Biophysical Perspective

Structure-Encoded Global Motions and Their Role in Mediating Protein-Substrate Interactions

Ivet Bahar,[1,][*] Mary Hongying Cheng,[1] Ji Young Lee,[1] Cihan Kaya,[1] and She Zhang[1]
[1]Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

ABSTRACT   Recent structure-based computational studies suggest that, in contrast to the classical description of equilibrium fluctuations as wigglings and jigglings, proteins have access to well-defined spectra of collective motions, called intrinsic dynamics, encoded by their structure under native state conditions. In particular, the global modes of motions (at the low frequency end of the spectrum) are shown by multiple studies to be highly robust to minor differences in the structure or to detailed interactions at the atomic level. These modes, encoded by the overall fold, usually define the mechanisms of interactions with substrates. They can be estimated by low-resolution models such as the elastic network models (ENMs) exclusively based on interresidue contact topology. The ability of ENMs to efficiently assess the global motions intrinsically favored by the overall fold as well as the relevance of these predictions to the dominant changes in structure experimentally observed for a given protein in the presence of different substrates suggest that the intrinsic dynamics plays a role in mediating protein-substrate interactions. These observations underscore the functional significance of structure-encoded dynamics, or the importance of the predisposition to favor functional global modes in the evolutionary selection of structures.

Proteins perform their function via chemical and physical changes. Chemical changes include catalysis, posttranslational modification, cross-link formation, and covalent binding/unbinding of ligands. Physical changes involve domain rearrangements, allosteric changes in conformations (intramolecular), and protein-ligand interactions, multimerization, and formation of complexes and assemblies (intermolecular). The combination of chemical and physical (or mechanical) properties leads to unique mechanochemical behavior essential to biological activity (1). Some activities require high precision (e.g., positioning of catalytic residues in enzymes), whereas others rely on modular structures (or substructures) that are adaptable to different functionalities, e.g., binding different substrates by ubiquitin, or different antigens by antibodies. Conformational flexibility accompanied by sequence variations ensures in the latter case the adaptation to binding different substrates, and to mediate substrate specificity (2,3). Conformational flexibility that is highly specific in its directionality is essential for allosteric responses, too. It is, usually, via a conformational switch triggered by a first ligand binding (to an allosteric site) that another (recruitment) site reconfigures to facilitate its recognition by another ligand (4,5). How does the protein strike the right balance between rigidity and flexibility? How does it elicit the specific type of conformational switch that leads the way to its biological function? How does it ensure that those functional responses are robustly maintained?

Studies attempting to bridge between structure and function have led to the prominent view that the link between structure and function is through dynamics (6). Dynamics refers to changes in the chemical or physical state (reaction dynamics or interaction dynamics, respectively). The perspective of our study focuses on physical changes. We consider the changes in conformation because of substrate binding or allosteric interactions under physiological conditions—the conditions under which biomolecules accomplish their biological functions. We will submit the views that 1) each protein fold has a unique dynamics, called intrinsic dynamics, encoded by its structure under equilibrium conditions; 2) current structure-based computational approaches can provide a good first glimpse of proteins' intrinsic dynamics at the microscopic scale, broadly consistent with experimental data, despite inherent limitations and simplifications in the models and methods; and 3) predicted dynamics has functional significance, and could be exploited for design, engineering, and therapy purposes (e.g., for exploring the druggability of target proteins). We conclude by inviting attention to the significance of intrinsic dynamic propensities in protein design and evolution; and we discuss recent studies suggesting that structure-encoded dynamics is evolutionary optimized and regulated by robust mechanisms.

Each protein has a unique intrinsic dynamics, which may be quantitatively explored by structure-based computational models and methods

In principle, the dynamics of each protein is governed by the forces experienced by its atoms as a result of intramolecular

*Correspondence: bahar@pitt.edu

Editor: H. Jane Dyson

CrossMark

and intermolecular interactions; and to the extent that the force field provides an adequate description of these interaction potentials, the way the protein samples the conformational space in molecular simulations is a deterministic process (assuming Newtonian dynamics): it is fully determined by the atomic positions, or the instantaneous structure of the protein in the examined force field. Although the highly nonlinear interactions and in particular the aqueous environment introduce an apparent stochasticity, a given protein in a given environment has in principle its own dynamic character uniquely defined by the spatial distribution of its atoms.

One method that found wide applications in characterizing the unique dynamics of each protein near its native state conditions is normal mode analysis (NMA) (7). NMA applies to the close neighborhood of the native state (energy minimum). It yields the spectrum of modes intrinsically accessible under equilibrium conditions. These vary over a broad range, from global (modes that cooperatively involve significant portions of the structure, if not the entire structure, usually at the low-frequency end of the spectrum) to local (at the high-frequency end). Each structure has a unique spectrum of modes, also called collective motions. Note that in a strict sense, the normal modes hold in the infinitesimal proximity of the global energy minimum, and as such, they provide insights into the intrinsic tendency or predisposition of the protein to undergo particular changes in structure near its native state.

An important feature of collective motions extracted by NMA is the robustness of the computed global modes. It has been widely established and confirmed by many studies that global modes are insensitive to precise atomic coordinates or detailed force field parameters, but they are robustly defined by the overall fold/architecture of the biomolecular system, or by the topology of interresidue contacts. This fundamental concept, first shown in the pioneering coarse-grained NMA of Tirion (8), prompted the introduction of elastic network models (ENMs) for delineating the equilibrium dynamics of biomolecular systems, starting from the Gaussian Network Model (GNM) (9,10), followed by coarse-grained NMA with ENM variants by Hinsen, Perahia, and coworkers (11–13). A natural extension broadly used has been the anisotropic network model (ANM) (14). The main advantage of ANM analysis, like other ENM-NMA, is the ability to efficiently yield a unique solution for the collective motions of the examined structure without the need to perform any simulations or energy minimization. This is achieved by eigenvalue decomposition of a simple, closed-form expression for the Hessian—a $3N \times 3N$ matrix composed of $N \times N$ submatrices of the following form (14):

$$\boldsymbol{H_{ij}} = -\frac{\gamma_{ij}}{R_{ij}^2} \begin{bmatrix} x_{ij}^2 & x_{ij}y_{ij} & x_{ij}z_{ij} \\ x_{ij}y_{ij} & y_{ij}^2 & y_{ij}z_{ij} \\ x_{ij}z_{ij} & y_{ij}z_{ij} & z_{ij}^2 \end{bmatrix},$$

for $i \neq j$, where x_{ij}, y_{ij}, z_{ij} are the x-, y-, and z-components of the distance vector $\boldsymbol{R_{ij}}$ between nodes i and j, directly taken from the known (e.g., crystal) structure; the spring constant $\gamma_{ij} = \gamma$, if the distance R_{ij} between nodes i and j is shorter than a cutoff distance r_c, and zero otherwise; and $\boldsymbol{H_{ii}} = \Sigma_j \boldsymbol{H_{ij}}$, where the summation is over off diagonal elements. There have been various formalisms in the literature for choosing the spring constants, ranging from this simplest formulation to detailed approaches customized per protein (15). In particular, the use of a distance-dependent expression for γ that helps eliminate the parameter r_c has been broadly adopted (11,16). Most of these approaches usually emphasize the need to use stiffer springs for covalently bonded pairs of residues. However, the results from ENMs, and in particular the predicted global modes, have been consistently shown to be robust to changes in spring function and parameters (see, for example, (17)).

As a corollary to the deterministic nature of molecular motions and interactions, molecular dynamics (MD) simulations would, in principle, provide us with an accurate description of biomolecular systems dynamics near their equilibrium (native) coordinates—if the generated trajectories are long enough to thoroughly sample the conformational space. Unfortunately, this is not the case, even with the most advanced simulation hardware and software. Remarkable successes have been made in developing MD methodologies applicable to systems of the order of 10^8 atoms (18,19), or conducting millisecond-range simulations (20). Yet, the former is restricted to the timescale of nanoseconds, and the latter, to small proteins only. However, MD convergence studies clearly show that it is impossible to obtain statistically significant data from full atomic MD simulations on the time evolution of many biological processes of interest (21,22). In contrast, ENM offers significant advantage for its computational efficiency and easy usage with orders of magnitude less computational expensive.

ENMs have their own limitations. They are coarse-grained—they usually describe the protein at the level of one-node-per-residue, each residue being identified by the position of its C^α-atom. They do not take account of specific interactions: all interresidue interactions are represented by a uniform harmonic potential between residue pairs connected in the network (which are within a cutoff separation). The network connectivity, or the spatial distribution of the nodes and springs, fully defines the dynamics of the structure represented as an ENM. Third, ENMs do not (usually) take account of the solvent or lipid environment. Finally, by definition, they do not incorporate nonlinear effects nor couplings between modes. The ENM formalism assumes fluctuations about a single well. As such, it may not be suitable for modeling the conformational dynamics of flexible objects such as loops that have multiple populated states.

Despite all limitations and simplifications, structure-based computations can provide a mechanistic description of the unique dynamics, often consistent with experiments

Despite all these limitations, computational models and methods are useful. The question often reduces to deciding which particular properties, or processes, are of interest, and choosing the proper model and methods depending on the specific time and length windows that are being examined. For example, if the goal is to have a detailed atomic understanding of the mechanism of extracellular (EC) gate opening by a given neurotransmitter transporter, one essentially examines by MD simulations the interactions near the binding site, or near the extracellular (EC) vestibule (Fig. 1, A–C) in the presence of the whole structure. Simulations up to 100s of nanoseconds are sufficient to consistently view substrate binding and EC gate closure or opening events in transporters, for example, in the presence of explicit membrane, water molecules, ions, and substrate molecules. In addition to elucidating the behavior of individual proteins, MD data permit us to make inferences on the common mechanisms of function selected by family members that share similar folds despite sequence dissimilarities, or the distinctive mechanisms for binding different ligands. The comparison of the behavior of dopamine transporter

(DAT) and leucine transporter (LeuT), both sharing the LeuT fold, clearly elucidates the equivalent role of particular residues in EC (or intracellular, IC) gating (Fig. 1, A–C).

On the other hand, if one seeks to gain insights into the conformational space accessible on a global scale, ENM-based analyses turn out to be exceptionally powerful. For example, ENM-based computations help visualize the transition between the outward-facing (OF) and inward-facing (IF) states of neurotransmitter transporters (23,28) (Fig. 1, D and E), consistent with the alternating access mechanism (25,29–31). Other examples (from the past 5 years) are the transition between the open or closed conformers of enzymes in the respective ligand-free and ligand-bound forms (32), the reconfiguration between the alternative (e.g., R and T) functional forms of allosteric proteins such as hemoglobin or GroEL (33–35) (Fig. 2), the ATP-coupled conformational mechanics of motor proteins such as myosin (34,36,37), kinesin, G-actin (38), or F1-ATPase (34), the functional mechanisms of transporters (39–41) including the above illustrated OF ↔ IF transition (Figs. 1, C and D, and 3), pore opening or allosteric gating of ion channels (42–47), or the global structural changes or concerted domain rearrangements that are stabilized upon ligand/inhibitor binding by certain proteins (6,32,48–50) (Fig. 4).
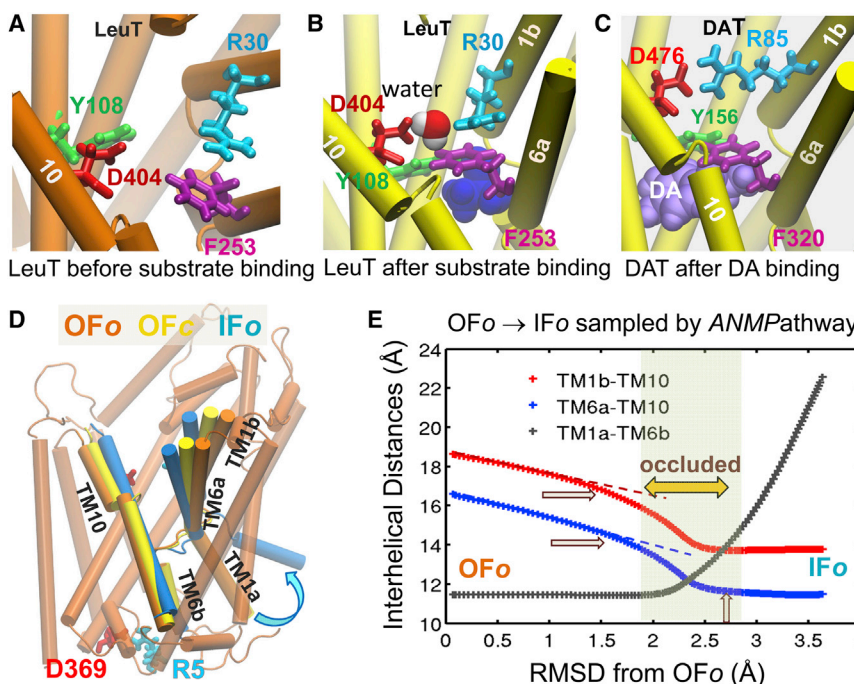

FIGURE 1 Local and global structural changes captured by full-atomic and coarse-grained structure-based computations. (A and B) Closure of the extracellular (EC) gate after substrate binding, observed in MD simulations of leucine transporter (LeuT) (23). LeuT is known to transport alanine (most efficiently), leucine, and other amino acids. Panel A displays its substrate binding site in the outward-facing, open (OFo) crystal structure (24), before substrate binding. The oppositely charged residues R30 and D404, originally far apart (A), closely interact in upon substrate binding (leucine, *blue space filling*) (B), as illustrated for the LeuT in substrate-bound outward-facing closed (OFc) state (25). Likewise, the two aromatic residues Y108-F253 closely associate to form another layer further consolidating the EC gate (23). (C) MD simulations of a LeuT structural homolog, human dopamine transporter (hDAT, modeled after the OFo dDAT structure (26)) show local changes in conformation upon dopamine (DA) (*light violet space filling*) binding (27), which closely resemble those stabilized in substrate-bound LeuT. R85-D476 and Y156-F320 serve as the EC gates in this case. Transmembrane (TM) helical segments TM1a-b, TM6a-b, and TM10 that line the binding cavity, exhibit significant structural rearrangements. (D) Global structural differences between LeuT in OFo state before substrate binding (*dark orange*), OFc state after substrate binding (*yellow*), and inward-facing open (IFo) state (*blue*) after substrate release to the intracellular (IC) medium (24) (E) Global structural changes observed in ENM-based simulations of OFo ↔ IFo transition (28). TM1b-TM10, TM6a-TM10, and TM1a-TM6b center-of-mass (COM) distances serve as metrics for probing the extent of reconfiguration. The passage over an occluded state where the transporter is closed to both EC and IC media, consistent with experimental data (25), is highlighted.
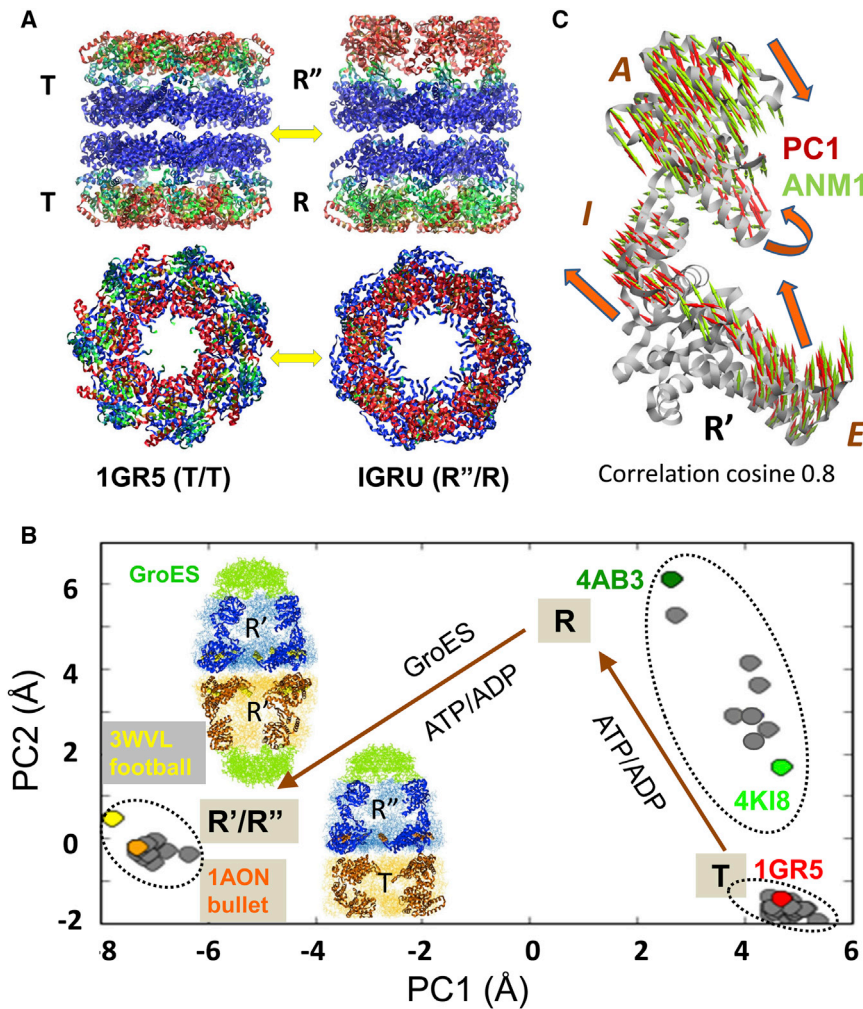
FIGURE 2 Allosteric changes in the bacterial chaperonin complex GroEL/GroES structure are facilitated by intrinsically accessible ANM modes. GroEL consists of two heptameric rings, which assume the states (33,51): T: ATP-free; R: ATP-bound before substrate protein (SP) and co-chaperonin (GroES) binding; R′: ATP-, SP-, and GroES-bound; R″: ADP-, SP-, and GroES-bound. (A) Crystal structures of GroEL in T/T (PDB: 1GR5) and R″/R (PDB: 1GRU) states, side view (*top*) and top view (*bottom*). Residues are color-coded by experimental B-factor, as low (*blue*), intermediate (*green*), and high (*red*). (B) Projection of GroEL monomers on the conformational sub-space spanned by the principal modes PC1 and PC2 derived from the principal component analysis (PCA) of an ensemble of 39 GroEL and GroEL/GroES structures deposited to date in the PDB. Computations are performed using ProDy (52). Examples of resolved structures, including the symmetric football-like GroEL:GroES$_2$:ATP$_{14}$ (53), are shown in inset. (C) Comparison of the directions of PC1 (*red arrows*) deduced from the PCA of the experimental dataset and the first ANM mode (ANM1) (*green arrows*) calculated using a single monomeric structure (in the R′ state). ANM1 and PC1 yield a correlation cosine of 0.8.

Are computationally predicted changes in structure meaningful? Are they functional?

Many studies in the past decade, including those cited above, have drawn attention to the physical and biological significance of computational approaches that utilize ENMs or their variants, either in NMA per se, or in hybrid methods in combination with simulation methods such as MD or Monte Carlo (MC). Structural data on proteins resolved in more than one functional state clearly support the hypothesis that computationally predicted modes of motions strongly resemble to the differences between experimentally resolved structures, and they furthermore support the view that computations based on a single structure may provide an accurate, albeit low-resolution, first description of the conformational space accessible to that structure under native state conditions.

One might argue, however, that ENMs (or NMA) predict a multitude of modes, and it may be hard to assess which ones are functionally relevant; and furthermore, given the fact that the modes represent a complete orthonormal basis set of *3N*-dimensional directional vectors, there may always

be among them one or more that are relevant to the structural difference $d_{3N} = \{R_{S1} - R_{S2}\}_{3N}$ between the selected substates *S1* and *S2* of the same protein stabilized under different conditions. We present two major observations that refute this argument.

First, the collective modes that usually exhibit good agreement with experiments are consistently among the lowest frequency modes (6). They are the softest or the most probable modes from energetic point of view (note that the mode eigenvalue represents the curvature of the quadratic energy change along that mode coordinate, and provides a metric for the particular mode's previous probability). The experimentally observed structural change d_{3N} is usually observed to yield a correlation cosine of ~0.6 or higher with one of the soft modes u_k computed for *S1* or *S2* (e.g., $1 \leq k \leq 10$, with mode index $k = 1$ referring to the softest mode) (6). Furthermore, a small subset ($m \sim$ tens) of softest modes usually generates a cumulative overlap $Co(m) = [\Sigma_k \, cos^2(d_{3N}, u_k)]^{1/2}$ ($1 \leq k \leq m$) of 0.9 or higher. For example, the change d_{3N} between the unbound and inhibitor-bound structures of HIV-1 reverse
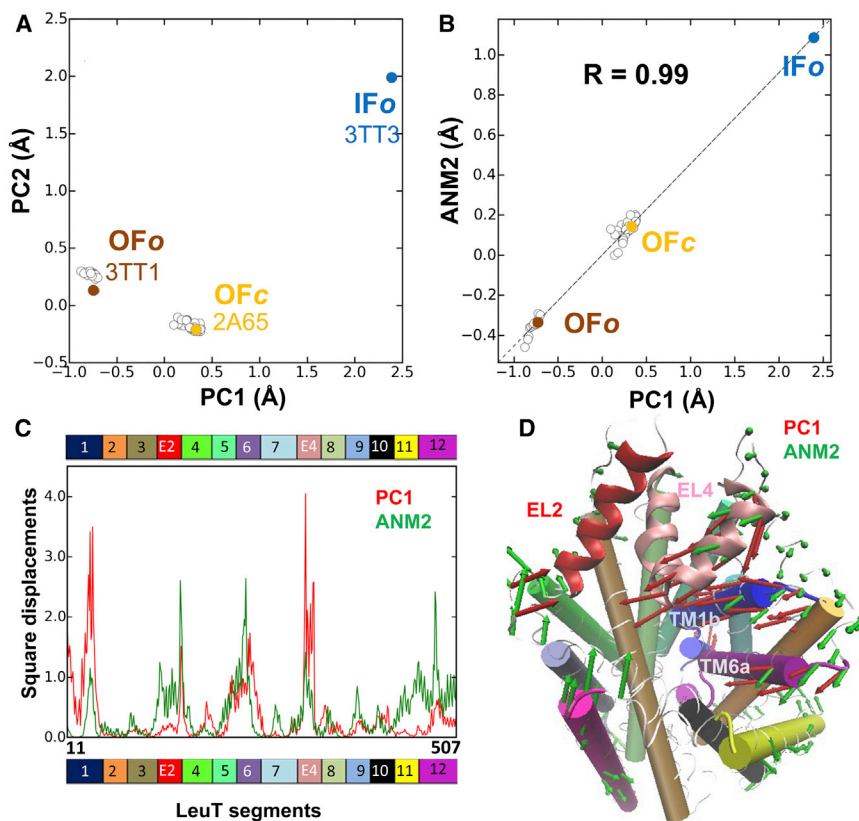
FIGURE 3 Correlation between collective motions predicted by ANM and experimentally observed structural changes in LeuT. (*A*) Projection of the 50 crystal structures of LeuT monomer onto the space of conformations spanned by the principal components PC1 and PC2 shows three clusters, which broadly correspond to the crystallographically resolved LeuT structures in the OF*o*, OF*c*, and IF*o* states. (*B*) Close correspondence between the structural variations along PC1 and the structural changes along ANM2. The close agreement between PC1 and ANM2 is further illustrated in (*C*). Here the distribution of square displacements of residues along the dominant structural difference (PC1; *red*) is compared with the predicted soft mode (ANM2; *green*). (*D*) Comparison of residue displacements along PC1 (*red arrows*) and ANM2 (*green arrows*). We used ProDy (52) for analysis and visualization.

transcriptase (RT) (Fig. 4 *A*), yields a correlation of 0.79 with the second-softest mode (ANM2; see Movie S1 in the Supporting Material) predicted for the RT complexed with nevirapine (56), and a *Co(m)* = 0.9 for *m* = 6 (Fig. 4, *B* and *C*).

To assess the significance of these numbers, we consider the relationship $\Sigma_k \, cos^2(\boldsymbol{d}_{3N}, \boldsymbol{u}_k) = 1$, where the summation is over all modes $\boldsymbol{u}_k$, $1 \leq k \leq 3N\text{-}6$. This equality implies that the average contribution of a random mode $\boldsymbol{u}_i$ would be $<cos^2(\boldsymbol{d}_{3N}, \boldsymbol{u}_i)> = (3N\text{-}6)^{-1}$. The correlation of 0.79 between $\boldsymbol{u}_2$ and $\boldsymbol{d}_{3N}$ for HIV-1 RT, for example, indicates an enhancement by a factor of $f > 0.79/(3N\text{-}6)^{-1/2} = 42$ (using $N \approx 950$ for RT). Therefore, the observed correlation in RT or many other applications (6) between the easy modes of motion that are intrinsically accessible and the structural changes relevant to function (or dysfunction), is far from random: it shows the predisposition of the protein to selectively undergo those conformational changes. This also signals the evolution of the structure to favor these changes, as will be discussed below.

Secondly, we consider an even more stringent test using the complete ensemble of structures resolved (for the same protein) under different conditions/states. The PDB contains several such examples, i.e., proteins for which hundreds of structures have been resolved. Typical examples are thoroughly studied drug targets resolved in the presence of different agonists/antagonists. The idea is to analyze the

space of conformations experimentally detected for each of these proteins and compare the dominant changes in structure observed in this dataset with those predicted by the ANM. This is accomplished by the following three-step analysis (32,57,58).

1) We first examine the principal structural differences between the members of the ensemble of structures resolved for a given protein. This is achieved by principal component analysis (PCA) of aligned structures (52,58); Figs. 3 *A* and 4 *D* display such ensembles of structures resolved for LeuT and RT, respectively (50 and 201 respective structures each represented by a dot), projected onto the subspace spanned by the first two principal components PC1 and PC2. The structures that cluster in different regions are characteristic of different substates: outward-facing open (OF*o*), outward-facing closed (OF*c*), and inward-facing open (IF*o*) for LeuT (see also Fig. 1), or complexes with inhibitors (*blue dots*), with DNA/RNA (*green*), or unliganded (*red*) for RT.

2) We select a representative structure from the ensemble (e.g., one centrally located in a given cluster) and perform an ANM analysis to determine the dominant modes (ANM1, ANM2, etc.) accessible to that structure. The goal is to see if the transition into (or motion toward) other functional structures (or clusters) is enabled by motions along one or more soft modes.
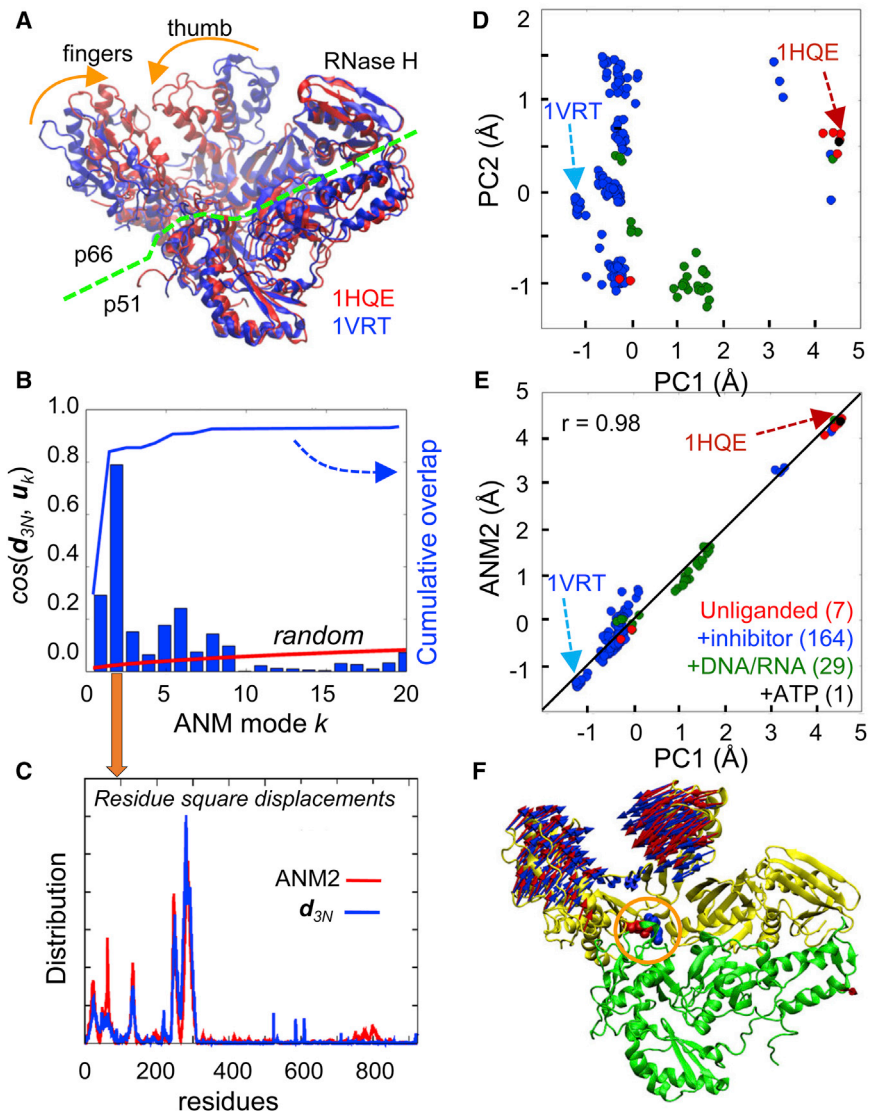
FIGURE 4 Comparison of theoretical predictions and experimental observations for HIV-1 RT heterodimer. (A) Difference d_{3N} between an unliganded (1HQE) [54] and NNRTI-bound (1VRT) [56] structures. Major structural difference occurs at the fingers and thumb subdomains of the subunit p66. The green dashed line indicates the approximate boundary between p66 and p51 subunits. (B) Correlation cosines (blue bars) between d_{3N} and soft ANM modes (u_k, for $k \leq 20$) computed for 1VRT. The cumulative overlap is shown by the blue line. The red line shows the cumulative overlap expected in the absence of any orientational correlation between d_{3N} and ANM modes. (C) Comparison of the residue square-displacement profiles based on experiments (d_{3N}, blue line) and computations (ANM2, red line). (D) Projection of the ensemble of 7 unliganded (red), 164 inhibitor bound (blue), 29 DNA/RNA bound (green), and 1 ATP-bound (black) RT structures in the PDB onto the subspace of conformations spanned by the two principal components of structural differences, PC1 and PC2. (E) Close agreement between the structural changes along PC1 (experiments) and ANM2 (theory; predicted for the unliganded RT (1RTJ) [55]). (F) Comparison of the displacements along PC1 (blue arrows) and ANM2 (red arrows). The p66 and p51 subunits are colored yellow and green, respectively. Hinge residues, E138 (p51; red space filling) and K101 (p66, blue space filling), at the intersubunit interface are highlighted by the orange circle.

3) We compare the principal components PC1, PC2, etc., deduced from experimental data and the dominant modes ANM1, ANM2, etc., predicted by the ANM. The former gives the structural variations detected to date; and the latter, the structural changes most likely to occur, based on ANM predictions. Fig. 3, B–D, shows that ANM2 computed for LeuT OFo structure correlates strongly with PC1, in support of the intrinsic ability of the OFo LeuT to reconfigure into OFc, and then into IFo. Likewise ANM calculations for unliganded RT show the close correspondence between PC1 and ANM2 (Fig. 4, E and F).

The benchmarking against large sets of structures provides solid, unambiguous evidence for the relevance of the ENM-predicted motions, based on a single resolved structure, to cooperative changes in conformations that proteins and their complexes and assemblies potentially undergo, while maintaining their native fold under physiological conditions. These motions are enabled by hinge sites, or key interactions at the hinge regions, usually at the interface between domains. Not surprisingly, drug-binding sites coincide with the global hinge sites in many proteins.

Future directions: bridging structural dynamics and sequence evolution

Computational studies illustrated above highlight the ability of proteins to sample a continuum of structures, rather than discrete substates, while maintaining their native fold. Clearly, different extents of deformations (usually along one of the principal modes of reconfiguration) are stabilized depending on the identity of the substrate. The ability to adapt to different extents of deformations is indeed a requirement for optimizing the interactions with substrates,
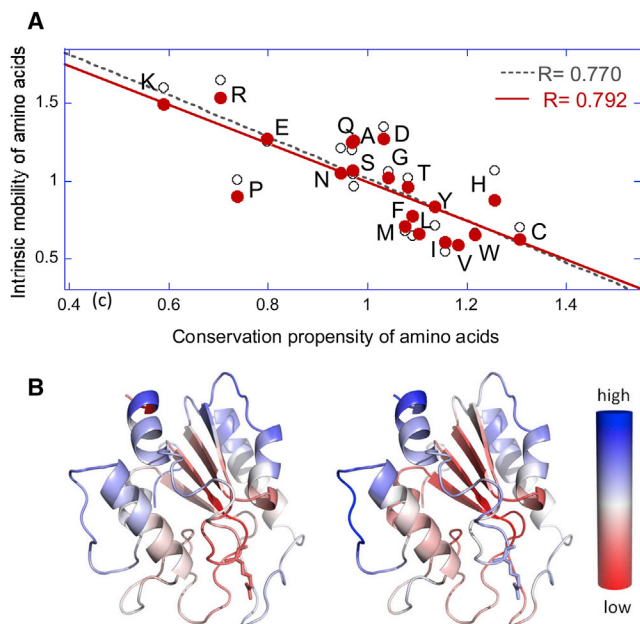
FIGURE 5 Correlation between residue conservation and conformational mobility. (*A*) Results from statistical analysis of the conservation propensities and intrinsic mobilities of amino acids. Both properties are expressed as enrichments relative to average behavior of all residues over a representative set of proteins (see Liu and Bahar (3) for details). Conservation is based on Shannon entropy, lower entropy indicating higher conservation; and mobility refers to ANM-predicted root-mean-square fluctuations in residue position. (*B*) Color-coded ribbon diagrams for human uracil-DNA glycosylase, illustrating the close similarity between Shannon entropy (*left*) and mobility (*right*) profiles of residues.

endogenous or exogenous as they accomplish their activities. These observations bring us to the significance of intrinsic dynamics in evolution.

The question is: Does the sequence evolve to favor the structural dynamics that lends itself to function? A large body of studies is directed at interpreting evolutionary properties in the light of molecular structure and stability. For example, core residues tend to be conserved as they play a key role in stabilizing the fold; local packing density appears to be a major determinant of protein sequence evolution rate (59). Amino acid positions at secondary structures are usually maintained during evolution, whereas disordered regions are less conserved per amino acid position—although the composition of amino acids at disordered regions, as well as the length and position of those regions tend to be conserved (60). In addition, a multitude of methods have been developed for deriving information on coevolving pairs (or correlated mutations), and making inferences on structure based on the hypothesis that those pairs of amino acids that exhibit strong coevolutionary signals are likely to make contacts in the three-dimensional structure (61–65). A recent study (66) provides an extensive comparison of the predictive ability of several such methods for detecting coevolutionary patterns, which also confirms the utility of such methods for inferring structural contacts.

Although the relationship of structure and evolution is widely recognized, there are now increasing numbers of studies drawing attention to the functional significance of dynamics and exploring the role of structural dynamics, or structural flexibility in enabling functional events and in evolution (2,3,67–70). The reasoning behind these studies is simple: if a major incentive for sequence selection (or mutation) is to enable (or restore) function, and if the intrinsic dynamics is key to accomplishing functional changes in structure, evolutionary selection is shaped by not only the stability of the structure, but also by its inherent flexibility. The mechanical properties of the structure are indeed as important as their chemical properties—hence the term mechanochemical entities for biomolecular systems. It is widely established that chemically active sites (e.g., catalytic sites) are conserved. Why would it not be the same for the mechanically key sites, e.g., for residues enabling hinge-bending movements between cooperative domain rearrangements? Several studies now show that it is not sufficient, nor adequate, to design extremely stable structures if the goal is to mimic/capture biological functions; on the contrary, marginal stability, or intrinsic flexibility, or elasticity, is key to functionality.

A simple comparison of the mobility scale of amino acids with their evolutionary conservation propensities yields a correlation coefficient of 0.77 (Fig. 5 *A*). Residues that exhibit large fluctuations in their coordinates, or those enjoying high mobility under native state conditions, are sequentially variable, whereas those highly constrained (low mobility) are conserved. Note that mobility here refers to the displacements of residues, rather than their flexibility. Not all flexible residues are highly mobile. For example, glycine is very flexible and it is often located at hinge regions; the hinge regions are usually fixed in space. So, an average over all glycines (in a set of proteins) yields an intermediate mobility. Likewise, residues that tend to be buried in the core exhibit low mobilities, whereas charged residues appear to have high mobilities. The color-coded ribbon diagrams in Fig. 5 *B* illustrate the close similarity between the conservation (*left*) and mobility (*right*) profiles of residues in an enzyme. Although the relation between sequence conservation and structural mobility may contain other effects (e.g., surface residues tend to be more mobile, in contrast to those buried in the hydrophobic core), the existence of an (anti)correlation between conservation and intrinsic flexibility cannot be overlooked. Our systematic study of 34 families of enzymes showed that the correlations become even more pronounced when focusing on the mobilities in the global modes (3). In a series of studies, the dynamic flexibility index determined from ENMs as a measure of mobility is used to understand the sequence evolution and functional enhancement (52,71–73).

Further analysis of coevolving residues shows that many such pairs do indeed make contacts in the three-dimensional structure. A systematic analysis of coevolving pairs showed

that a large fraction of coevolved residues are located close to each other (74). In particular, 80% of charge compensatory substitutions are within very close proximity (74), if not making direct contacts. However, calculations indicate other correlated pairs, which occupy distant positions in the structure. It remains to be systematically examined whether those correlated mutations originate from allosteric effects, whether they could be rationalized on the basis of long-range couplings characteristic of ENM global modes of motions. Systematic studies of allosteric proteins may shed light to the possible functional origin of those coevolving distal sites.

SUPPORTING MATERIAL

One movie is available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00586-X.

AUTHOR CONTRIBUTIONS

I.B. designed the research. M.H.C., J.Y.L, C.K., and S.Z. performed the research. I.B. wrote the manuscript with input from all authors.

ACKNOWLEDGMENTS

REFERENCES

1. Yang, L. W., and I. Bahar. 2005. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure.* 13:893–904.

2. Tokuriki, N., and D. S. Tawfik. 2009. Protein dynamism and evolvability. *Science.* 324:203–207.

3. Liu, Y., and I. Bahar. 2012. Sequence evolution correlates with structural dynamics. *Mol. Biol. Evol.* 29:2253–2263.

4. Smock, R. G., and L. M. Gierasch. 2009. Sending signals dynamically. *Science.* 324:198–203.

5. Liu, Y., and I. Bahar. 2010. Toward understanding allosteric signaling mechanisms in the ATPase domain of molecular chaperones. *Pac. Symp. Biocomput.* 15:269–280.

6. Bahar, I., T. R. Lezon, …, E. Eyal. 2010. Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.* 39:23–42.

7. Cui, Q., and I. Bahar. 2005. Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems. Q. Cui and I. Bahar, editors Chapman & Hall/CRC,, Boca Ratan, FL.

8. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.

9. Haliloglu, T., I. Bahar, and B. Erman. 1997. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79:3090.

10. Bahar, I., A. R. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* 2:173–181.

11. Hinsen, K. 1998. Analysis of domain motions by approximate normal mode calculations. *Proteins.* 33:417–429.

12. Hinsen, K., A. Thomas, and M. J. Field. 1999. Analysis of domain motions in large proteins. *Proteins.* 34:369–382.

13. Thomas, A., K. Hinsen, …, D. Perahia. 1999. Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study. *Proteins.* 34:96–112.

14. Atilgan, A. R., S. R. Durell, …, I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.

15. Lyman, E., J. Pfaendtner, and G. A. Voth. 2008. Systematic multiscale parameterization of heterogeneous elastic network models of proteins. *Biophys. J.* 95:4183–4192.

16. Yang, L., G. Song, and R. L. Jernigan. 2009. Protein elastic network models and the ranges of cooperativity. *Proc. Natl. Acad. Sci. USA.* 106:12347–12352.

17. Leioatts, N., T. D. Romo, and A. Grossfield. 2012. Elastic network models are robust to variations in formalism. *J. Chem. Theory Comput.* 8:2424–2434.

18. Zhao, G., J. R. Perilla, …, P. Zhang. 2013. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature.* 497:643–646.

19. Chandler, D. E., J. Strumpfer, …, K. Schulten. 2014. Light harvesting by lamellar chromatophores in Rhodospirillum photometricum. *Biophys. J.* 106:2503–2510.

20. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2011. How fast-folding proteins fold. *Science.* 334:517–520.

21. Grossfield, A., and D. M. Zuckerman. 2009. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.* 5:23–48.

22. Zuckerman, D. M. 2011. Equilibrium sampling in biomolecular simulations. *Annu. Rev. Biophys.* 40:41–62.

23. Cheng, M. H., and I. Bahar. 2013. Coupled global and local changes direct substrate translocation by neurotransmitter-sodium symporter ortholog LeuT. *Biophys. J.* 105:630–639.

24. Krishnamurthy, H., and E. Gouaux. 2012. X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature.* 481:469–474.

25. Yamashita, A., S. K. Singh, …, E. Gouaux. 2005. Crystal structure of a bacterial homologue of Na^+/Cl^- dependent neurotransmitter transporters. *Nature.* 437:215–223.

26. Penmatsa, A., K. H. Wang, and E. Gouaux. 2013. X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature.* 503:85–90.

27. Cheng, M. H., E. Block, …, I. Bahar. 2015. Insights into the mechanisms of dopamine transporter function modulation by emphetamine, orphenadrine and cocaine biding. *Front. Neurol.* 6:134.

28. Das, A., M. Gur, …, B. Roux. 2014. Exploring the conformational transitions of biomolecular systems using a simple two-state anisotropic network model. *PLOS Comput. Biol.* 10:e1003521.

29. Forrest, L. R. 2013. Structural biology. (Pseudo-)symmetrical transport. *Science.* 339:399–401.

30. Jardetzky, O. 1966. Simple allosteric model for membrane pumps. *Nature.* 211:969–970.

31. Perez, C., C. Koshy, …, C. Ziegler. 2012. Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter BetP. *Nature.* 490:126–130.

32. Bakan, A., and I. Bahar. 2009. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc. Natl. Acad. Sci. USA.* 106:14349–14354.

33. Yang, Z., P. Májek, and I. Bahar. 2009. Allosteric transitions of supramolecular systems explored by network models: application to chaperonin GroEL. *PLOS Comput. Biol.* 5:e1000360.

34. Tekpinar, M., and W. Zheng. 2010. Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model. *Proteins.* 78:2469–2481.

35. Lee, H., S. Seo, …, M. K. Kim. 2014. Opening and closing of a toroidal group II chaperonin revealed by a symmetry constrained elastic network model. *Protein Sci.* 23:703–713.

36. Ovchinnikov, V., B. L. Trout, and M. Karplus. 2010. Mechanical coupling in myosin V: a simulation study. *J. Mol. Biol.* 395:815–833.

37. Düttmann, M., Y. Togashi, …, A. S. Mikhailov. 2012. Myosin-V as a mechanical sensor: an elastic network study. *Biophys. J.* 102:542–551.

38. Düttmann, M., M. Mittnenzweig, …, A. S. Mikhailov. 2012. Complex intramolecular mechanics of G-actin—an elastic network study. *PLoS ONE.* 7:e45859.

39. Li, C. H., Y. X. Yang, …, C. X. Wang. 2014. Allosteric transitions of the maltose transporter studied by an elastic network model. *Biopolymers.* 101:758–768.

40. Jiang, J., I. H. Shrivastava, …, S. G. Amara. 2011. Large collective motions regulate the functional properties of glutamate transporter trimers. *Proc. Natl. Acad. Sci. USA.* 108:15141–15146.

41. Lezon, T. R., and I. Bahar. 2012. Constraints imposed by the membrane selectively guide the alternating access dynamics of the glutamate transporter GltPh. *Biophys. J.* 102:1331–1340.

42. Yeheskel, A., T. Haliloglu, and N. Ben-Tal. 2010. Independent and cooperative motions of the Kv1.2 channel: voltage sensing and gating. *Biophys. J.* 98:2179–2188.

43. Bertaccini, E. J., J. R. Trudell, and E. Lindahl. 2010. Normal mode gating motions of a ligand-gated ion channel persist in a fully hydrated lipid bilayer model. *ACS Chem. Neurosci.* 1:552–558.

44. Zhu, F., and G. Hummer. 2010. Pore opening and closing of a pentameric ligand-gated ion channel. *Proc. Natl. Acad. Sci. USA.* 107:19814–19819.

45. Zheng, W., and A. Auerbach. 2011. Decrypting the sequence of structural events during the gating transition of pentameric ligand-gated ion channels based on an interpolated elastic network model. *PLOS Comput. Biol.* 7:e1001046.

46. Cheng, M. H., and R. D. Coalson. 2012. Energetics and ion permeation characteristics in a glutamate-gated chloride (GluCl) receptor channel. *J. Phys. Chem. B.* 116:13637–13643.

47. Gofman, Y., C. Schärfe, …, N. Ben-Tal. 2014. Structure, dynamics and implied gating mechanism of a human cyclic nucleotide-gated channel. *PLOS Comput. Biol.* 10:e1003976.

48. Meireles, L., M. Gur, …, I. Bahar. 2011. Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. *Protein Sci.* 20:1645–1658.

49. Rai, B. K., G. J. Tawa, …, C. Humblet. 2010. Modeling G protein-coupled receptors for structure-based drug discovery using low-frequency normal modes for refinement of homology models: application to H3 antagonists. *Proteins.* 78:457–473.

50. Seckler, J. M., N. Leioatts, …, A. Grossfield. 2013. The interplay of structure and dynamics: insights from a survey of HIV-1 reverse transcriptase crystal structures. *Proteins.* 81:1792–1801.

51. Clare, D. K., D. Vasishtan, …, H. R. Saibil. 2012. ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell.* 149:113–123.

52. Bakan, A., A. Dutta, …, I. Bahar. 2014. Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics.* 30:2681–2683.

53. Koike-Takeshita, A., T. Arakawa, …, T. Shimamura. 2014. Crystal structure of a symmetric football-shaped GroEL:GroES2-ATP14 complex determined at 3.8Å reveals rearrangement between two GroEL rings. *J. Mol. Biol.* 426:3634–3641.

54. Hsiou, Y., J. Ding, …, E. Arnold. 2001. The Lys103Asn mutation of HIV-1 RT: a novel mechanism of drug resistance. *J. Mol. Biol.* 309:437–445.

55. Esnouf, R., J. Ren, …, D. Stuart. 1995. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat. Struct. Biol.* 2:303–308.

56. Ren, J., R. Esnouf, …, D. Stammers. 1995. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.* 2:293–302.

57. Yang, L., G. Song, …, R. L. Jernigan. 2008. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure.* 16:321–330.

58. Yang, L. W., E. Eyal, …, A. Kitao. 2009. Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. *Bioinformatics.* 25:606–614.

59. Yeh, S. W., T. T. Huang, …, J. Echave. 2014. Local packing density is the main structural determinant of the rate of protein sequence evolution at site level. *Biomed Res. Int.* 2014:572409.

60. Forman-Kay, J. D., and T. Mittag. 2013. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure.* 21:1492–1499.

61. Morcos, F., A. Pagnani, …, M. Weigt. 2011. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. USA.* 108:E1293–E1301.

62. Marks, D. S., T. A. Hopf, and C. Sander. 2012. Protein structure prediction from sequence variation. *Nat. Biotechnol.* 30:1072–1080.

63. Jones, D. T., D. W. Buchan, …, M. Pontil. 2012. PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics.* 28:184–190.

64. Hopf, T. A., C. P. Schärfe, …, D. S. Marks. 2014. Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife.* 3:e03430.

65. Kaján, L., T. A. Hopf, …, B. Rost. 2014. FreeContact: fast and free software for protein contact prediction from sequence co-evolution. *BMC Bioinformatics.* 15:85.

66. Mao, W., C. Kaya, …, I. Bahar. 2015. Comparative study of the effectiveness and limitations of current methods for detecting sequence coevolution. *Bioinformatics.* 15:1929–1937.

67. Bhabha, G., J. Lee, …, P. E. Wright. 2011. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science.* 332:234–238.

68. Maguid, S., S. Fernandez-Alberti, and J. Echave. 2008. Evolutionary conservation of protein vibrational dynamics. *Gene.* 422:7–13.

69. Marsh, J. A., and S. A. Teichmann. 2015. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* 84:5.1–5.25.

70. Marsh, J. A., and S. A. Teichmann. 2014. Protein flexibility facilitates quaternary structure assembly and evolution. *PLoS Biol.* 12:e1001870.

71. Gerek, Z. N., S. Kumar, and S. B. Ozkan. 2013. Structural dynamics flexibility informs function and evolution at a proteome scale. *Evolut. Applic.* 6:423–433.

72. Glembo, T. J., D. W. Farrell, …, S. B. Ozkan. 2012. Collective dynamics differentiates functional divergence in protein evolution. *PLOS Comput. Biol.* 8:e1002428.

73. Zou, T., V. A. Risso, …, S. B. Ozkan. 2015. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol. Biol. Evol.* 32:132–143.

74. Chakrabarti, S., and A. R. Panchenko. 2010. Structural and functional roles of coevolved sites in proteins. *PLoS ONE.* 5:e8591.