Annual Review of Biophysics

The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution

Paul Campitelli,[1] Tushar Modi,[1] Sudhir Kumar,[2,3,4] and S. Banu Ozkan[1]

[1]Center for Biological Physics, Department of Physics, Arizona State University, Tempe, Arizona 85281, USA; email: paul.campitelli@asu.edu, tmodi@asu.edu, Banu.Ozkan@asu.edu

[2]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, Pennsylvania 19122, USA; email: s.kumar@temple.edu

[3]Department of Biology, Temple University, Philadelphia, Pennsylvania 19122, USA

[4]Center for Excellence in Genome Medicine and Research, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Annu. Rev. Biophys. 2020. 49:267–88

First published as a Review in Advance on February 19, 2020

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

https://doi.org/10.1146/annurev-biophys-052118-115517

Keywords

conformational dynamics, allostery, protein evolution, protein flexibility, molecular dynamics simulation, protein design

Abstract

Advances in sequencing techniques and statistical methods have made it possible not only to predict sequences of ancestral proteins but also to identify thousands of mutations in the human exome, some of which are disease associated. These developments have motivated numerous theories and raised many questions regarding the fundamental principles behind protein evolution, which have been traditionally investigated horizontally using the tip of the phylogenetic tree through comparative studies of extant proteins within a family. In this article, we review a vertical comparison of the modern and resurrected ancestral proteins. We focus mainly on the dynamical properties responsible for a protein's ability to adapt new functions in response to environmental changes. Using the Dynamic Flexibility Index and the Dynamic Coupling Index to quantify the relative flexibility and dynamic coupling at a site-specific, single-amino-acid level, we provide evidence that the migration of hinges, which are often functionally critical rigid sites, is a mechanism through which proteins can rapidly evolve. Additionally, we show that disease-associated mutations in proteins often result in flexibility changes even at positions distal from mutational sites, particularly in the modulation of active site dynamics.

Contents

INTRODUCTION

Proteins are incredible machines in living systems on the molecular level. Not only are they accurate, proficient, and specific in their function, but they also can acquire new functions and structures. The capability for protein evolution is profound when we consider the fact that a vast majority of today's proteins diverged from but a few common ancestors. Moreover, recent evolutionary events like the emergence of drug resistance or enzymes with the capacity to degrade new chemicals emphasize the need to expand our understanding of these distinctive and admirable features of proteins.

Modern proteins have evolved through a series of small changes from ancient times. Computational analysis of the evolutionary record of proteins offers a tractable and highly effective solution to better understand protein function and diversification. Evolution has been a single, massive, ongoing experiment in diversification and optimization of the protein sequence–structure–function relationship over billions of years, the outcomes of which are present in the sequences, structures, and functions of modern-day (i.e., extant) protein families. Comparative analysis of sequences in protein families sheds light on mutations that lead to changes in functions and provides insight into the sequence–structure–function relationship. Thus, one can work backward to uncover protein evolution as encoded in the present-day proteomes; this process is also key to working forward to genetically engineer new modified proteins by substantially speeding up future evolution.

Indeed, comparative analyses of proteins (29, 38, 91, 92, 102, 106) and protein engineering studies (70, 73) were the first attempts to provide insight into the diversity of protein structures, along with their sequence variation, to understand the sequence–structure–function relationship in protein evolution. Recent advances in sequencing, along with high-resolution structure determination, have enhanced the ability to predict ancestral sequences and their 3D structures, thus allowing us to integrate evolutionary history encoded in sequences into the study of proteins to identify the mechanism by which mutations generate new functions (2, 17, 20, 34, 67, 68, 85, 103). Furthermore, enhanced computational and experimental techniques in protein dynamics highlight the crucial role of structure-encoded dynamics in function (18, 53, 66–68, 95) and are poised to answer questions about the role that protein dynamics play in protein evolution.

NATURE UTILIZES CONFORMATIONAL DYNAMICS FOR PROTEIN EVOLUTION

The obsolete view of the single native structure has been long replaced by the view of an ensemble of states that accurately represent the native state (61, 62, 96, 108). In the ensemble model, a protein samples a variety of conformations through local changes such as loop motions, side-chain rotations, or global changes through domain rearrangement. Indeed, what makes proteins uniquely different from other macromolecules in nature is their vast conformational diversity. Allostery, commonly known as regulation at a distance, is a widely used emergent property of this ensemble picture. Rather than forming a new structure, a ligand binding to a remote site promotes a shift in the dynamics of all residues in the protein, changing the distribution of accessible conformational states in the ensemble and, thus, promoting easier access to certain conformers for allosteric regulations while restricting others (1, 8, 16, 28, 31, 32, 55, 61, 62, 74, 87, 100, 108, 109). Furthermore, the ensemble view also matches with evolution, in which the same conserved 3D native fold can adopt new functions by modulating the conformational sample space (26, 61, 89). However, identifying substitutions that modify conformational dynamics toward a new function or adaptation to a new environment remains a challenge. Many questions are left to explore: How do conformational dynamics evolve as proteins evolve? Can we relate evolutionary conservation (substitution rates per position) to conformational dynamics? If functionally critical sites are more conserved throughout evolution, then how do proteins adapt to a new environment or evolve to perform new functions? Do directed laboratory and natural evolution share the same physical principles? We explore some of these questions in this article.

First Dominant Collective Modes Are Conserved to Maintain Function

The relationship between protein evolution and conformational dynamics is complex (5, 6, 30, 47, 76, 96, 105). Protein engineering studies have shown some principles for the emergence of a new function. Notably, laboratory-directed evolution studies have demonstrated that stabilizing mutations are crucial for compensating function-altering mutations (10, 40). In addition to stabilizing mutations, the vital role of conformational dynamics has also been reported (15, 43, 96, 109). Moreover, many advances have been made through the study of homologous proteins corresponding to a family unified by a common fold but possessing different (sometimes wildly so) function(s) from one another (44, 82). Experimental and computational studies have shown that the timescales and motions of enzymatic activity can be widely different among enzyme homologs of different species, indicating that these enzymes possess fundamentally different conformational dynamics while maintaining similar folds (9, 37, 59). These studies also highlight drastic differences in specific key regions, which exhibit differences in flexibility that may be mostly responsible for functional divergence. Additionally, a correlation between coevolving residues and the local dynamics of substrate recognition sites has also been presented, highlighting the role of local dynamics in the design of substrate interactions (56).

In addition, the large-scale, global motions of a protein are determined by its specific 3D architecture. Thus, when the fold is conserved among members of a protein family, these collective, low-frequency modes should be, as well (21, 59, 60, 112, 114). These low-frequency (i.e., slow) global fluctuations of residues in a protein play an important role in synchronizing coherent motions at more considerable distances, aiding in long-distance interactions, which are allosteric in nature (114). It follows that the functional differences among structural homologs relate to changes in nonglobal modes, modes that are primarily dominated by movements in subsets of the structure rather than by the entire protein chain in an intermediate frequency range. In fact, for a majority of protein enzyme families, the signature collective motion (i.e., functional dynamics)

associated with catalysis is conserved and can be identified through common, low-frequency modes (59, 60, 112, 114). Moreover, the specific functional changes among these homologous proteins could arise from the differences in motions in the range of low-to-intermediate frequency modes that are specific to individual proteins (63, 112).

This type of horizontal approach, based on a comparison of modern-day proteins on the tips of the phylogenetic tree, far away from their ancestors, is insightful but incomplete. The protein functions have evolved vertically, as mutations throughout their history have accumulated in their ancestral protein lineages. Therefore, it is essential to incorporate this historical background, which contains both neutral and function-altering mutations. As correctly predicted by Pauling, Zuckerkandl, and their colleagues (78) in early 1960s, recent advances in statistical methods, along with the sequencing of full genomes, have made it possible to obtain ancestral sequences through protein sequence alignments in a phylogenetic framework using a variety of statistical frameworks (33, 85, 86). Most probable ancestral sequences are now constructed by synthesizing DNA molecules, and the protein is subsequently expressed, which allows for robust experimental and computational characterization of ancient proteins (86).

When a similar dynamical analysis is performed on these ancestrally reconstructed enzymes, the principle that global dynamics of a protein are conserved in evolution still holds (112). The comparison of slow global fluctuations of ancestral enzymes with their extant homologs indicates that the root mean square fluctuations obtained from slower modes of extant enzymes significantly correlate with those of their 3-billion-year-old ancestral counterparts: (*a*) The *Escherichia coli* thioredoxin (Thrx) and its last bacterial common ancestor (LBCA) show a correlation of 0.76, and (*b*) extant TEM-1 β-lactamase and its Gram-negative common ancestor (GNCA) show a correlation of 0.79. Indeed, in both cases, the fold is strictly conserved, yet the ancestral proteins are much more stable and function completely differently than their extant homologs. However, the intermediate-to-higher-frequency modes exhibit different behavior, with limited correlations of 0.12 and 0.37 for Thrx and β-lactamase, respectively, highlighting the shift in conformational ensemble during evolution (**Figure 1**).

Sequence Conservation Correlates with the Flexibility of a Position

Functional specificity among structural homologs can be connected to the evolution of intermediate frequency modes governed by the motion of local regions within a protein structure. This means that a relationship should exist between the variation in conformational dynamics of specific positions and evolutionary rates. Studies involving specific protein families and subsets of enzymes have shown that residues that act as hinges (i.e., sites with low flexibility) are generally more evolutionarily conserved than other positions for specific protein families or a subset of enzymes (23, 56, 59, 60, 63).

While the above analyses suggest that dynamics play a role in evolution, we understand that, ultimately, evolution of proteins involves a continuous accumulation of amino acid substitution at different positions, some of which are functionally critical. This begs the question of how we can relate the role of dynamics at the level of individual residue positions to evolution. A significant reason for the lack of methods incorporating dynamical changes with amino acid substitutions in evolution, despite growing realization of their importance, has been the absence of amino acid site–specific measures that can statistically quantify each position's contribution to and its substitution's impact on the structural dynamics of the protein.

Many techniques revolve around an attempt to reconstruct the networking maps of communication between regions of a given protein (31, 61, 65, 98); one such technique is the allosteric wiring diagram (AWD), which captures the most relevant residue networks participating in

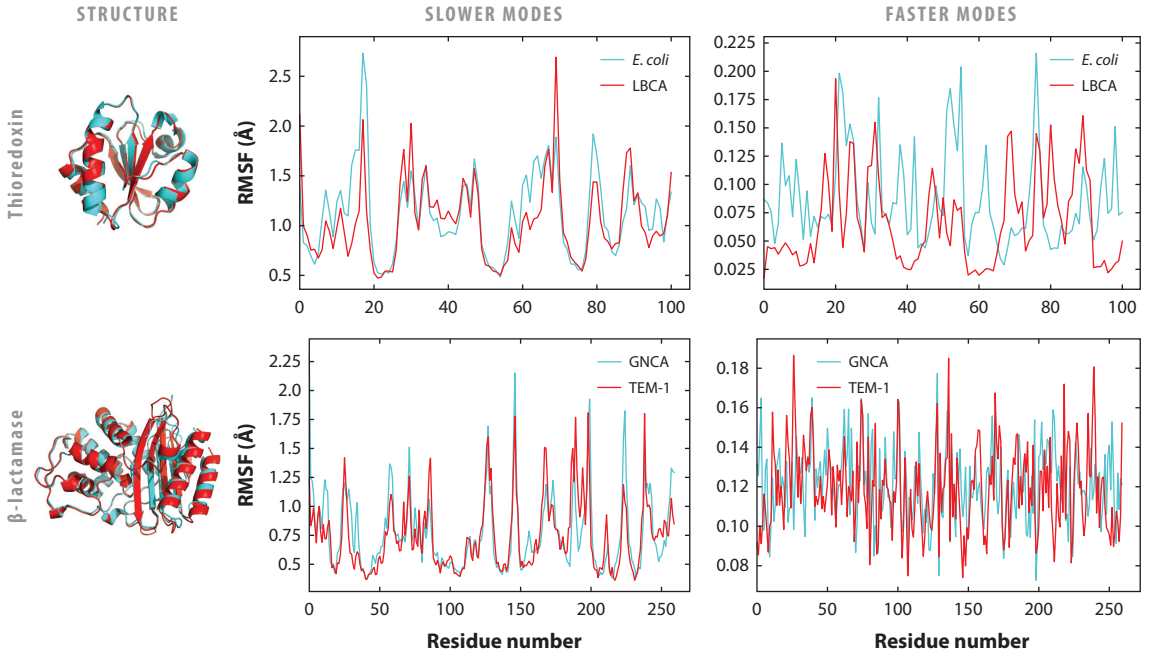

Figure 1

Slow, global motions in proteins are evolutionarily conserved, while faster modes can change dramatically between ancestral and extant proteins. Although the modern versions of thioredoxin and β-lactamase enzymes have evolved from their ancestors toward a new function, comparison of their 3D folds and the RMSF from slower modes show striking similarity, as evident from a high correlation between them (0.76 and 0.79, respectively). However, there are significant differences in the RMSF from intermediate to fast modes, indicating that the function has been fine-tuned through the modulation of intermediate or fast mode dynamics. Abbreviations: *E. coli*, *Escherichia coli*; GNCA, Gram-negative common ancestor; LBCA, last bacterial common ancestor; RMSF, root mean square fluctuation.

allosteric signal processing. Some of these methods incorporate simulated perturbations and perturbation responses to amino acids within a protein structure, such as the structural perturbation method (SPM) (98). The SPM can employ a variety of general energy functions, often using the elastic network model (ENM), in which the protein is treated as a network of nodes (residues) connected by elastic springs with harmonic potentials. In this model, the magnitude of response of a specific amino acid to a perturbation will be proportional to the elastic energy of a given mode as a result of the springs connected to that amino acid residue or position. Amino acids that respond strongly are residues that are functionally important in allosteric signaling. The SPM has been used to successfully identify AWDs in a variety of structures, including bacterial chaperonin (114), molecular motors, and DNA polymerase (113).

The success of methods such as the one mentioned above emphasizes the importance of utilizing simulated forces for exploring protein conformational dynamics. Forces are used ubiquitously in biology for important protein–protein and protein–ligand interactions, as well as in protein chaperones, which can assist in the folding and unfolding of other proteins or biological macromolecules and even facilitate the refolding of misfolded proteins (54, 57, 79). To emulate the effect of such forces, a technique has been developed that employs explicit forces to capture protein dynamics at the amino acid level: the dynamic flexibility index (DFI), which combines the ENM and linear response theory (LRT) (4, 24, 25) (see Equation 3 below). In the DFI, force is used as an additional probe to estimate the fluctuation response profile of a protein upon exertion of directed

random forces on selected residues; it allows one to sample the native ensemble efficiently and to identify long-range dynamics that propagate or modulate the allosteric communication.

The protein network of interactions is modeled either as an elastic network, as described above, or by incorporating dynamics from all-atomistic force fields for estimating the DFI. A unit force perturbation is applied to the representative node of each amino acid, acting as a random Brownian kick. This creates a response to the perturbation that then propagates through the rest of the structure and causes other positions to fluctuate through the interaction network. The fluctuation response, ΔR, of each position can be calculated through LRT (see Equation 1 below), from which a response vector is constructed to measure the magnitude and direction (x, y, z) of displacement of every residue from its equilibrium position in the native state. Averaging this response over multiple unit forces in different directions simulates an isotropic perturbation. This approach, under the harmonic approximation, closely mimics the response of a protein to an approaching substrate (11, 12, 51):

$$[\Delta R]_{3N\times 1} = [H]^{-1}_{3N\times 3N}[F]_{3N\times 1}. \quad 1.$$

In the above equation, H is the Hessian, a $3N \times 3N$ matrix that can be constructed from atomic coordinates and is composed of the second derivatives of the harmonic potential from the ENM with respect to the components of the Cartesian position vectors of length 3N. In the DFI, perturbations are introduced as random external forces exerted on selected residues, in contrast other approaches such as modifying the distances between pairs of nodes or spring constants (98). This enables us to analyze residues affected by the perturbation in a manner similar to naturally occurring regulatory motions that regulate dynamics in the cell. This approach also allows us to capture the coupling between different amino acids measured by response at one amino acid as a result of perturbations at another site.

Additionally, H can be extracted directly from molecular dynamics simulations as the inverse of the covariance matrix, which implicitly captures specific physiochemical properties of amino acids and more accurate residue–residue interactions via atomistic force fields and subsequent all-atom simulation data. Each position in the structure is perturbed sequentially, repeating the process above, for generating a perturbation response matrix A,

$$A_{N\times N} = \begin{bmatrix} |\Delta R^1|_1 & \cdots & |\Delta R^N|_1 \\ \vdots & \ddots & \vdots \\ |\Delta R^1|_N & \cdots & |\Delta R^N|_N \end{bmatrix}, \quad 2.$$

where $|\Delta R^j|_i = \sqrt{\langle(\Delta R)\rangle^2}$ is the magnitude of fluctuation response at position i due to the perturbations at position j. Subsequently, the DFI value of position i is calculated as the displacement response of position i relative to the net displacement response of the entire protein:

$$DFI_i = \frac{\sum_{j=1}^{N} |\Delta R^j|_i}{\sum_{i=1}^{N}\sum_{j=1}^{N} |\Delta R^j|_i}. \quad 3.$$

The DFI quantifies the resilience of a position to perturbations exerted at other parts of the chain through simulated mechanical force perturbations to residues in the chain. By repeatedly applying these random perturbations to each of the positions in the chain one at a time, we can compute the normalized response profile (i.e., DFI value) for every residue in the protein (**Figure 2*a***). Thus, the DFI is a relative value, being higher or lower than the average response to perturbations observed at any position. Residues with very low DFI are dynamically stable; they do not exhibit

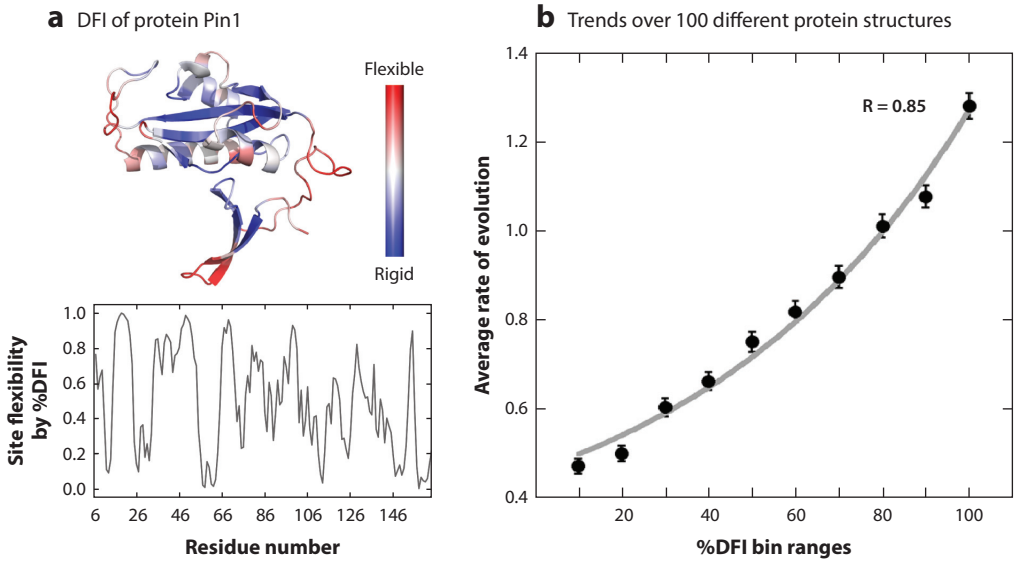

Figure 2

The dynamic flexibility index (DFI) captures amino acid flexibility at every position and correlates with evolutionary conservation. (*a*, *top*) DFI values mapped onto protein Pin1 (Protein Data Bank ID: 1PIN), where it is colored within a spectrum of red (flexible) to blue (rigid). (*Bottom*) DFI values ranked as a percentile, plotted against the residue index. Positions with %DFI values under 0.2 are considered hinge regions and are often functionally important sites. (*b*) Evolutionary rates calculated over 100 sequences from the human proteome versus %DFI values. Amino acid flexibility and evolutionary conservation are highly correlated, with more rigid positions often being more conserved.

large fluctuations upon external force perturbations, but they do play a pivotal role in transferring these perturbations throughout the chain in a cascading fashion. Thus, they will often be treated as the hinge parts of the protein that control and mediate the motion, similar to joints in a skeleton. In contrast, sites with very high DFI are structurally flexible, are prone to amino acid chain perturbations, and can play an important role in biochemical functions such as ligand recognition.

The DFI has been used successfully in a variety of contexts, including the establishment of a broad relationship between the structural dynamics of individual amino acids and their evolutionary conservation (13). In a large-scale study, the flexibility of 39,813 residues from 100 different proteins was analyzed using the DFI. It was observed that %DFI (DFI scores of residues ranked with their percentiles) strongly correlates with position-specific rates of evolutionary change obtained from multispecies sequence analysis. This result indicates that positions that are more important to protein dynamics, such as hinges, are in fact under stronger natural selection, and thus that nature permits fewer amino acid substitutions at these positions (**Figure 2*b***).

Nature Uses a Hinge-Shift Mechanism to Create New Dynamics through Evolution

While there is a correlation with the flexibility of a position and its evolutionary rate, the question of how flexibility profiles evolve throughout evolution needs further investigation. Ancestrally reconstructed proteins allow us to explore the typical behavior associated with adaptation to a new environment or the emergence of a new function in different species. For example, Thrxs are ubiquitous oxidoreductase enzymes present in all living organisms, from Archaebacteria to humans. Recently, the ancestral forms of Thrxs (i.e., Precambrian Thrxs) have been resurrected and

analyzed (80, 83, 84, 88). In agreement with ancestral protein resurrection studies, ancestral Thrxs share the same canonical 3D structure, and even similar chemical mechanisms of reduction, with modern enzymes (**Figure 3*a*,*b***). However, in accordance with their melting temperatures, they are 32°C more stable than the modern extant proteins and also have higher catalytic efficiency (i.e., higher activity) at a pH of 5. In other words, Thrxs evolved toward both lower stability and lower activity to adapt to the changes in temperature and ocean acidity, which occurred through environmental evolution from the ancient ambient conditions to the current conditions on Earth. This brings up the question of how they achieved these adaptations through sequence variation while conserving their 3D fold. To answer this question, we first performed all-atom molecular dynamics simulations and then obtained the DFI profiles of each Thrx. Comparison of the distribution of flexibility of residues in each protein reveals the differences between the ancestral and extant Thrxs for both human and *E. coli* branches. Notably, as Thrxs evolved, there was a redistribution of residues with medium flexibilities with a gain in rigid and highly flexible sites, suggesting that proteins fine-tune their activity following the functional requirement. This characteristic pattern of increasing distribution width with evolution is further supported by the high correlation between the variance of DFI distributions over time (a correlation of $R = 0.77$; $p = 1.6 \times 10^{-2}$) (**Figure 3*c***). In addition, since melting temperature correlates negatively with evolutionary time, it correlates significantly with the variance of DFI distributions (a correlation of $R = -0.86$; $p = 3.2 \times 10^{-3}$) (67) (**Figure 3*d***). Moreover, projecting the hierarchical clustering of the DFI profiles of all nine Thrxs onto the 2D map of their stability and catalytic rates indicates that these biophysical attributes are indeed associated with the flexibility of their residues (**Figure 3*e*,*f***). Thus, nature sculpts the native ensemble to adapt and alter function, a result that is in agreement with our earlier work on protein evolution (41, 115).

Comparison of how the DFI scores of the residues differ between ancestral and extant Thrxs on human and *E. coli* branches provides a plausible molecular mechanism for their adaptation to lower temperatures. In the bacterial branch of Thrxs, the increased flexibility in the α3 region (a hinge loss), which contributes most to stability, is compensated by a loss in the flexibility of α4, which is critical for folding (**Figure 4*a***). This may explain the fact that modern Thrxs have decreased stability while maintaining their canonical 3D fold (67). This mechanism of hinge shift, that is, the migration of a hinge (a position with low flexibility) from one region to another within a protein, has not been observed or investigated in other ancestral studies of Thrxs.

The same hinge-shift mechanism has also been witnessed in the evolution of TEM-1 β-lactamase (**Figure 4*b***). Unlike the modern TEM-1, which can only degrade penicillin, the last Gram-positive and Gram-negative common ancestor (PNCA) GNCA bacteria could degrade both penicillin and second-generation antibiotics with similar efficiency. The enhanced substrate promiscuity of ancestral enzymes (i.e., GNCAs and PNCAs) is not accompanied by significant changes in the active site region; the 3D crystal structure is also conserved throughout their evolution. However, the substrate-promiscuous ancestral β-lactamases exhibit high flexibility around regions close to the active site, which is significantly more rigid in penicillin-specific extant TEM-1, emphasizing the flexibility required for the binding of different ligands (115). The decreased flexibility of this region is compensated for by the rigidity of the N-terminal helix, indicating the migration of hinges in functional evolution of TEM-1 β-lactamase.

Finally, a study using reconstructed ancestors of green fluorescent protein (GFP) shows that the evolution of red color from a green ancestor was a result of migration of the hinge positions from the active site diagonally across the β-barrel fold, making sites near the chromophore more flexible and subsequently accommodating the sizeable conformational change necessary for red chromophores (41) (**Figure 4*c***). This dynamics-driven evolutionary mechanism modifies the flexibility profile of the barrel-shaped β-protein with a large number of tertiary contacts and a

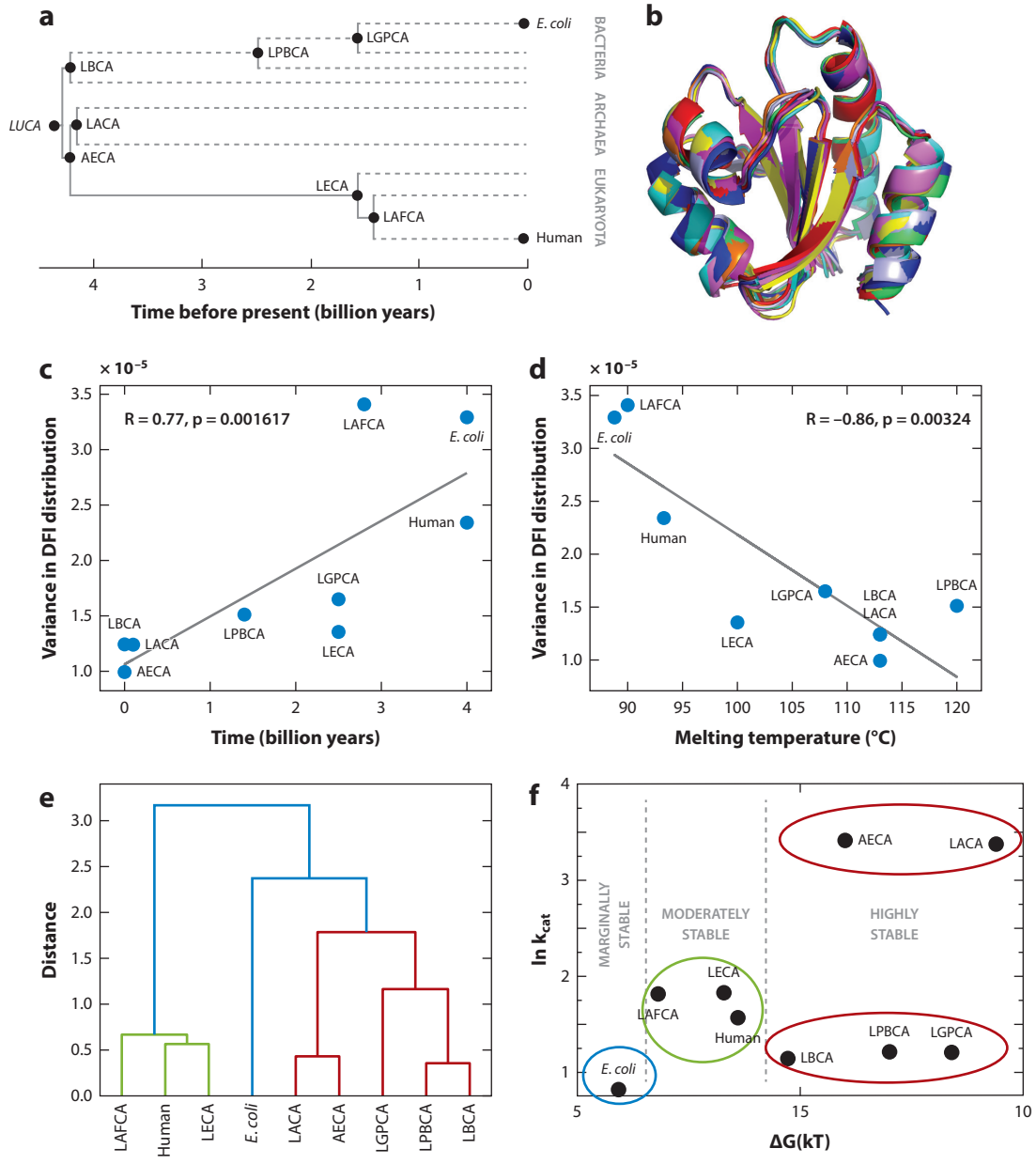

Figure 3

Proteins can evolve new functions by keeping the same 3D fold but changing the dynamical properties, which are quantified by the dynamic flexibility index (DFI). (*a*) Phylogenic tree showing human and *Escherichia coli* thioredoxin (Thrx) back to their common ancestor. (*b*) Structural overlay of modern Thrxs and their ancestors. While they exhibit significant sequence variations, the 3D folds are conserved. Variance in DFI distributions of ancestral and modern-day Thrx proteins correlates strongly with both (*c*) evolutionary time (R = 0.77) and (*d*) melting temperature (R = −0.86). (*e*) Clustering the DFI profiles of all nine Thrx variants and (*f*) projecting them onto a 2D map of catalytic rates and stability show that the DFI accurately groups the variants by these biophysical properties.

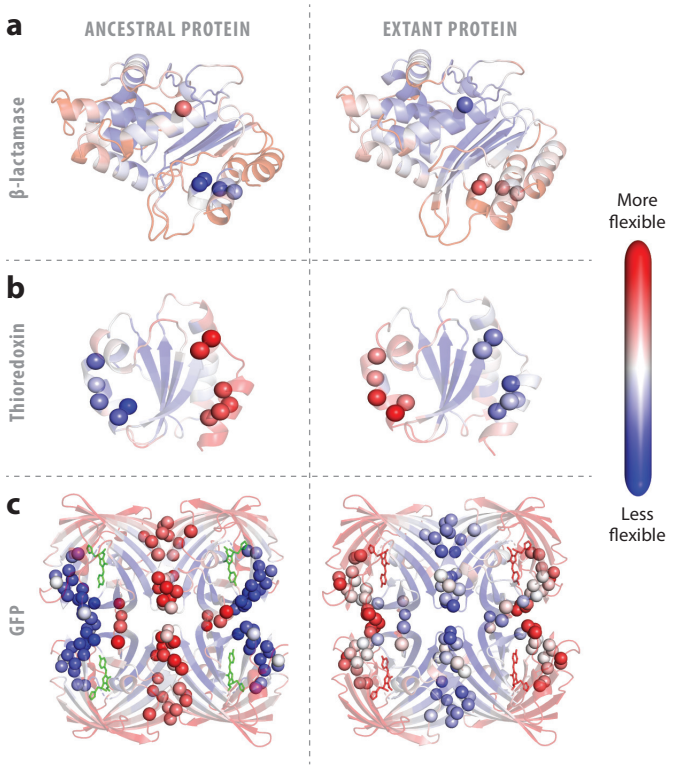

Figure 4

Nature can modulate function by relocating hinges within a protein through evolutionary changes in dynamical properties without altering the 3D structure. The hinge-shift mechanism is illustrated through the comparison of a modern protein with a reconstructed ancestral version for three different proteins, (*a*) β-lactamase, (*b*) thioredoxin, and (*c*) green fluorescent protein (GFP), and the dynamic flexibility index values are mapped onto the structures. In all cases, the location of some significant, highly flexible and highly rigid sites shifts dramatically. For each protein, the structure is conserved throughout evolution, but properties such as (*a*) ligand-binding promiscuity, (*b*) stability and catalytic rates, and (*c*) photochromatic activity all change substantially as a result of a hinge shift.

relatively closed topology, harboring a buried active site during the evolution of green color to red.

Taken together, these ancestral studies indicate that a hinge shift, i.e., enhancement of the flexibility of some rigid sites (loss of hinges at some sites) compensated for by decreased flexibility of some other distal sites (gain of hinges at some other sites), leading to change in dynamics, is a common mechanism in evolution of different protein systems with different folds. These observations prompt the next question: Do the mutational sites themselves exhibit the most substantial changes in flexibility, or does nature utilize another mechanism?

DYNAMICALLY COUPLED ALLOSTERIC POSITIONS PLAY A CRITICAL ROLE IN MODULATING DYNAMICS

Functionally critical amino acids within a protein, such as those directly involved with ligand binding, are very often highly evolutionarily conserved positions because mutations at these positions often result in dramatic and deleterious changes (49, 94). However, studies have shown that, even

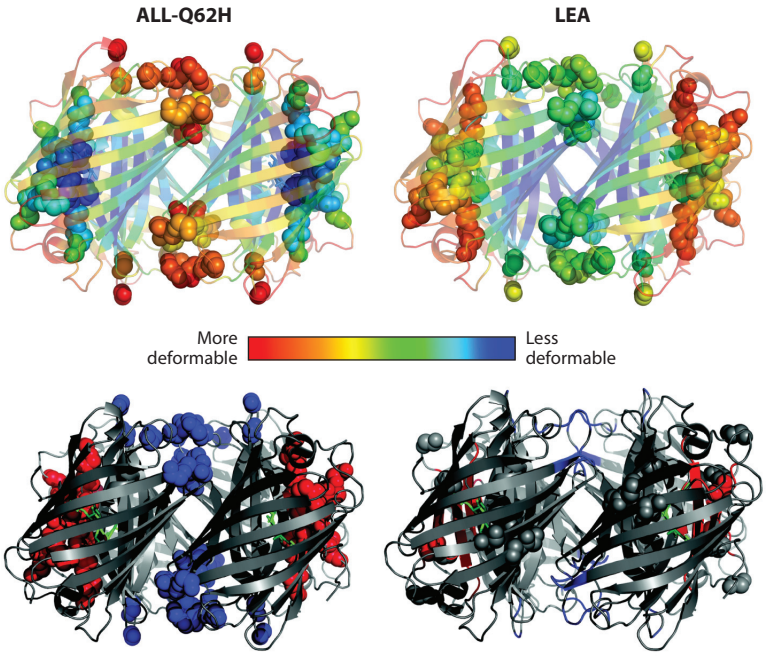

Figure 5

Mutations can induce significant changes in flexibility at distal regions, while the sites of mutations themselves can remain relatively unchanged. (*Top*) Structures of green fluorescent protein common ancestor variant ALL-Q62H (*green*) and its LEA (*red*) colored by %DFI values. Residues with significantly different flexibility are represented as space-filling spheres using main and side-chain atoms. (*Bottom*) The top 15% of ΔDFI (difference of DFI values between the LEA and ALL-Q62H) values mapped onto the ALL-Q62H structure (*right*), where red residues are more flexible and blue residues are more rigid in the LEA as compared to the ancestral homolog; the sites that do not exhibit significant change in flexibility are colored gray. This mapping presents the formation of new hinge sites that may regulate changes in function between the two enzymes. The sites of sequence variation are marked as spheres (*left*). A majority of mutations are colored gray, indicating that mutations do not significantly impact the flexibility of these positions. Note that the most significant changes in flexibility occur at regions that were not directly mutated. Abbreviations: DFI, dynamic flexibility index; LEA, least evolved ancestor.

at regions distal to binding sites or catalytic regions, point mutations may have a dramatic effect on the function of a protein (19, 32, 35, 68, 89, 100). In the absence of structural changes or significant changes to local dynamics at the region surrounding the mutation site, these distal mutations are a reliable indicator of the presence of allostery or allosteric regulation.

The importance of these allosteric mutations has been observed when the ancestral proteins are incorporated in studies (52, 89). This is particularly pronounced, as explained above, in the study of GFPs. This study showed that the mutational sites exhibit relatively small changes in flexibility, as measured by the DFI, yet these mutations significantly impact the flexibility of positions distal from these mutational sites (**Figure 5**). This indicates that the changes in amino acid sequence for green to red chromophore evolution usually did not occur at sites directly involved with photochromatic activity (i.e., at functionally critical catalytic sites).

In fact, amino acids that possess strong dynamic allosteric residue coupling (DARC) spots to other regions of the protein can affect protein function, regardless of the separation distance. As these sites are often less conserved than other, more crucial positions, it appears that nature

can take advantage of functional modulation via DARC spot mutations that can impact protein dynamics through changes in allosteric networking. The dynamic coupling index (DCI) (see Equation 4 below) is one metric that captures the complex effects of allosteric interactions regulated by DARC spots, as well as helping to describe the emergent changes in the functional behavior of DARC spot mutations (13, 14, 25, 47, 68).

Similar to the DFI, the DCI captures the strength of the displacement response of a given position i upon perturbation to a single functionally important position (or subset of positions) j, relative to the average fluctuation response of position i calculated using perturbations to all other positions within a structure:

$$DCI_i = \frac{\sum_j^{N_{functional}} |\Delta R^j|_i / N_{functional}}{\sum_{j=1}^{N} |\Delta R^j|_i / N}. \quad 4.$$

As such, the DCI can be considered a measure of the dynamic coupling between residues i and j upon perturbation to residue j.

The Distal, Non-Conserved Sites Coupled to Active Sites Are Used in Evolution to Modulate Function

The DCI has been able to identify critical allosteric interactions in a wide variety of systems (46, 47, 50, 68). For example, in a comparison study of the modern extant enzyme TEM-1, it was revealed that the majority of the mutations that contribute to resistance are distally located from the catalytic sites (68) yet strongly dynamically coupled to the active sites, as indicated by higher DCI scores. Comparison of the DFI analysis of the antibiotic-resistant mutations with that of wild-type TEM-1 revealed that these distal mutations remotely alter the flexibility of the active site to accommodate the hydrolysis of newer antibiotics (68), as observed in antibiotic-resistant ancestral β-lactamases (115). Additionally, the analysis of dynamic coupling of an exhaustive set of approximately 5,000 mutations in TEM-1 has shown that the mutations that contribute most to the emergence of a function for the degradation of a new antibiotic are those that exhibit mid-range flexibility and high dynamic coupling with the active sites. While the medium flexibility of these sites allows for substitutions, their higher coupling with the active site creates a cascading set of changes in the interaction network, leading to a change in the flexibility profile of regions that play a critical role in function.

Additionally, a DARC spot analysis of ancestral Thrxs, comparing the coupling of residues to the catalytic site residues in an ancestor with that of an extant enzyme, showed that regions far from functionally important catalytic sites contributed significantly to enzymatic activity and overall stability. Specifically, one of the α helices (α3), previously shown to impact structural stability when its formation was disrupted, exhibits much weaker dynamic coupling to these catalytic sites in extant Thrx than in the ancestors, which suggests that dynamic coupling networks could fine-tune function and stability during evolution.

Overall, small, subtle perturbations in such distally coupled sites cascade a set of changes toward the functionally active sites. Nature utilizes this principle of minimum perturbation with maximum response by allosterically altering the dynamics of the functionally critical sites, rather than acquiring new, large-effect mutations.

While allosteric mutations fine-tune the dynamics of functionally critical positions toward an evolved function, one must also wonder if the sped-up version of evolution practiced in directed evolution experiments follows the same principle. Indeed, enzymes engineered through directed evolution have yielded many cases in which distal mutations in regions that were previously

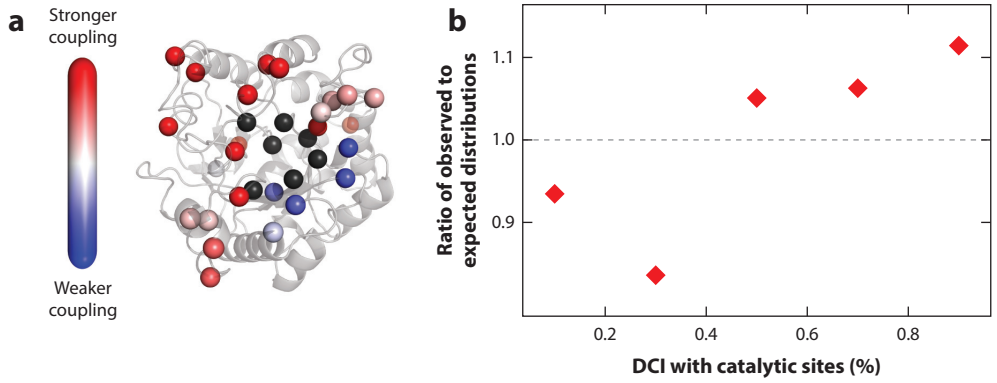

Figure 6

Function-altering mutations often occur at amino acid positions dynamically coupled to critical regions, such as catalytic sites, regardless of the structural separation distance between them. (*a*) The dynamic coupling index (DCI) values of the substitutions observed in the directed evolution of wild-type bacterial phosphotriesterases (R0) to R22 are color-coded within a spectrum of blue–white–red, where blue indicates weaker coupling, and red indicates stronger coupling. The catalytic sites are shown in black. R22 has evolved to exhibit a 10^9-fold change in the ratio of its activity to the hydrolysis of different organophosphates. We observe that a large number of acquired mutational positions exhibit higher coupling with the catalytic sites regardless of their distance, indicating the critical role of allosteric interactions of dynamic allosteric residue coupling (DARC) spots. (*b*) The observed-to-expected ratio obtained over the DCI distribution of the remote mutational sites with the DCI distribution of all 5,200 amino acid positions across 18 proteins. The distal positions that were acquired throughout the directed evolution experiments are overabundant at high DCI values (ratio = 1.11), whereas those with lower DCI values show a significant deficit (ratio = 0.84), suggesting that the majority of function-altering mutations can be DARC spots. Furthermore, it also emphasizes the importance of identifying DARC spots when designing proteins engineered toward specific functions.

thought not to affect function have actually been functionally beneficial. Some of these mutations improve thermal stability and protein expression (7, 39, 69, 99), while others improve catalytic efficiency by modulation of conformational space, thus impacting active site dynamics (15, 18, 36, 76, 97, 110). In fact, a large number of function-altering mutations that are distally positioned from catalytic sites are indeed dynamically coupled to these sites. For example, the promiscuous, low activity of bacterial phosphotriesterase (PTE) for arylester hydrolysis was carried out through 22 generations of directed evolution experiments toward arylesterase (AE) activity. A majority of mutations leading to an approximately 40,000-fold increase in AE activity and 40,000-fold decrease in the activity of PTE were far from the active site (15).

Upon analysis of the dynamic coupling of the substituted residue positions with the catalytic site via DCI analysis, the mutations selected for the emergence of new function exhibited a higher coupling with the catalytic site despite considerable separation distance (**Figure 6*a***). Additionally, in a similar analysis performed on all 18 engineered proteins containing over 100 remote and function-altering mutations (constructed from the data set in Reference 107), it was observed that a large number of distal mutations impacting function occur at residues that are highly coupled to active sites, suggesting that they are DARC spots. In a robust statistical analysis comparing the DCI distribution of the remote mutational sites with the DCI distribution of all positions (5,200 amino acid positions across 18 proteins), the observed-to-expected ratio was calculated for the remote mutations by categorizing the DCI into five bins (**Figure 6*b***). Under the null hypothesis of no effect, the ratio of the expected to observed numbers of residue positions hosting function-altering remote mutations should be close to 1.0 for each category, but the null hypothesis is soundly rejected (**Figure 6*b***). Positions exhibiting the strongest dynamic coupling show the highest enrichment of remote functions altering mutations (ratio = 1.11), whereas those with the

lower DCI values show a deficit of these variants (ratio = 0.84). This analysis also highlights the importance of DARC spots in engineering enzymes toward a desired function.

Disease-Associated Variants in the Human Exome Also Modulate Conformational Dynamics and Use Allostery

With advancements in genome sequencing efforts, there has been an exponential growth in the number of known nonsynonymous single nucleotide variants (nSNVs). Indeed, new mutations occur randomly in nature and are constantly subjected to natural selection. While many of the mutations that significantly impact organismal fitness (owing to the disruption of protein function) manifest themselves in the form of diseases in populations, mutations with small or insignificant fitness effects are found as polymorphisms (48, 49). Capturing changes in protein dynamics at the level of individual amino acids could shed light on the mechanisms underlying human sequence variations associated with disease.

It is known that disease-associated variants alter the stability of a protein (3, 27, 111). Conversely, a recent study based on high-throughput functional assays of over 2,000 variants revealed that only one-third of mutations led to a decrease in protein stability (90). Rather than affecting stability, a significant fraction of disease-associated variants impairs protein-ligand function or enzymatic activity (13, 46, 104). Additionally, disease-associated variants are not always located at highly conserved positions and/or at the positions close to or at functionally critical domains. To further complicate the problem, studies that combine evolutionary approaches with biochemistry for protein design have also revealed that disease-causing mutations at non-conserved sites can involve very complex and poorly understood mechanisms.

The basic evolutionary principle that biochemically similar substitutions at non-conserved sites do not alter function does not necessarily hold. On the contrary, regardless of biochemical similarity, amino acid substitutions at non-conserved sites can lead to a wide range of outcomes, with changes of increasing or decreasing functional activity at up to three orders of magnitude (i.e., rheostatic pattern of change) (94). Indeed, recent evolutionary analysis has revealed that approximately one-third of the residues in human proteins are relatively fast evolving (93), which means that these positions vary considerably among species with many different amino acid types allowed by natural selection. Thus, many disease-associated variants implicated in Mendelian and complex diseases are also present in one or more species. A majority of the computational tools using sequence alignments with or without structural information to diagnose such variants are likely to produce a wrong diagnosis for known disease-associated variants, which is indeed the case, as the true positive rates have been reported to be rather low (49, 64). Incorporation of conformational dynamics is therefore fundamental to more accurately identify human variants that impact biological function (e.g., non-neutral, disease-associated) and those that do not (neutral nSNVs) (14, 23, 46, 81).

Furthermore, from a biophysics perspective, variations in a human exome, which ultimately determine the constellation of proteins expressed by an individual, are already associated with more than a thousand major diseases. Because the disease-associated variants in proteins remain the part of our genome that provides the best potential for understanding how sequence relates to function via known phenotypic impact, they represent our best chance to bridge genomics and evolution with biophysical aspects of proteins and, particularly, conformational dynamics.

The first question is whether flexibility per position could be useful to identify the sites that are more susceptible to damaging, disease-associated mutation. To this end, a human proteome-wide DFI analysis of 792 disease-associated variants and 788 neutral-associated variants has shown that mutations associated with disease generally occurred at positions with low %DFI, which are

structurally important rigid regions or hinge sites, whereas neutral-associated variants exhibited opposite behavior, more often occurring at sites of intermediate or high flexibility (23). These contrasting patterns establish that structural dynamics continuously shape the protein variation present in the human population. They also suggest that metrics that measure protein dynamics, such as the DFI, have the potential to provide information that is independent of multispecies sequence alignment.

As discussed above, rigid hinge sites are often mechanistically essential and evolutionarily conserved. However, not all disease-causing mutations occur at hinge sites, or in more complex situations, at non-hinge sites located distally from other important binding or catalytic regions. A compelling test case is the human ferritin protein; disease variants are linked to a broad range of conditions including neurodegenerative diseases such as Parkinson's disease and Huntington's disease and early developed cataract syndrome. In the study of the wild-type human ferritin protein (47), which contains neutral and disease-associated variants, it was observed that the mutated sites are neither located at nor in the vicinity of the experimentally identified functional regions that act as hinges in controlling the overall motion. While neutral variants exhibit similar flexibility (i.e., DFI) profiles to those of the wild-type protein, disease mutations soften these distal and functionally critical regions of human ferritin by increasing their DFI values (**Figure 7*a***). Notably, the disease-associated variants allosterically induce changes in flexibility at two particular functionally critical regions: the C-terminal end and a regulatory loop denoted L1 (46). Thus, disease mutations may loosen hinges, in this case, impairing the allosterically regulated structural dynamics [e.g., GFPs (41)]. Indeed, DCI analysis shows that the dynamic coupling among loop L1, the C terminus, and the rest of the structure varied dramatically between the wild-type and disease variants. This suggests that cataract syndrome–associated mutations disrupt necessary allosteric regulation (**Figure 7*b***). Overall, DCI and DFI analyses suggest that disease-associated mutations soften the functionally critical regions, leading to a floppy protein with the loss of allosterically regulated conformational dynamics. This is similar to the hinge of a door; if the hinge is loosened, then the motion will not be adequately transmitted, and the door cannot function properly.

Proteins are not isolated within a cell, and they interact with one another to engage in essential biological functions (46, 47, 74, 75). Considering the critical role that protein interactions play in cellular functions, a recent study of large-scale characterization of disease variants indicates that the majority of disease variants do not alter structure or folding stability, but rather impair protein interactions (90, 104). Different variants in the same gene lead to different interaction profiles, often resulting in distinct disease phenotypes. Based on experimental analysis, one could expect that positions with the most significant impact on binding dynamics contribute the most to binding interactions, and that mutations at these positions may impair binding and, thus, function (i.e., disease causing). Indeed, in a study over the full human proteome, DFI profiles of over a thousand positions harboring neutral and disease variants revealed that interface residues have a lower average %DFI (31%) than those present at non-interfaces (50%) when complex forms are used, rather than single monomeric units alone, indicating the critical role in protein interaction played by the interface residues between the monomeric units. Interestingly, interface sites with disease-associated variants have significantly lower average %DFI (23%) compared to those of neutral nSNVs (42%), a result that directly relates structural dynamics to functional importance (13).

While mutations occur directly at important regions such as the positions that contribute most to binding free energies (protein interface hotspots) (22, 101), distal mutations far from binding sites could impact function through allosteric regulation. Some mutations allosterically impair posttranslational modification, as observed in driver mutations in cancer (13). Disease-associated

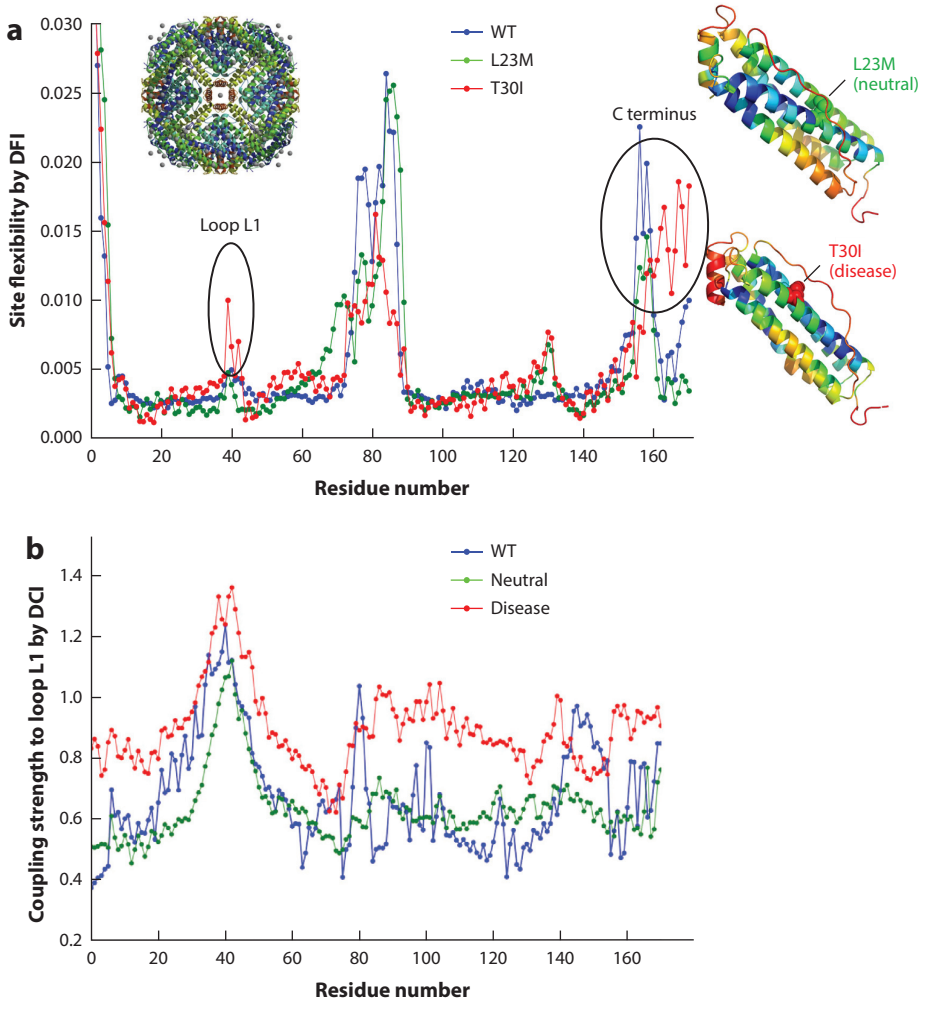

Figure 7

Disease-associated mutations can change flexibility at distant sites, as well as the overall dynamic coupling between residues within a protein. (*a*) Dynamic flexibility index (DFI) values of the light chain subunit of human ferritin protein (the full complex is shown in the inset) for wild-type (WT) variants (*blue*), average DFI of neutral variants (*green*), and average DFI of disease variants (*red*). Flexibilities for disease variants differ from the WT at functionally critical regions (loop L1, C terminus) while maintaining relatively similar flexibilities in neutral variants. Note that these significant flexibility changes occur at regions far from the mutational sites themselves. (*b*) Dynamic coupling profiles for loop L1 of WT variants (*blue*), average dynamic coupling index (DCI) of neutral variants (*green*), and average DCI of disease variants (*red*). Disease-associated variants exhibit overall stronger coupling to this region across the entire protein structure, which disrupts necessary allosteric regulation.

variants can also change the functionally active (ON) versus inactive (OFF) populations by altering the stability of particular conformations and/or conformational dynamics, as observed in cancer driver mutations of kinases (45). Furthermore, they can lead to disease by shifting allosteric pathways (71). In these cases, positions not exactly identified as hotspots could be DARC spots that may act as important nodes in the interaction network by remotely modulating the dynamics of

the binding interface; mutations at these sites would alter the interface interactions, leading to disease (46). Thus, distal allosteric mutations that modulate conformational dynamics may impact function and cause disease.

CONCLUSIONS AND FUTURE PERSPECTIVES

Proteins are the machines of living systems that carry out a diverse set of essential biochemical functions. Furthermore, the diversity of their functions has grown over time via molecular evolution. Analyses of protein families indicate that proteins evolve for different functions through sequence variation while conserving their 3D structures. This becomes even more apparent when the 3D native folds of resurrected ancestral proteins are obtained. Since ancestral proteins were adapted to intracellular and extracellular environments that are likely different from the environments hosting modern proteins, resurrected ancestral proteins exhibit high stability, substrate and catalytic promiscuity, and altered patterns of interaction with other subcellular components (46). Detailed computational conformational analyses support the hypothesis that proteins may have evolved to new or more specific modern functions by altering their ensemble of conformational states in their native (functional) state. Comparison of the distribution of flexibility of residues between ancestral proteins and their extant homologs reveals that the population density of highly flexible and rigid sites increased as they evolved. This common feature of changing the flexibility of specific positions observed in evolution suggests a fine-tuning of their native ensemble. In addition, functionally critical positions such as catalytic pockets or critical binding hotspots are sequentially conserved. To adapt or to create a new function, nature uses substitutions of distal positions that are dynamically coupled to functionally critical sites (DARC spot positions) rather than substitutions at functional sites. Mutations at DARC spots allow proteins to evolve toward a specific network of interactions that enables communication between the active sites and the rest of the protein through conformational dynamics. The conformational dynamics analysis of disease-associated variants in the human exome also suggests that disease-associated mutants hijack the same physical principles of modulation of protein dynamics that lead to loss or gain in function.

While natural selection involves beneficial mutations, the evolution of proteins involves many other critical stochastic forces (58, 72). Indeed, neutral theory has become central to the study of evolution at the molecular level. As stated by Kimura, ". . .the overwhelming majority of evolutionary changes at the molecular level are not caused by selection acting on advantageous mutants, but by random fixation of selectively neutral or very nearly neutral mutants through the cumulative effect of sampling drift (due to finite population number) under continued input of new mutations" (42, p. 381). Although this is still under debate, a bulk of conformational dynamics studies obtained by directed evolution experiments and resurrection of ancestral proteins also support the presence of stochastic forces that fine-tune function incrementally and even work to maintain function in the face of the accumulation of deleterious variations due to population-level processes (77). By bridging the fields of biophysics and evolutionary biology, we can explore how these stochastic forces shape the biophysical landscape of proteins and address emerging questions about complex nonadditive (epistatic) relationships among mutations that lead to interactions and dependence among positions and proteins in evolution.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work is supported in part by grants from the National Science Foundation (NSF-MCB 1715591) and the Moore Foundation to S.B.O. and from the National Science Foundation (Growing Convergence Research Award 1934848) to S.K. We thank I. Can Kazan and Nicholas Ose for a careful review of the manuscript.

LITERATURE CITED

1. Abrusán G, Marsh JA. 2019. Ligand-binding-site structure shapes allosteric signal transduction and the evolution of allostery in protein complexes. *Mol. Biol. Evol.* 36(8):1711–27
2. Agozzino L, Dill KA. 2018. Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *PNAS* 115(37):9092–97
3. Alber T. 1989. Mutational effects on protein stability. *Annu. Rev. Biochem.* 58:765–98
4. Atilgan C, Gerek ZN, Ozkan SB, Atilgan AR. 2010. Manipulation of conformational change in proteins by single-residue perturbations. *Biophys. J.* 99(3):933–43
5. Bahar I, Atilgan AR, Demirel MC, Erman B. 1998. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* 80(12):2733–36
6. Bahar I, Cheng MH, Lee JY, Kaya C, Zhang S. 2015. Structure-encoded global motions and their role in mediating protein-substrate interactions. *Biophys. J.* 109(6):1101–9
7. Baier F, Hong N, Yang G, Pabis A, Miton CM, et al. 2019. Cryptic genetic variation shapes the adaptive evolutionary potential of enzymes. *eLife* 8:e40789
8. Ben-David M, Huang H, Sun MGF, Corbi-Verge C, Petsalaki E, et al. 2019. Allosteric modulation of binding specificity by alternative packing of protein cores. *J. Mol. Biol.* 431(2):336–50
9. Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM, et al. 2013. Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat. Struct. Mol. Biol.* 20(11):1243–49
10. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. 2006. Protein stability promotes evolvability. *PNAS* 103(15):5869–74
11. Bolia A, Gerek ZN, Keskin O, Ozkan SB, Dev KK. 2012. The binding affinities of proteins interacting with the PDZ domain of PICK1. *Proteins* 80(5):1393–408
12. Bolia A, Ozkan SB. 2016. Adaptive BP-Dock: an induced fit docking approach for full receptor flexibility. *J. Chem. Inf. Model.* 56(4):734–46
13. Butler BM, Gerek ZN, Kumar S, Ozkan SB. 2015. Conformational dynamics of nonsynonymous variants at protein interfaces reveals disease association: the role of dynamics in neutral and damaging nsSNVs. *Proteins Struct. Funct. Bioinform.* 83(3):428–35
14. Butler BM, Kazan IC, Kumar A, Ozkan SB. 2018. Coevolving residues inform protein dynamics profiles and disease susceptibility of nSNVs. *PLOS Comput. Biol.* 14(11):e1006626
15. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, et al. 2016. The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* 12(11):944–50
16. Cooper A, Dryden DT. 1984. Allostery without conformational change: a plausible model. *Eur. Biophys. J.* 11(2):103–9
17. Copp JN, Anderson DW, Akiva E, Babbitt PC, Tokuriki N. 2019. Exploring the sequence, function, and evolutionary space of protein superfamilies using sequence similarity networks and phylogenetic reconstructions. *Methods Enzymol.* 620:315–47
18. Coyle SM, Flores J, Lim WA. 2013. Exploitation of latent allostery enables the evolution of new modes of MAP kinase regulation. *Cell* 154(4):875–87
19. Curado-Carballada C, Feixas F, Osuna S. 2019. Molecular dynamics simulations on *Aspergillus niger* monoamine oxidase: conformational dynamics and inter-monomer communication essential for its efficient catalysis. *Adv. Synth. Catal.* 361(11):2718–26
20. Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW. 2012. Evolution of increased complexity in a molecular machine. *Nature* 481(7381):360–64
21. Fuglebakk E, Echave J, Reuter N. 2012. Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics* 28(19):2431–40

22. Geng C, Xue LC, Roel-Touris J, Bonvin AMJJ. 2019. Finding the ΔΔG spot: Are predictors of binding affinity changes upon mutations in protein-protein interactions ready for it? *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 9(5):e1410
23. Gerek ZN, Kumar S, Ozkan SB. 2013. Structural dynamics flexibility informs function and evolution at a proteome scale. *Evol. Appl.* 6(3):423–33
24. Gerek ZN, Ozkan SB. 2010. A flexible docking scheme to explore the binding selectivity of PDZ domains. *Protein Sci.* 19(5):914–28
25. Gerek ZN, Ozkan SB. 2011. Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. *PLOS Comput. Biol.* 7(10):e1002154
26. Gobeil SMC, Ebert MCCJC, Park J, Gagné D, Doucet N, et al. 2019. The structural dynamics of engineered β-lactamases vary broadly on three timescales yet sustain native function. *Sci. Rep.* 9:6656
27. Guerois R, Nielsen JE, Serrano L. 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* 320(2):369–87
28. Guo J, Zhou H-X. 2016. Protein allostery and conformational dynamics. *Chem. Rev.* 116:6503–15
29. Halabi N, Rivoire O, Leibler S, Ranganathan R. 2009. Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138(4):774–86
30. Haliloglu T, Bahar I. 2015. Adaptability of protein structures to enable functional interactions and evolutionary implications. *Curr. Opin. Struct. Biol.* 35:17–23
31. Hawkins RJ, McLeish TCB. 2004. Coarse-grained model of entropic allostery. *Phys. Rev. Lett.* 93(9):098104
32. Hilser VJ. 2010. Biochemistry: an ensemble view of allostery. *Science* 327(5966):653–54
33. Hobbs JK, Shepherd C, Saul DJ, Demetras NJ, Haaning S, et al. 2012. On the origin and evolution of thermophily: reconstruction of functional Precambrian enzymes from ancestors of Bacillus. *Mol. Biol. Evol.* 29(2):825–35
34. Ingles-Prieto A, Ibarra-Molero B, Delgado-Delgado A, Perez-Jimenez R, Fernandez JM, et al. 2013. Conservation of protein structure over four billion years. *Structure* 21(9):1690–97
35. Jiménez-Osés G, Osuna S, Gao X, Sawaya MR, Gilson L, et al. 2014. The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat. Chem. Biol.* 10(6):431–36
36. Kaltenbach M, Jackson CJ, Campbell EC, Hollfelder F, Tokuriki N. 2015. Reverse evolution leads to genotypic incompatibility despite functional and active site convergence. *eLife* 4:e06492
37. Keskin O, Bahar I, Jernigan RL, Beutler JA, Shoemaker RH, et al. 2000. Characterization of anticancer agents by their growth inhibitory activity and relationships to mechanism of action and structure. *Anticancer Drug Des.* 15(2):79–98
38. Khersonsky O, Roodveldt C, Tawfik DS. 2006. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* 10(5):498–508
39. Khersonsky O, Röthlisberger D, Dym O, Albeck S, Jackson CJ, et al. 2010. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J. Mol. Biol.* 396(4):1025–42
40. Khersonsky O, Tawfik DS. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* 79:471–505
41. Kim H, Zou T, Modi C, Dörner K, Grunkemeyer TJ, et al. 2015. A hinge migration mechanism unlocks the evolution of green-to-red photoconversion in GFP-like proteins. *Structure* 23(1):34–43
42. Kimura M. 1991. The neutral theory of molecular evolution: a review of recent evidence. *Idengaku Zasshi* 66(4):367–86
43. Knies JL, Cai F, Weinreich DM. 2017. Enzyme efficiency but not thermostability drives cefotaxime resistance evolution in TEM-1 β-lactamase. *Mol. Biol. Evol.* 34(5):1040–54
44. Knudsen M, Wiuf C. 2010. The CATH database. *Hum. Genom.* 4(3):207–12
45. Köhler S, Bauer S, Horn D, Robinson PN. 2008. Walking the interactome for prioritization of candidate disease genes. *Am. J. Hum. Genet.* 82(4):949–58
46. Kumar A, Butler BM, Kumar S, Ozkan SB. 2015. Integration of structural dynamics and molecular evolution via protein interaction networks: a new era in genomic medicine. *Curr. Opin. Struct. Biol.* 35:135–42
47. Kumar A, Glembo TJ, Ozkan SB. 2015. The role of conformational dynamics and allostery in the disease development of human ferritin. *Biophys. J.* 109(6):1273–81

48. Kumar S, Dudley JT, Filipski A, Liu L. 2011. Phylomedicine: an evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends Genet*. 27(9):377–86
49. Kumar S, Suleski MP, Markov GJ, Lawrence S, Marco A, Filipski AJ. 2009. Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations. *Genome Res*. 19(9):1562–69
50. Larrimore KE, Kazan IC, Kannan L, Kendle RP, Jamal T, et al. 2017. Plant-expressed cocaine hydrolase variants of butyrylcholinesterase exhibit altered allosteric effects of cholinesterase activity and increased inhibitor sensitivity. *Sci. Rep.* 7(1):10419
51. Li Z, Bolia A, Maxwell JD, Bobkov AA, Ghirlanda G, et al. 2015. A rigid hinge region is necessary for high-affinity binding of dimannose to cyanovirin and associated constructs. *Biochemistry* 54(46):6951–60
52. Liang Z, Verkhivker GM, Hu G. 2019. Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. *Brief. Bioinform.* In press
53. Liberles DA, Teichmann SA, Bahar I, Bastolla U, Bloom J, et al. 2012. The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci*. 21(6):769–85
54. Lieberman RL, Wustman BA, Huertas P, Powe AC, Pine CW, et al. 2007. Structure of acid β-glucosidase with pharmacological chaperone provides insight into Gaucher disease. *Nat. Chem. Biol.* 3(2):101–7
55. Liu X, Golden LC, Lopez JA, Shepherd TR, Yu L, Fuentes EJ. 2019. Conformational dynamics and cooperativity drive the specificity of a protein-ligand interaction. *Biophys. J.* 116(12):2314–30
56. Liu Y, Bahar I. 2012. Sequence evolution correlates with structural dynamics. *Mol. Biol. Evol.* 29(9):2253–63
57. Liu Y, Gierasch LM, Bahar I. 2010. Role of Hsp70 ATPase domain intrinsic dynamics and sequence evolution in enabling its functional interactions with NEFs. *PLOS Comput. Biol.* 6(9):e1000931
58. Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, et al. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* 17(11):704–14
59. Maguid S, Fernandez-Alberti S, Echave J. 2008. Evolutionary conservation of protein vibrational dynamics. *Gene* 422(1–2):7–13
60. Maguid S, Fernández-Alberti S, Parisi G, Echave J. 2006. Evolutionary conservation of protein backbone flexibility. *J. Mol. Evol.* 63(4):448–57
61. McLeish TCB, Rodgers TL, Wilson MR. 2013. Allostery without conformation change: modelling protein dynamics at multiple scales. *Phys. Biol.* 10(5):056004
62. McLeish TCB, Schaefer C, von der Heydt AC. 2018. The 'allosteron' model for entropic allostery of self-assembly. *Philos. Trans. R. Soc. B* 373(1749):20170186
63. Mikulska-Ruminska K, Shrivastava I, Krieger J, Zhang S, Li H, et al. 2019. Characterization of differential dynamics, specificity, and allostery of lipoxygenase family members. *J. Chem. Inf. Model.* 59(5):2496–508
64. Miller M, Bromberg Y, Swint-Kruse L. 2017. Computational predictors fail to identify amino acid substitution effects at rheostat positions. *Sci. Rep.* 7:41329
65. Mishra SK, Jernigan RL. 2018. Protein dynamic communities from elastic network models align closely to the communities defined by molecular dynamics. *PLOS ONE* 13(6):e0199225
66. Mishra SK, Kandoi G, Jernigan RL. 2019. Coupling dynamics and evolutionary information with structure to identify protein regulatory and functional binding sites. *Proteins Struct. Funct. Bioinform.* 87(10):850–68
67. Modi T, Huihui J, Ghosh K, Ozkan SB. 2018. Ancient thioredoxins evolved to modern-day stability-function requirement by altering native state ensemble. *Philos. Trans. R. Soc. B* 373(1749):20170184
68. Modi T, Ozkan SB. 2018. Mutations utilize dynamic allostery to confer resistance in TEM-1 β-lactamase. *Int. J. Mol. Sci.* 19(12):3808
69. Morley KL, Kazlauskas RJ. 2005. Improving enzyme properties: When are closer mutations better? *Trends Biotechnol.* 23(5):231–37
70. Moult J. 2008. Comparative modeling in structural genomics. *Structure* 16(1):14–16
71. Navlakha S, Kingsford C. 2010. The power of protein interaction networks for associating genes with diseases. *Bioinformatics* 26(8):1057–63

72. Nei M, Kumar S. 2000. *Molecular Evolution and Phylogenetics*. Oxford, UK: Oxford Univ. Press
73. Neuwald AF. 2007. The CHAIN program: forging evolutionary links to underlying mechanisms. *Trends Biochem. Sci.* 32(11):487–93
74. Nussinov R, Tsai C-J. 2013. Allostery in disease and in drug discovery. *Cell* 153(2):293–305
75. Nussinov R, Tsai C-J, Liu J. 2014. Principles of allosteric interactions in cell signaling. *J. Am. Chem. Soc.* 136(51):17692–701
76. Otten R, Liu L, Kenner LR, Clarkson MW, Mavor D, et al. 2018. Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat. Commun.* 9:1314
77. Patel R, Scheinfeldt LB, Sanderford MD, Lanham TR, Tamura K, et al. 2018. Adaptive landscape of protein variation in human exomes. *Mol. Biol. Evol.* 35(8):2015–25
78. Pauling L, Zuckerkandl E, Henriksen T, Lövstad R. 1963. Chemical paleogenetics: molecular "restoration studies" of extinct forms of life. *Acta Chem. Scand.* 17(Suppl.):9–16
79. Penkler D, Sensoy Ö, Atilgan C, Tastan Bishop Ö. 2017. Perturbation-response scanning reveals key residues for allosteric control in Hsp70. *J. Chem. Inf. Model.* 57(6):1359–74
80. Perez-Jimenez R, Inglés-Prieto A, Zhao Z-M, Sanchez-Romero I, Alegre-Cebollada J, et al. 2011. Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat. Struct. Mol. Biol.* 18(5):592–96
81. Ponzoni L, Bahar I. 2018. Structural dynamics is a determinant of the functional significance of missense variants. *PNAS* 115(16):4164–69
82. Ponzoni L, Zhang S, Cheng MH, Bahar I. 2018. Shared dynamics of LeuT superfamily members and allosteric differentiation by structural irregularities and multimerization. *Philos. Trans. R. Soc. B* 373(1749):20170177
83. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. 2013. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. *J. Am. Chem. Soc.* 135(8):2899–902
84. Risso VA, Manssour-Triedo F, Delgado-Delgado A, Arco R, Barroso-delJesus A, et al. 2015. Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history. *Mol. Biol. Evol.* 32(2):440–55
85. Risso VA, Sanchez-Ruiz JM. 2017. Resurrected ancestral proteins as scaffolds for protein engineering. In *Directed Enzyme Evolution: Advances and Applications*, ed. M Alcalde, pp. 229–55. Berlin: Springer
86. Risso VA, Sanchez-Ruiz JM, Ozkan SB. 2018. Biotechnological and protein-engineering implications of ancestral protein resurrection. *Curr. Opin. Struct. Biol.* 51:106–15
87. Rocks JW, Ronellenfitsch H, Liu AJ, Nagel SR, Katifori E. 2019. Limits of multifunctionality in tunable networks. *PNAS* 116(7):2506–11
88. Romero-Romero ML, Risso VA, Martinez-Rodriguez S, Ibarra-Molero B, Sanchez-Ruiz JM. 2016. Engineering ancestral protein hyperstability. *Biochem. J.* 473(20):3611–20
89. Saavedra HG, Wrabl JO, Anderson JA, Li J, Hilser VJ. 2018. Dynamic allostery can drive cold adaptation in enzymes. *Nature* 558(7709):324–28
90. Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-Huntington J, et al. 2015. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* 161(3):647–60
91. Sethi A, O'Donoghue P, Luthey-Schulten Z. 2005. Evolutionary profiles from the QR factorization of multiple sequence alignments. *PNAS* 102(11):4045–50
92. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, Ranganathan R. 2005. Evolutionary information for specifying a protein fold. *Nature* 437(7058):512–18
93. Subramanian S, Kumar S. 2006. Evolutionary anatomies of positions and types of disease-associated and neutral amino acid mutations in the human genome. *BMC Genom.* 7:306
94. Swint-Kruse L. 2016. Using evolution to guide protein engineering: The devil IS in the details. *Biophys. J.* 111(1):10–18
95. Tan L, Serene S, Chao HX, Gore J. 2011. Hidden randomness between fitness landscapes limits reverse evolution. *Phys. Rev. Lett.* 106(19):198102
96. Tawfik DS, Tokuriki N. 2009. Protein dynamism and evolvability. *Science* 324(5924):203–7
97. Taylor JL, Price JE, Toney MD. 2015. Directed evolution of the substrate specificity of dialkylglycine decarboxylase. *Biochim. Biophys. Acta Proteins Proteom.* 1854(2):146–55

98. Thirumalai D, Hyeon C, Zhuravlev PI, Lorimer GH. 2019. Symmetry, rigidity, and allosteric signaling: from monomeric proteins to molecular machines. *Chem. Rev.* 119(12):6788–821
99. Tokuriki N, Tawfik DS. 2009. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* 19(5):596–604
100. Townsend PD, Rodgers TL, Glover LC, Korhonen HJ, Richards SA, et al. 2015. The role of protein-ligand contacts in allosteric regulation of the *Escherichia coli* catabolite activator protein. *J. Biol. Chem.* 290(36):22225–35
101. Tuncbag N, Gursoy A, Keskin O. 2009. Identification of computational hot spots in protein interfaces: Combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics* 25(12):1513–20
102. Turner JM, Graziano J, Spraggon G, Schultz PG. 2006. Structural plasticity of an aminoacyl-tRNA synthetase active site. *PNAS* 103(17):6483–88
103. Wang S-W, Bitbol A-F, Wingreen NS. 2019. Revealing evolutionary constraints on proteins through sequence analysis. *PLOS Comput. Biol.* 15(4):e1007010
104. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H. 2012. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat. Biotechnol.* 30(2):159–64
105. Wei G, Xi W, Nussinov R, Ma B. 2016. Protein ensembles: How does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell. *Chem. Rev.* 116(11):6516–51
106. Weinreich DM, Delaney NF, Depristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312(5770):111–14
107. Wilding M, Hong N, Spence M, Buckle AM, Jackson CJ. 2019. Protein engineering: the potential of remote mutations. *Biochem. Soc. Trans.* 47(2):701–11
108. Wodak SJ, Paci E, Dokholyan NV, Berezovsky IN, Horovitz A, et al. 2019. Allostery in its many disguises: from theory to applications. *Structure* 27(4):566–78
109. Woldeyes RA, Sivak DA, Fraser JS. 2014. E pluribus unum, no more: from one crystal, many conformations. *Curr. Opin. Struct. Biol.* 28(1):56–62
110. Yang G, Hong N, Baier F, Jackson CJ, Tokuriki N. 2016. Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects. *Biochemistry* 55(32):4583–93
111. Yue P, Li Z, Moult J. 2005. Loss of protein structure stability as a major causative factor in monogenic disease. *J. Mol. Biol.* 353(2):459–73
112. Zhang S, Li H, Krieger JM, Bahar I. 2019. Shared signature dynamics tempered by local fluctuations enables fold adaptability and specificity. *Mol. Biol. Evol.* 36(9):2053–68
113. Zheng W, Brooks BR, Doniach S, Thirumalai D. 2005. Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. *Structure* 13(4):565–77
114. Zheng W, Brooks BR, Thirumalai D. 2006. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *PNAS* 103(20):7664–69
115. Zou T, Risso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB. 2015. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol. Biol. Evol.* 32(1):132–43

Annual Review of Biophysics
Volume 49, 2020

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Biophysics* articles may be found at http://www.annualreviews.org/errata/biophys