model of the *E. Coli.* system, we established a simple continuum model for the
**E. Coli.** MscL.\textsuperscript{55} As shown in Fig.5, the model contains all the essential structural motifs including the transmembrane (TM1,TM2)/cytoplasmic (S1-S3) helices and periplasmic loops; in an earlier “minimalist” model,\textsuperscript{53} only the TM1 and TM2 were included since it was speculated that due to their extensive interactions with the lipid, the transmembrane helices are the most important components in MscL; the performances of the minimalist and full models of protein are compared below. The helices are treated as homogeneous (i.e., no sequence dependence has been included here) and isotropic rods, while the loops as elastic springs. The Young’s moduli of the rods are taken from the estimate based on all-atom simulations of Sun et al.\textsuperscript{104} using the CHARMM force field, and those for the springs are established by fitting the lowest three normal modes of the isolated springs from the continuum model and from an all-atom CHARMM calculation.\textsuperscript{105} The materials properties of the elastic membrane are approximated by values in the literature for DPPC.\textsuperscript{106,107} Non-bonded interactions between different continuum components were estimated based on CHARMM force field energy calculations and fitted into simple functional forms similar to the Lennard-Jones interactions. The membrane sheet-helix rod interaction was estimated by rotating the corresponding helix in an implicit membrane using the Generalized Born model.\textsuperscript{108} The fitted parameters can be applied to several structural configurations, or different relative orientations of the channel in the close, open, and intermediate states.\textsuperscript{109}

Once the model is parameterized, different mechanical stress can be applied to the membrane and quasi-static structural response of the channel can be solved using the finite element framework. In the published studies so far,\textsuperscript{53,55} further simplifications were made in which the deformation of the lipid hole that contains the channel and structural response of the channel were calculated separately. This simplification is based on the assumption that the deformation of lipid dominates that of the protein, which reduces computational cost and can be easily removed by considering the full coupling between lipid and protein; in fact, such a comparison may yield important insights regarding the dominance of lipid mechanics during gating (Tang et al., work in progress). Due to limited space, we do not discuss the quantitative aspects (e.g., estimate of proper tension, pore evolution profile) of the simulations, which can be found in Refs.\textsuperscript{53,55} We restrict ourselves to two examples that illustrate the unique value of the FEM framework.

With the “minimalist” model, the structural response of MscL to in-plane biaxial stretching and out-of-plane bending was studied.\textsuperscript{53} With the proper magnitude (∼35 MPa) of in-
plane stretching consistent with the experimental value, the channel was indeed observed to make a transition from the closed to the open configuration. The structural evolution from the FEM simulations compare very favorably to the structural model of Guy and co-workers;\textsuperscript{100} this is significant because the continuum mechanics model was established largely based on the closed state only! With the out-of-plane bending, although the tilting angle of the transmembrane helices changed significantly, due to the lack of in-plane membrane deformation, the structural response of the protein model is minimal. This drastic difference clearly illustrates that the structural response of MscL depends sensitively on the form of the mechanical perturbation and the gating transition relies critically on the in-plane tension in the membrane rather than the curvature of the membrane.

As the second example, we turn to recent results with the more complete structural model\textsuperscript{55} at the continuum level and focus on the effect of in-plane stretching of the membrane. Similar to the results for the “minimalist” model, the channel fully opens under the proper magnitude of in-plane tension. At the same membrane strain level, the pore radius of the final state is, however, about 20\% smaller than that in the “minimalist” model, which indicates that the presence of additional structural motifs other than the transmembrane helices (most likely the periplasmic loops, see below) tend to reduce pore opening. Nevertheless, the relatively small difference between results from the complete and the “minimalist” model indicates that the gating process is dominated by the iris-like expansion of transmembrane helix bundles. Interestingly, the structural variation of the S3 helical bundle during the gating transition is very small, which supports the recent modification of the structural model.\textsuperscript{101} The cytoplasmic S1 helical bundle, on the other hand, moves into the transmembrane region and opens up; the periplasmic loops also closely follow the trajectory of the transmembrane helices.

Another type of interesting study is to remove specific structural component and observe the effect on the gating transition; this is an unique aspect of computational analysis since the corresponding exercise with experiments will be complicated by factors such as major structural distortions prior to channel activation. Here we have tested the role of three structural motifs: the S3 helical bundle, the periplasmic loops and the cytoplasmic loops that connect S1 and TM1 helices. As expected based on the above discussion, removing S3 helices did not cause much change in the gating behavior, once again confirming the insignificant role of S3 for the opening of MscL. Removing the periplasmic loops causes
major variations in the configurations (e.g., tilting angle) of the transmembrane helices and therefore the final pore size; in addition, removing the cytoplasmic loops makes the S1 bundle distorted, thus also affecting the pore radius. Both observations regarding the importance of these loops are consistent with the recent experimental observation of Sukharev and co-workers.98,101

IV. CONCLUDING DISCUSSIONS AND FUTURE OUTLOOKS

There is little doubt that biomolecules are flexible objects and rich in motions of different temporal and spatial scales. Characterizing the nature of these motions and how they are perturbed by changes in the environment (e.g., osmotic stress) or ligation state is a fundamental challenge in structural biology and biophysics. It is even more challenging, however, to identify *functional motions* that in fact play a major role in facilitating the function of biomolecules, which can be striking domain-scale rearrangements that “propel” a molecular motor forward or subtle local changes that set up the proper active site or interface for the subsequent catalysis or binding.

Through the examples in this chapter, we hope to illustrate that modern computational approaches are making rapid advances so that processes at multiple scales can be investigated. As a result, computational analysis can play a major role in the study of functional motions, in terms of both helping better interpret experimental data and stimulating new hypotheses regarding the nature of such motions and mechanisms by which they are regulated. For example, our study of RNase A helped establish a concrete hypothesis regarding how the perturbation of a catalytic residue’s motion may lead to a significant decrease in the observed catalytic rate for the D121A mutant. This hypothesis, which is more molecular in nature than the original proposal58 that underlines the importance of “coordinated” and “time” global dynamics (however, see discussions below regarding current challenges), should be tested by further mutation studies.

In most cases, the functional motion implicates multiple structural changes and therefore an important issue is to understand how these structural changes are coupled and whether there is distinct “causality” (or sequence of event) between them. In CheY, for example, a key question is whether the Tyr rotation is dependent on the hydrogen-bond formation between Thr87-phosphate. In the molecular motor myosin, a fundamental question regards
whether ATP hydrolysis in the active site triggers structural transitions that eventually propagate to the converter or the converter rotation occurs first,\textsuperscript{109} which then leads to changes in the nucleotide binding site that activate the ATP hydrolysis. Since it remains difficult to directly observe those events in real-time, either computationally or experimentally, the best approach is to characterize the \textit{energetic coupling} between different processes, which can be achieved with careful potential of mean force (PMF) computations.

In CheY, extensive multi-dimensional PMF simulations\textsuperscript{78} revealed the energetics of different motions and therefore how they are coupled. The results support the idea that the isomerization of a key Tyr residue can occur prior to the activation event (hydrogen-bonding formation between Thr87-phosphate) with a modest barrier; unlike the original experimental interpretation of the similar NtrC,\textsuperscript{64} however, the calculations clearly indicate that the Tyr rotation and the activation event are coupled. In myosin, PMF calculations\textsuperscript{89} for the active-site open/close transition with different x-ray structures also convincingly showed how converter rotation propagates to structural changes near the active site such that the open/close energetics get affected significantly. As a result, the ATP hydrolysis activity is tightly coupled to the converter rotation, despite the large separation of more than 40 Å.

Due to the complex nature of functional motions, it is productive to combine multiple computational techniques, similar to the use of multiple approaches in an experimental investigation. In the study of CheY activation,\textsuperscript{78} the collection of hundreds of reactive trajectories from TPS simulations are instructive but not conclusive due to the local nature of the employed TPS algorithm. However, these \textit{natural} reactive trajectories played a major role in identifying the proper coordinates for the subsequent multi-dimensional PMF simulations. In the study of myosin,\textsuperscript{90} the recovery stroke is a highly complex process that involves both domain-scale motions and extensive rearrangements at the loop or sidechain levels. To identify residues that play a key role in the process, combining normal mode based hinge analysis, targeted molecular dynamics and the statistical coupling analysis was productive because these techniques are based on different fundamental assumptions and therefore complement each other well.

Another form of combining different methods is propagating information from all-atom simulations to an effective coarse-grained model, which is illustrated here with the continuum mechanics model of MscL.\textsuperscript{53,55} Even with rather simple proof-of-concept type of models, new insights have been obtained regarding the impact of different forms of mechanical per-
turbation on the gating transition of MscL and the role of various structural motifs in the process. Although still in its infancy, if done carefully and properly, this type of strategy can be very powerful for analyzing the functional motions of very large biomolecular complexes at very large length (even cellular) scales.

**Outstanding and emerging challenges** Although it is always presumptuous to speculate too much into the future, we briefly ponder on several subjects for which we would like to see further studies. Instead of discussing from the perspective of technical developments, which clearly will continue on multiple fronts and at multiple scales, we point out a number of questions regarding “functional motions” that the authors believe are particularly interesting to explore.

- **What are the roles of “slow (\(\mu s - ms\)) motions” in enzyme catalysis?** A significant body of computational studies has been focused on analyzing the impact of motion on enzyme catalysis. However, essentially all calculations focused on relatively fast motions on the order of pico- to nano-seconds, due to either limits in the computational resources or the fact that the goal was to study the impact of enzyme motions on the barrier crossing process, which does occur at the picosecond time scale for most chemical reactions. Therefore, an important issue that has not been extensively analyzed at the molecular level concerns the possibility that slow (\(\mu s - ms\)) motions may significantly modulate the enzyme (active-site) structure so that a significant number of chemical turnovers, in fact, occurs in “excited state” conformation(s). In this regard, we note that we do not consider “simple” cases where a conformational change (e.g., closure of the active-site upon substrate binding) is a kinetically distinct step prior to catalysis;\(^1\(^{10}\) rather, we focus on systems where the most catalytically active conformation is rarely populated and distinct from the most stable Michaelis-complex as observed by, for example, crystallography. Another way of stating the issue is that the most populated conformation observed under a specific experimental condition may not be the most functionally active one. Single molecule experiments demonstrated that the “rate-constant” (or apparent barrier) of an enzyme (e.g., cholesterol oxidase) catalyzed reaction is in fact time-dependent,\(^1\(^{11}\) presumably due to structural transitions between different conformational substates; the molecular nature of such transition has not been illustrated and the magnitude of the apparent barrier fluctuation is usually small. The existence of important “global” \(\mu s - ms\) motion has been suggested in several enzyme systems based on NMR relaxation measurements,\(^1\(^{11}\) including the RNase system
discussed earlier.\textsuperscript{57,58} Although alternative explanations \textit{may} exist, a conclusive analysis of
the nature of “functional motions” is not complete unless motions on the $\mu$s – $ms$ time-scale
are accessed and their impact on the catalysis analyzed at the atomic level. This is clearly
a challenging task for computations because a meaningful description of such long-time motions
requires both extensive sampling and reliable force fields. Before \textit{tour de force} analyses
can be done, it is likely that combining QM/MM analysis and enhanced sampling techniques
such as “conformational flooding”\textsuperscript{113} may produce instructive insights.

\textbf{What are the bottlenecks for large-scale functional motions?} To completely
understand functional motion, it is important to identify the kinetic bottleneck of the process
among all the implicated structural transitions. The above discussion of myosin made it
clear that functional motions likely involve both domain-scale changes \textit{and} important local
structural transitions. Domain motions are more striking in scale while the local transitions
more subtle, but the spatial magnitude of changes does \textit{not} necessarily correlate with kinetic
significance. As noted above, many studies found that large-scale structural transitions
are correlated with low-frequency modes, which implies that biomolecules tend to have
intrinsic structural flexibilities that ensure domain-scale motions to be largely diffusive in
nature; therefore, the kinetic bottleneck of a functional transition may, in fact, consist of
key local structural changes that are thermally activated. Such considerations highlight the
importance of revealing the free energy landscape of functional motions, for which detailed
computational analysis beyond a simple harmonic picture is indispensable.

The realization that local structural changes may constitute the kinetic bottleneck of com-
plex structural transitions has important implications regarding strategies for constructing
meaningful coarse-grained models in the context of studying functional motions. For in-
stance, although it seems sensible to coarse-grain biomolecules into rigid domains, the pre-
dictive power of such models might be significantly compromised if important local features
(e.g., repacking of hydrophobic sidechains in the rely loop/helix in myosin) are ignored. In
this regard, an important emerging challenge is to make quantitative connections, at multiple
resolutions and scales, between computational models and experiments that report on the
time and spatial scales of biomolecular motions; most notable examples include small-angle
x-ray scattering,\textsuperscript{114,115} diffuse x-ray scattering,\textsuperscript{116} fluorescence resonance transfer (FRET),\textsuperscript{117}
electron spin resonance\textsuperscript{118} and 2-dimensional infrared spectroscopy,\textsuperscript{119} which span a broad
range of time resolutions and spatial scales. $\Phi$ analysis, which is commonly used in protein
folding analysis\textsuperscript{120} and recently applied to study motions in the acetylcholine receptor,\textsuperscript{121} also provides extremely valuable data regarding whether specific residues are involved in the transition state ensemble of the transition. Making explicit comparison to experiments provides not only important validations for the computational model but also the opportunity of gleaning additional information from experimental data. Recent applications of elastic network models in the refinement of x-ray structures, EM structures, electron tomography and FRET data are good examples.\textsuperscript{32,33,122}

\textit{Can functional motions be modulated in a predictive manner?} Although significant motions are implicated in the functional cycle of biomolecules, protein engineering studies have largely been guided by static structural considerations, which reflects our lack of a thorough understanding of factors that dictate the features of functional motions. Although the situation will improve steadily, the most productive avenue for incorporating molecular motions into protein design in the near future likely involves combining clever genetic approaches, molecular simulations and informatics motivated models.

In an impressive recent study,\textsuperscript{123} for example, a novel gene synthesis approach was used to construct chimeras between the mesophilic and thermophilic adenylate kinases, in which different domains from the two enzymes are combined randomly (eight were considered). Measurement of thermostability and enzyme activity (which is limited by a structural transition that implicate the active site closure) revealed that it is possible to enhance the flexibility of key domains without affecting the thermostability. Further studies of this sort, supplemented by simulation and informatics analysis,\textsuperscript{124} may lead to new avenues of manipulating protein functions through rationally modulating essential motions. In addition to enzymes, interesting targets are molecular motors and other allosteric systems, in which specific mutations are known to disrupt functional motions such that communication between different sites is abolished.\textsuperscript{125–127} Since these mutations often lead to serious diseases, devising effective methods for restoring key functional motions has great biomedical implications.

\textit{Are there major differences between “functional motions” in vitro and those in vivo?} Finally, a recent trend in biophysical studies is to contrast molecular behaviors under \textit{in vitro} and \textit{in vivo} conditions. It has been recognized for some time that the cellular environment is far from the dilute solution condition in \textit{in vitro} experiments; molecular crowding, non-specific binding and other features associated with non-ideality may signifi-
cantly affect the structure, stability and association of biomolecules in cells.\textsuperscript{128–130} Functional motions of biomolecules, especially those large-scale motions that strongly implicate nearby water/solute molecules, might be substantially different \textit{in vivo} compared to \textit{in vitro}. To what extent this is true and and what are the corresponding functional implications are clearly very interesting questions that deserve careful analysis. Along this line, recent studies of both the protein\textsuperscript{131} and solvent\textsuperscript{132} dynamics in reverse micelles provided interesting clues. Clearly, more studies are needed from both the theoretical/computational and experimental perspectives to fully understand biomolecular motions that are most important under the physiological condition.

\textbf{Acknowledgments}

Q.C. and X.C. acknowledge other group members and collaborators who made significant contributions to the work discussed here. The studies reviewed have been generously supported from a Research Innovation Award from the Research Corporation and the National Institutes of Health (R01-GM071428-01) to Q.C. and a CAREER award to X.C. (CMMI-0643726) from the National Science Foundation. Q.C. also acknowledges a Alfred P. Sloan Research Fellowship. Computational resources from the National Center for Supercomputing Applications at the University of Illinois are greatly appreciated.

\begin{enumerate}
\end{enumerate}

26


69 A. Szabo and M. Karplus, J Mol Biol 72, 163 (1972).
77 M. S. Formaneck, L. Ma, and Q. Cui, Proteins 63, 846 (2006).
FIG. 1: Illustration of the computational methods discussed in this Chapter. (a) Potential of mean force (PMF) simulations along chosen reaction coordinates ($\xi_1, \xi_2$), from which the kinetics ($k_{R\rightarrow P}$) and thermodynamics ($K_{R\rightarrow P}^{eq}$) of the relevant transitions can be estimated, are most useful for probing local motions; (b) Transition Path Sampling (TPS), which probes reactive (either chemical or conformational) trajectories between different states with proper weights, does not require the pre-selection of reaction coordinates and is suitable for studying complex local structural transitions in systems with rough energy landscape; (c) Normal Mode Analysis (NMA) is an approximate method well suited for describing collective motions at the domain scale; (d) Further coarse-grained methods such as those based on continuum mechanics can, if parameterized carefully, probe functional motions at very large length and time scales. All structural figures are made using VMD.\textsuperscript{133}

FIG. 2: Study of functional motions in RNase A. (a) The basic catalytic cycle; (b) The active site of RNase A with both the His119 A and B site rotamers present; the hydrogen bonding in both the A site (Asp 121) and the B site (Ala 109 main chain and Glu 111) is indicated. (c)-(d) Potential of mean forces from different calculations along the His 119 $\chi_1$ angle that characterizes the transition between the A & B sites.

FIG. 3: Study of functional motions in CheY.\textsuperscript{78} (a) Comparison of the inactive\textsuperscript{75} and active\textsuperscript{76} structures of CheY. Overlay of key residues between the phosphorylation (Asp 57) and response sites (Tyr 106). Residues in the active structure are colored according to atom types, while those in the inactive structure are colored as ice-blue. The inactive and active configurations of the $\beta 4-\alpha 4$ loop are colored as dark-blue and yellow, respectively. (b) Four configurations along an exemplary activation trajectory harvested using TPS; note that the intrinsic time scale of barrier crossing, which is different from the reaction time, is short and on the pico-second scale. Several important residues including Tyr 106, Ala 90, Ile 95 and Val 108 are shown in the van der Waals scheme; the phosphorylated Asp 57 is shown in the licorice form; the $\beta 4-\alpha 4$ loop is shown as the blue ribbon. (c) A three-dimensional scheme that illustrates the energetics and possible pathways for CheY activation based on the computed 2-dimensional PMFs along the key degrees of freedom. The expected fully-active state, $A_f^*$, is not a local free-energy minima in the simulations, presumably due to the absence of the FlhM peptide in the model.
FIG. 4: Study of mechanochemical coupling in myosin.\textsuperscript{89,90} (a) The difference between the post-rigor (1FMW\textsuperscript{86}, in blue) and pre-powerstroke (1VOM\textsuperscript{87}, in light green) states; the structures are aligned based on backbone atoms in the first 650 residues. Also shown is the superposition of the active site, where Mg-ATP is in the van der Waals form, and key loops (P-loop, Switch I/II) as ribbons; the active site is “closed” with the salt-bridge between Arg 238 and Glu 459, and “open” when the salt-bridge is broken; (b) Involvement coefficient ($I_k$) and cumulative involvement coefficient ($CI_k$) from normal mode calculations (Eqs. 8 and 9) for the structural transitions between the two x-ray structures using the modes of either structure; (c) PMF for the open/close transition of the active site with different x-ray structures; (d) Mapping of the 52 strongly coupled core residues (in van der Waals form) to the structure of \textit{Dictyostelium} myosin motor domain. The coupled residues are colored based on residue type; blue: basic, orange: acidic, ice-blue: polar, white: non-polar.

FIG. 5: Study of structural response of MscL from \textit{E. Coli} based on a continuum mechanics model. (a) The basic philosophy of the coarse-graining procedure (replacing transmembrane helices by homogeneous and isotropic elastic rods, bilayer by elastic sheet) and set-up of the finite element simulation;\textsuperscript{53} (b) Comparison of the transmembrane helices in terms of their packing and tilting angles during the gating transition from the structural model of Guy et al.\textsuperscript{100} and from the finite element simulations;\textsuperscript{53} (c) The sideview of the structure at the end of the simulated gating transition using a more complete continuum model at the continuum level,\textsuperscript{55} along with the corresponding structure when the periplasmic loops, the S3 helical bundle and the cytoplasmic loops, respectively, are removed from the model. The dotted lines in (b)-(c) approximately indicate the location of the membrane-water interface.
Coarse Graining

Fig. 1

(a) 

\[ k_{R \rightarrow P}^{TST} = A e^{-\frac{W(TS) - W(R)}{k_B T}} \]

\[ K_{R \rightarrow P}^{eq} = e^{-\frac{W(P) - W(R)}{k_B T}} \]

\[ W(\xi_1, \xi_2) = -k_B T \ln P(\xi_1, \xi_2) + C \]

(b) 

\[ P_{acc}[x^{(o)}(T) \rightarrow x^{(n)}(T)] = h_R[x^{(n)}_0] h_P[x^{(n)}_T] \min \left[ \frac{1}{\rho(x^{(n)}_i)}, \frac{\rho(x^{(o)}_i)}{\rho(x^{(o)}_i)} \right] \]

(c) 

\[ \mathbf{H} \mathbf{L}_i = \omega_i^2 \mathbf{L}_i \]

(d) 

\[ \int_{\Omega} G_{ij}(\hat{\phi}) d\Omega + \int_{\Gamma} g(\hat{\phi}) d\Gamma = 0 \]
Fig. 2
Asp 57-phosphate

Ala 90

Thr 87

Tyr 106

"out"

"in"

"4"#4 loop

FliM binding

(a)

(b)

(c)

Fig. 3
Relay Helix
Converter Rotation
(a) (b)
(c) (d)
Pre-powerstroke
Post-rigor
Fig. 4
Post-rigor $ Pre-powerstroke
Pre-powerstroke $ Post-rigor
Fig. 5

(a) (b) (c)

Periplasmic loops w/o Periplasmic loops w/o S3 helical bundle w/o Cytoplasmic loops

TM1 TM2

Coarse graining

E. Coli

Periplasmic loops

S1

S3