THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Global Dynamics Renders Protein Sites with High Functional Response

Published as part of The Journal of Physical Chemistry virtual special issue "Ruth Nussinov Festschrift".

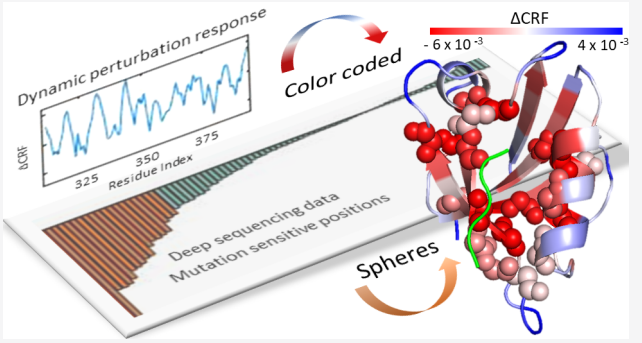
Yiğit Kutlu, Nir Ben-Tal, and Turkan Haliloglu*

Cite This: https://doi.org/10.1021/acs.jpcb.1c02511

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Deep mutational scanning enables examination of the effects of many mutations at each amino acid position in a query protein, readily disclosing positions that are particularly sensitive. Mutations in these positions alter protein function the most. Here, on the premise that dynamics underlie function, we explore to what extent the measured sensitivity to mutations could be linked to—perhaps be explained by—the structural dynamics of the protein. We employ a minimalist perturbation−response approach based on the Gaussian Network Model (GNM) on a data set of seven proteins with deep mutational scanning data. The analysis shows that the mutation-sensitive positions are often of capacity to modulate the global dynamics and to intermediate allosteric interactions in the structure. With that, upon strain perturbation, these positions decrease residue fluctuations the most, affecting function via entropy changes. This is particularly relevant for positions that are distant from binding sites or other functional regions of the protein and are sensitive to mutations, nevertheless. Our results indicate that mutations in these positions allosterically manipulate protein function.



INTRODUCTION

DNA replication is a highly accurate process that consists of many control and repair mechanisms. Nevertheless, it is imperfect, rendering DNA susceptible to mutations in each replication. In addition, outside factors (chemicals, radiation, etc.) may cause mutations. Nonsynonymous (amino acids changing) mutations in DNA have a wide range of consequences. They may have no effect at all, result in gain/loss of function, or be catastrophic by preventing proper folding.[1−3] Mutations leading to loss of function and/or diseases (loss of fitness) are called deleterious.[4] Their prediction is rather complicated even for proteins of known structure because of the high level of noise in mutational data[5] and because it is difficult to accurately calculate changes in structural stability and dynamics.[6−8]

Until recently, experimental studies regarding the consequences of mutations were limited to the low amount of mutations. This has changed with the emergence of the high-throughput and massively parallel deep sequencing, also known as next generation sequencing (NGS), which revolutionarily impacts genomics.[9,10] It enables sequencing an entire human genome in a single day, whereas it would take almost a decade with traditional Sanger sequencing.[11] Small fragments of DNA sequenced multiple times and in parallel provides highly accurate data on DNA variations. Deep sequencing studies can be used to evaluate mutation effects on different protein functional activities such as binding to ligands, other proteins, or nucleotides, and also to evaluate enzymatic activities like ubiquitination or phosphorylation.[9,10,12] Sequence-to-function mapping provides enormous opportunity to reveal how intrinsic structural and dynamic behavior impact protein function.

There are numerous computational tools to predict the effect of single point mutations, which use various kinds of evolutionary and biophysical data. Mutations in highly conserved positions are often deleterious.[13] Thus, ConSurf calculations may help detecting deleterious mutations.[14] Some methods like Provean[15] and SIFT[16] use amino acid substitution matrixes, whereas others, such as VIPUR,[17] Polyphen-2,[18] SNAP,[19] MuD,[20] and Pro-Maya[21] use biophysical characteristics in addition to evolutionary data. There are also some methods, namely, DEOGEN[22] and SusPect,[23] which also integrate protein−protein interactions data to their prediction algorithms. On the other hand, EVmutation[24] and

Received: March 19, 2021
Revised: April 20, 2021

DeepSequence[25] use sequence covariation data to predict fitness effect of mutations and compare their predictions with deep sequencing data. Envision[26] uses the deep sequencing to predict mutational effects on other proteins in different organisms. Additionally, there are methods that focus on the mutations' energetic effect such as FoldX[27] or rely on molecular dynamics simulations[28,29] and elastic network models[30−34] to study mutations and allosteric interactions in perturbation−response schemes.

The success of dynamics-based approach to rationalize deep sequencing data is most relevant to the current study. Indeed, concepts and computational tools, from two-state conformational models to the dynamical conformational ensembles, have contributed a lot to our understanding of allosteric mechanisms.[13,34−37] For example, allosteric regulation has been explained in terms of the population shift model, and preexisting pathways in the free energy landscape.[38−44] Along these lines, elastic network models consider ensemble of conformations accessible around the native state of the protein, i.e., the global minimum of the free energy landscape, highlighting the importance of global modes in driving allosteric pathways, e.g., between binding sites.[45−48] Global fluctuation changes as a result of variations in the internal dynamics appear as a mechanism of dynamic allostery in the absence of conformational/structural changes.[45,46,49−52]

Here, we explore the ability of a minimalistic perturbation−response approach to reveal mutation-sensitive residues based on their effect on protein dynamics. For this, we conduct Gaussian Network Model (GNM) based analysis of a data set of seven selected proteins and correlate the results with deep mutational scanning data.[53−59] GNM describes the protein structure as an elastic network of Cα atoms connected by strings, assuming harmonic interactions.[60−63] To mimic the effect of a mutation, a perturbation is introduced on each residue by increasing/decreasing the strain in the respective springs of the elastic network, and the response is analyzed with respect to the changes in residue fluctuations.

MATERIALS AND METHODS

Deep Sequencing Data Set. A data set of seven proteins with deep mutational scanning data is assembled; PSD95, CcdB, GAL4 DNA binding region, PAB1-RRM2 domain, ubiquitin, TEM1 β-lactamase, and GTPase H-Ras[53−59] (Table 1). The deep sequencing data are summarized as an average functional cost, which amounts to the average effect of all mutations in an amino acid position on the specific activity (i.e., binding, enzymatic activity, etc.) of each protein in the set. The average functional costs for the seven proteins, used as our gold standard, are plotted in ranked order in Figures S1−S7. A number of mutation-sensitive positions with the highest functional costs are selected for the analysis of each protein in the data set. These are listed in Table 2. This arbitrary selection is complemented by considering different tiers of functional costs.

PSD95-PDZ Domain. Mutation sensitivity of positions are obtained with respect to ligand binding affinity. The ligand CRIPT is quantitatively linked to the expression of enhanced green fluorescent protein (eGFP). eGFP levels are measured and compared to wild type to get the average functional cost of amino acid substitutions.[54]

CcdB. When mutation sensitivity is considered, active/inactive phenotype data are used. WT CcdB shows an active phenotype when cells are killed and an inactive phonotype when cells survive. The survivability of cells is thus a measure of the mutation sensitivity of each position.[53]

GAL4−DNA Binding Domain. GAL4 binding to the DNA activates HIS3 expression, promoting cell survival.[57] Thus, mutation sensitivity is based on the survivability of cells depending on GAL4 binding to DNA.

PAB1-RRM2 Domain. Functional cost is based on the binding affinity of PAB-1 to initiation factor eIF4G.[55]

Ubiquitin. Mutation sensitivity of each position is obtained from the fitness landscape of all possible single mutations of ubiquitin in the presence of dimethyl sulfoxide.[58]

TEM1 β-Lactamase. Mutation-sensitive positions are determined from the fitness landscape of β-lactamase in the presence of ampicillin.[56] A special k^* score is provided, which is inversely proportional to fitness scores of mutations for each position. Position 3 is not available in the crystal structure.

GTPase H-Ras. The activity of the GTPase domain of the H-Ras protein is regulated by GAPs and GEFs. Thus, mutations in GTPase H-Ras positions are evaluated through their effect on bacterial growth in the presence of an antibiotic under GAP and GEF.[59]

GNM and GNM-Based Mode Perturbation Analysis. In the Gaussian Network Model (GNM),[60,61] the equilibrium correlation between fluctuations of two α carbons i and j is given by

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \left(\frac{3k_b T}{\gamma} \right) [\mathbf{\Gamma}^{-1}]_{ij} \tag{1}$$

where $\mathbf{\Gamma}$ is a symmetric matrix known as Kirchhoff (connectivity) matrix, T is the absolute temperature, and k_b is the Boltzmann constant. γ is the force constant of the Hookean pairwise potential that represents the interactions between the residues in the folded structure. The elements of $\mathbf{\Gamma}$ are given by

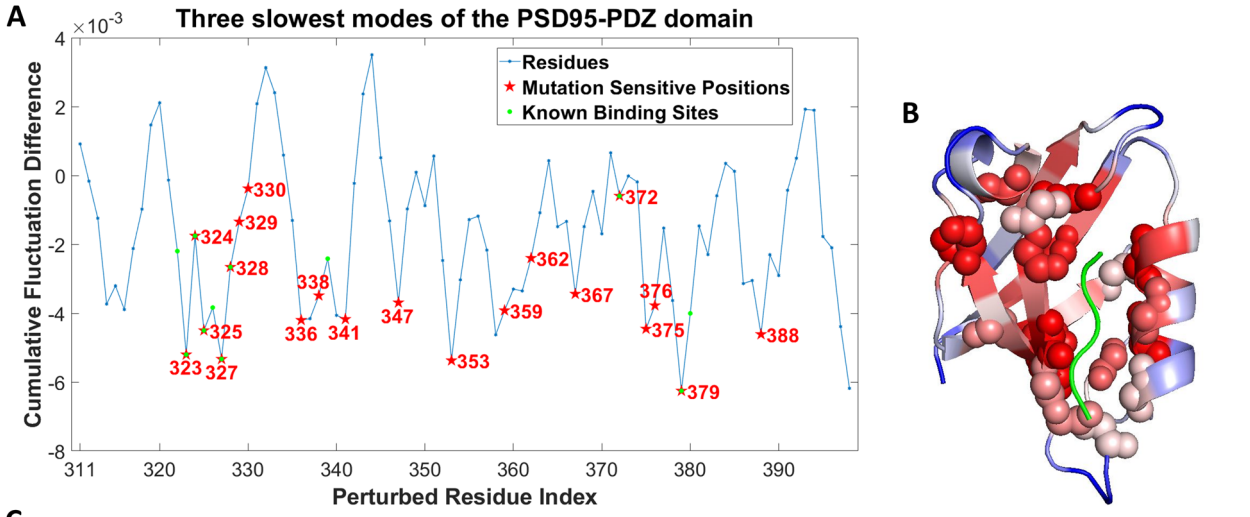
$$\mathbf{\Gamma}_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } \mathbf{R}_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } \mathbf{R}_{ij} > r_c \\ -\sum_{i,i\neq j} \mathbf{\Gamma}_{ij} & \text{if } i = j \end{cases} \tag{2}$$

where $\mathbf{R}_{ij}$ is the distance between the ith and jth C_α atoms and r_c is a distance threshold below which they interact. The inverse of $\mathbf{\Gamma}$ can be decomposed into its eigenvectors $\mathbf{U}$ and

Table 1. DeepSequencing Data Set

structure	crystal structure (PDB ID)	studied criteria	notes
PSD95-PDZ Domain[54]	1BE9	ligand binding	residues from 311 to 393
CcdB[53]	2VUB	toxin activity	
GAL4 - DNA binding region[57]	1D66	nucleotide binding	included the DNA
PAB1-RRM2 domain[55]	4F02	ligand/nucleotide binding growth rate	excluded the RNA Included the IF elF4G
Ubiquitin[58]	1UBQ	growth rate	residues from 1 to 73
TEM1 (β-lactamase)[56]	1XPB	enzyme activity	
H-Ras GTPase[59]	3K8Y	ligand binding/ growth rate	

https://doi.org/10.1021/acs.jpcb.1c02511
J. Phys. Chem. B XXXX, XXX, XXX−XXX



C

Mutation Sensitive Positions	Minimum Positions
323 324 **325** **327** 328 329 330 **336** 338 **341** **347** **353** 359 362 **367** 372 **375** 376 **379** **388**	314 316 **323** **325** **327** **336** **341** **347** **353** 358 361 **367** **375** **379** 382 386 **388** 390

Figure 1. Cumulative residue fluctuation difference profile upon strain perturbation of each position in the PSD95-PDZ domain, calculated using the three slowest modes. (A) Fluctuation differences per position as a function of the amino acid index. Mutation-sensitive positions, detected using the deep sequencing data, are marked as red asterisks, and binding site residues as green asterisks. (Binding site information is obtained from the crystal structure (PDB ID: 1BE9) entry in PDBsum[67]). (B) Mapping of the fluctuation differences on the structure of the domain using the blue-white-red color-code palette. The mutation-sensitive positions are indicated using space-filling models, and the bound peptide as green trace. (C) A list of the mutation-sensitive positions and fluctuation difference minima. Overlapping residues are highlighted in bold and underlined.

the diagonal matrix of eigenvalues λ_i. It is thus possible to express eq 1 as the sum of the contributions from $n-1$ individual modes, where n is the number of residues, as

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \left(\frac{3k_bT}{\gamma}\right)[\mathbf{U}(\mathbf{\Lambda}^{-1})\mathbf{U}^T]_{ij} = \left(\frac{3k_bT}{\gamma}\right)\sum_{k=1}^{n-1}[\lambda_k{}^{-1}\mathbf{u}_k\mathbf{u}_k{}^T]_{ij} \quad (3)$$

where k is the kth vibrational mode.[60,61] The modes are sorted by their associated eigenvalues, such that mode number 1 renders the slowest and most collective motion, and mode $n-1$ the fastest and most local motion. With that, the slowest modes, which are generally linked to allostery,[46] are most relevant here. The minima of the square residue fluctuation profile of the $\mathbf{u}_k$ eigenvector are the hinge residues that coordinate the motion defined by the kth mode. These are the sites where the sense of the correlation between the residue fluctuation changes.[64]

In the GNM-based perturbation analysis, we assume that a perturbation would affect every contact of selected residue; thus, we change the force constant of the harmonic interactions of the perturbed residue in the connectivity matrix of GNM. This perturbation is given as adding a stiffness $(-\epsilon)$ or softening $(+\epsilon)$ for the interactions of each selected residue p, otherwise being "-1" for all interacting residues, to mimic a mutation as

$$\mathbf{\Gamma}^*_{ij,p} = \begin{cases} -1 \pm \varepsilon & \text{if } i = p \neq j \text{ and } \mathbf{R}_{ij} \leq r_c \\ -1 & \text{if } i \neq p,\ i \neq j \text{ and } \mathbf{R}_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } \mathbf{R}_{ij} > r_c \\ -\sum_{i,i\neq j}\mathbf{\Gamma}_{ij} & \text{if } i = j \end{cases} \quad (4)$$

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle^*_p = \left(\frac{3k_bT}{\gamma}\right)[\mathbf{U}^*(\mathbf{\Lambda}^{*-1})\mathbf{U}^{*T}]_{ij,p} = \left(\frac{3k_bT}{\gamma}\right)\sum_{k=1}^{n-1}[\lambda^*_k{}^{-1}\mathbf{u}^*_k u^{*T}_K]_{ij,p} \quad (5)$$

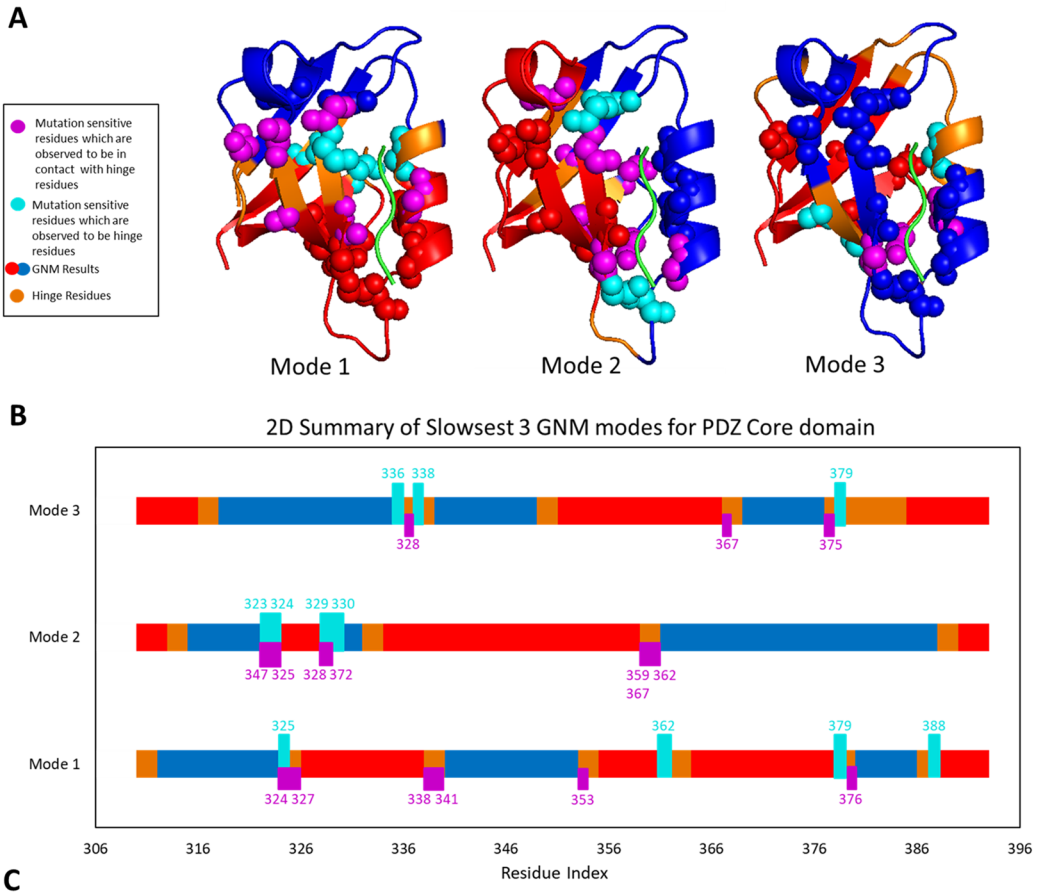
where the conventional distance cutoff of $r_c = 10$ Å is selected and a value of $\epsilon = -0.2$ is arbitrarily chosen.

This perturbation is repeated for each residue p and a new connectivity matrix $\mathbf{\Gamma}^*$ is obtained for each repeat, providing new eigenvectors $\mathbf{u}^*$ and new eigenvalues λ^*. The change in residue fluctuations between perturbed and unperturbed states, i.e., the difference in cumulative residue fluctuations upon perturbation on each residue p, is calculated as

$$\Delta CRF_p = \left(\frac{3k_bT}{\gamma}\right)\sum_{k=1}^{\#\text{ofmodes}-1}\sum_{i=1}^{n}([\lambda^*_k{}^{-1}u^*_k u^{*T}_k]_{ii,p} - [\lambda_k{}^{-1}u_k u_k{}^T]_{ii}) \quad (6)$$

Residue fluctuations are inversely proportional to eigenvalues and are mainly driven by the changes in eigenvalues. Strain perturbation with $\epsilon = -0.2$ shifts the eigenvalues to larger values that in return decrease cumulative residue fluctuations. Changing this value to lower or higher negative values affects

https://doi.org/10.1021/acs.jpcb.1c02511
J. Phys. Chem. B XXXX, XXX, XXX−XXX



C

Mutation Sensitive Residues	MODE-1	MODE-2	MODE-3
323 324 325 327 328 329 330 336 338 341 347 353 359 362 367 372 375 376 379 388	**Hinge:** 325 362 379 388	**Hinge:** 323 324 329 330	**Hinge:** 336 338 379
	Contact: 324 327 338 341 353 376	**Contact:** 325 328 347 359 362 367 372 388	**Contact:** 328 367 375

Figure 2. Mutation-sensitive positions often reside at or near hinges. (A) 3D representation of dynamic domains (red and blue) and hinge residues (orange) of the three slowest GNM modes. Mutation-sensitive positions that overlap (cyan) or are in contact (magenta) with hinges are highlighted. (B) Mapping of the same data on the amino acid sequence of the domain. (C) List of mutation-sensitive positions as overlapped or in contact with hinge residues.

the magnitude of cumulative residue fluctuations. However, it does not affect the profiles, and in particular the location of the minima, which our analysis relies on.[62]

In elastic network models, residue fluctuations are predicted around the native state minimum and these fluctuations mainly contribute to the entropy of the structure. Perturbation modifies residue fluctuations around this minimum. The eigenvalues of perturbed and unperturbed states can thus directly be used to define the vibrational entropy difference between the perturbed (p) and unperturbed (u) structure around equilibrium as inversely proportional to eigenvalues in logarithmic scale,[65,66] approximated as

$$\Delta S_{p-u} \approx \ln\left(\frac{\prod_{k=1}^{n-1} 1/\lambda_{k,p}}{\prod_{k=1}^{n-1} 1/\lambda_{k,u}}\right) \quad (7)$$

Here, in the case of constrained dynamics upon perturbation, stiffening of the structure means a shift in the eigenvalues to higher values (a steeper harmonic well), which in return leads to negative entropy changes. Thus, the difference of residue fluctuations around the native state upon perturbation directly reflects the vibrational entropy change of the system.

When GNM is performed for protein−DNA complexes, the C4′, C2, and P atoms of the nucleic acids are included as Cα atoms.

Statistical Significance Analysis. The statistical significance of the dynamic perturbation analysis is examined. Local/global minima positions are considered when the results of the cumulative amino acid fluctuation differences between perturbed and unperturbed states are analyzed. Minima positions are obtained via the "findpeaks" function of MATLAB with the mean residue fluctuation difference values as a threshold (MATLAB ver. R2018b). Then, the number of position-sensitive residues overlapping with the minima and their mean distances to the minima are calculated. Next, random sampling is used to estimate the statistical significance of the relationship between mutation-sensitive positions and these minima. To this end, the same number of residues as that of the mutation-sensitive positions is selected randomly 10 000 times. *Z*-scores are calculated as

https://doi.org/10.1021/acs.jpcb.1c02511
J. Phys. Chem. B XXXX, XXX, XXX−XXX

$$Z = \frac{(X - \mu)}{\sigma} \tag{8}$$

where X is the mean of the sample (mutation-sensitive positions), μ is the mean, and σ is the standard deviation of the population of random sampling. Then, p-values are calculated using one-tailed t test, where the goal is to estimate the significance of the values in one tail of the distribution (e.g., the right tail in Figure 3A or the left tail in Figure 3B).

RESULTS AND DISCUSSION

PSD95-PDZ Domain. As an exemplary case, the PSD95-PDZ domain is explored in depth using the GNM-based perturbation analysis, as well as conventional GNM.

Global Dynamic Response. The difference in cumulative residue fluctuations upon strain perturbation on each residue calculated for the three slowest, the five slowest, the ten slowest, and all-modes of motion display similar behavior. The results for the three slowest modes are presented in detail in Figure 1, and for the other mode selections in Figure S8. The results clearly demonstrate that individual residues vary in their dynamic effect, and that mutation-sensitive positions often reside at or near local/global minima of the cumulative residue fluctuation difference profile. The minima mark positions with the highest capacity to stiffen the structure among their sequence neighbors upon perturbation. That similar minima locations are obtained using the four different mode selections attests for the stability of the calculations, as well as the importance of these positions in the global dynamics. The correspondence between the mutation-sensitive positions and these key dynamics determinants indicates that mutations in these positions may harm the functional dynamics of the domain. The linkage between sensitivity to mutations and functional dynamics is further demonstrated by the correlation between the average functional cost of all the mutations per amino acid position with the cumulative residue fluctuation difference profiles upon strain perturbation (Figure S9).

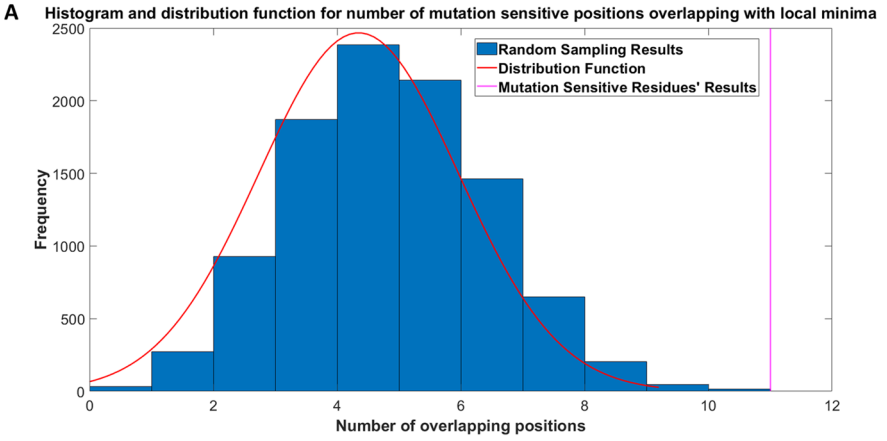
Mapping of the dynamic response (in the three slowest modes) and the mutation-sensitive positions on the 3D structure of the domain in complex with its ligand peptide (Figure 1B) further manifests the correlation. Interestingly, while some of the mutation-sensitive positions directly interact with the ligand, others are far from the binding site. That some of the latter are dynamics determinants may explain their evident sensitivity to mutations.

Strain perturbation mainly shifts the eigenvalues to higher values (Figure S10), which decreases residue fluctuations and the global/vibrational entropy of the structure with higher eigenvalues. Thus, roughly speaking, the peaks in the cumulative eigenvalue difference profile (Figure S10) translate into the dips in the cumulative residue fluctuation difference profile (Figure 1). The overall entropy increase in response to the rigid residue scan of a similar PDZ domain (PDZ2) was previously discussed with reference to Le Châtelier's (mass-action) principle of chemical equilibrium.[27] In this perspective, the present results show that strain perturbation causes some residues to increase their fluctuations, contributing to the softening of the structure, but the net effect in dynamic response is toward stiffening. Functionally unfavorable mutation-sensitive residues lead to the net strongest stiffening response upon strain perturbation. A constraint effect with negative entropy difference was also mainly seen upon ligand binding by the vibrational spectrum analysis of a large set of protein−ligand complexes.[68] Thus, the dynamic response of perturbation also reveals the dynamic capacity of each perturbed residue to change the entropy of the whole structure. Residues with capacity to modulate global dynamics facilitate allosteric interactions by causing other residues to increase/decrease their fluctuations.[69]

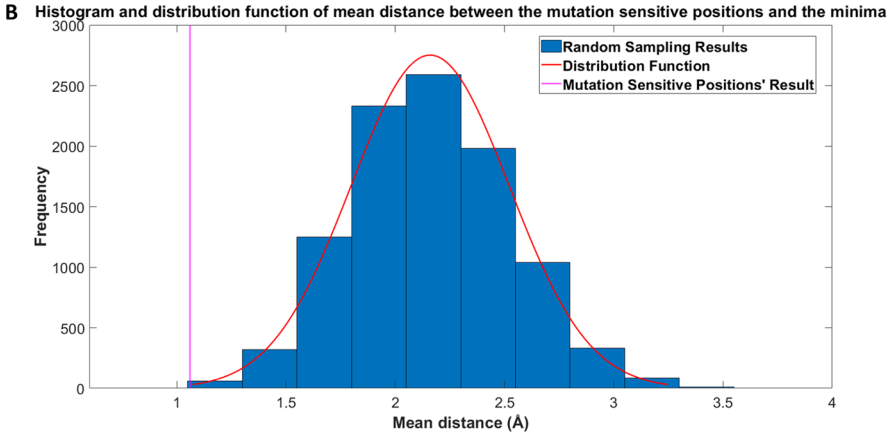
Association with Hinges. To further expand and complement the above analysis, we perform conventional GNM analysis on the PSD95-PDZ domain to observe the relationship between mutation-sensitive positions and the global modes' hinges. Dynamic domains (red and blue) dissected by hinge residues (orange) in each of the three slowest modes and mutation-sensitive positions (spheres) are shown in Figure 2. The hinge residues are at the minima of the square residue fluctuations in the corresponding modes (Figure S11). As seen, mutation-sensitive positions often reside at or right next to hinge residues, which rationalizes their causal effect on global fluctuations and function.[41,45,46,70−77] The hinges are thus mechanically important sites, coupled to the global dynamics of the structure.[78] Indeed, if the dynamic response to strain perturbation is evaluated on the basis of the individual behavior of each of the three slowest modes, the highest dynamic response (Figure S12) is obtained upon strain perturbation in the hinge residues (Figure S11). Because some of the hinges are unique to individual modes, using the average of the three slowest modes, as we do here (Figure 1), requires taking into account also local minima in addition to global minima.

Modulating Global Dynamics Intermediates Allosteric Interactions. Global modes mainly driven by entropic effects are particularly relevant for allosteric regulation.[46] The correspondence between the mutation-sensitive positions and the hinges in the three slowest modes of PSD95-PDZ suggest that they may mediate allosteric pathways linked to the ligand binding site. Indeed, of the mutation-sensitive positions, I336, I341, A347, I353, L367, A375, and I388 are distant from the ligand binding site and nevertheless correspond to minima of the cumulative residue fluctuation difference profile (Figure 1). The residues at minima here were also proposed to have allosteric importance and participate in allosteric networks by many other related studies.[79] A pathway that includes H372, I341, A347, L353, I327, and F325 was identified by anisotropic thermal diffusion model.[80] All these positions, except H372, reside in minima here (Figure 1). H372 is in the ligand binding site and resides at a minimum when more modes of motion are included (Figure S8). Statistical covariation analysis (SCA) suggested an allosteric network from the ligand binding site position H372 via A347, L353, V362, and V386 to position F329.[81] A347, L353, and V386 are at minima here and V362 is right next to the minima at S361. Similarly, the intramolecular allosteric network of F325, F340, I341, V362 and E373, A376, K380, V386, and A390, detected by NMR experiments,[82] include positions that are at minima (F325, F340, and I341) or in close vicinity to minima here (A376, K380, V386, and A390 are immediate neighbors of the minima at L379 and I388). Also, the majority of the positions that comprise the allosteric path identified on the basis of a dynamic response analysis[30] correspond to minima here. Finally, MCPath analysis suggested the four most probable paths starting from the ligand binding residue H372 (H372, F325, L353; H372, I327, I338, L353; H372, T7 (ligand), I327, I338, L353; H372, T7 (ligand), V9 (ligand), F325, A347, L353),[83] surfing through minima positions.

https://doi.org/10.1021/acs.jpcb.1c02511
J. Phys. Chem. B XXXX, XXX, XXX−XXX



Structure	Number of mutation sensitive positions overlapping with local minima	Mean number of random positions overlapping with local minima (random sampling)	Standard Deviation	p-value
PSD95-PDZ Domain	11	4.32	1.61	1.70×10^{-5}



Structure	Mean distance of mutation sensitive positions to local minima	Mean distance of randomly selected positions to local minima	Standard Deviation	p-value
PSD95-PDZ Domain	1.06 Å	2.16 Å	0.362 Å	0.0012

Figure 3. Statistical significance of the correlation between the mutation-sensitive positions of the PSD95-PDZ domain, detected in the deep sequencing experiments, and the dynamics perturbation analysis as reflected in the three slowest modes of motion. (A) Statistical significance of the number of mutation-sensitive positions that overlap with the minima in the cumulative residue fluctuation profiles. The observed overlap (pink) is far into the right tail of the random distribution (blue diagram and red density function). (B) Statistical significance of the mean distance between the mutation-sensitive positions and the minima. The observed average distance (pink) is far into the left tail of the random distribution (blue diagram and red density function).

It is noticeable that while some of the above studies highlight networks of spatially close positions, our analysis detects sparse allosteric pathways that dynamically couple residues that are most often remote from each other in 3D space. Reassuringly, many of these positions contribute to the previously described spatial networks.

Statistical Significance. A statistical analysis is performed to assess the significance of the apparent correlation between mutation-sensitive positions and the local/global minima of the cumulative residue fluctuation difference profiles. To this end, the 20 mutation-sensitive positions with highest functional costs are considered, compared to a null model, comprising randomly selected sets of 20 positions. Here we detail an examination of the correlation between the mutation-sensitive positions and the fluctuation difference profiles obtained for the three slowest modes. The analysis considers both the number of mutation-sensitive positions that overlap with the minima, and the mean distances between the mutation-sensitive positions and the minima. The histograms and distribution functions are given in Figure 3, together with detailed results and *p*-values from the statistical significance analysis. Eleven of the mutation-sensitive positions reside in minima, and the mean distance of the mutation-sensitive positions and the minima are 1.06 Å, with respective *p*-values of 1.7×10^{-5} and 1.2×10^{-3}, attesting for very high statistical significance. Equally significant correlations are observed when

https://doi.org/10.1021/acs.jpcb.1c02511
J. Phys. Chem. B XXXX, XXX, XXX−XXX

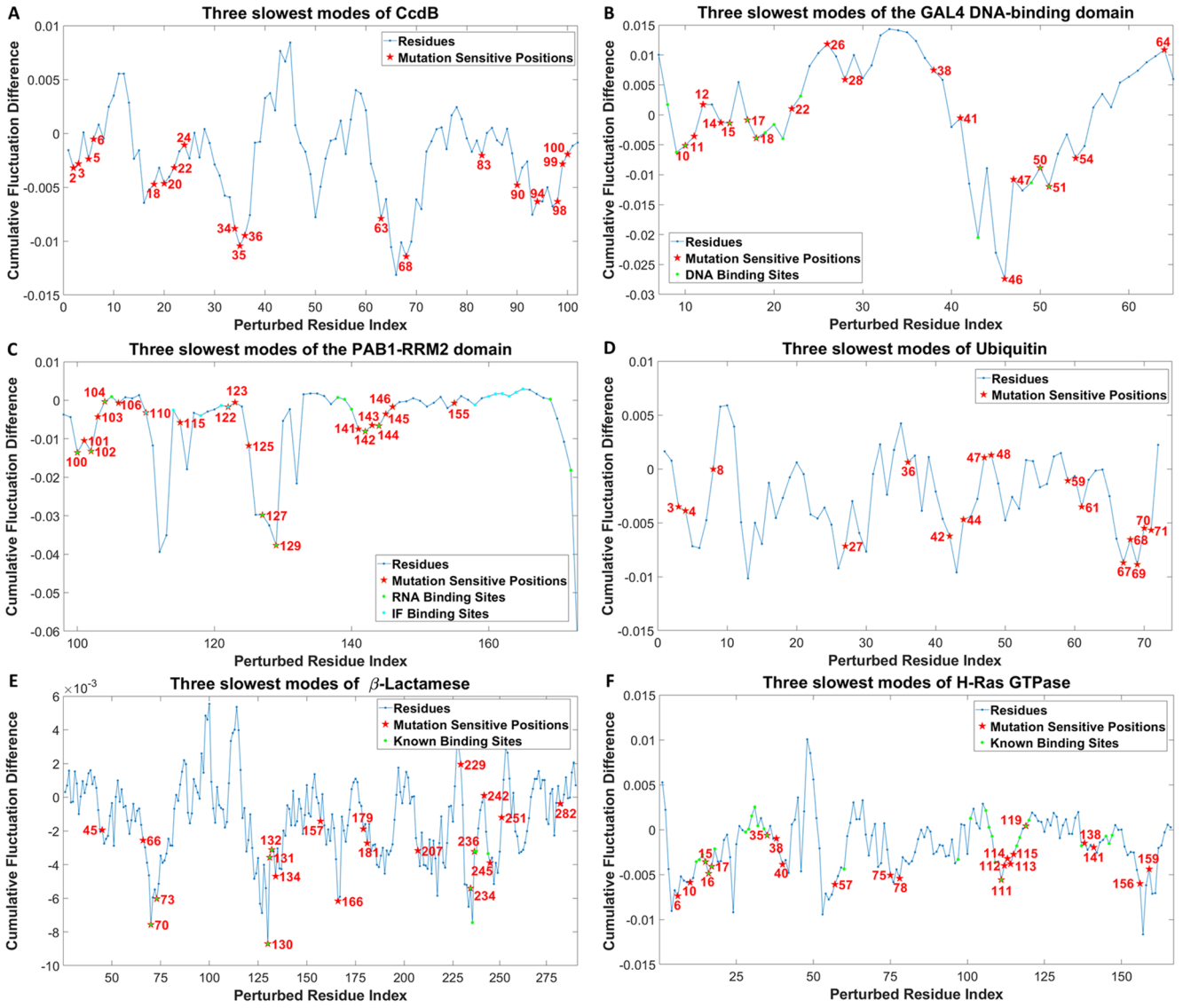


Figure 4. Cumulative residue fluctuation difference profile upon strain perturbation, calculated using the three slowest modes. Mutation-sensitive positions, detected using the deep sequencing data, are marked as red asterisks, and binding site residues as green asterisks. (A) CcdB. (B) GAL4 DNA binding region. (C) PAB1-RRM2 domain. (D) Ubiquitin. (E) TEM1 β-lactamase. (F) H-Ras GTPase. Binding site information is obtained from the related crystal structure entries on PDBsum.[67]

considering the five slowest-, the ten slowest-, and all-modes (Table S1).

Next, we extend the analysis to cover the whole spectrum of mutation-sensitive positions. First, we consider various selections of mutation-sensitive positions, including the 12, the 17, the 27, and the 36 most sensitive positions (Figure S1). Reassuringly, the statistically significant correlation observed between mutation-sensitive positions and the minima in the cumulative residue fluctuation profiles persists (Table S2). Then, we consider four different clusters of residues according to their functional costs: the first 20, the next 16, and the next 20 mutation-sensitive positions, respectively, with high, intermediate, and low functional costs. In addition, we considered the ten mutation-sensitive positions with the most positive effect on fitness. For each group of positions, we examined the correlation with both the minima and maxima of the cumulative residue fluctuation profiles. As we progress through the clusters from the most negative to positive effect on fitness, their correlation with the minima decreases (p-value increases), to a point where the cluster of positions with positive effect on fitness has a negative correlation with the minima (p-value 0.97). On the other hand, their correlations to the maxima are reversed in order. Positions with positive effect on fitness have strong correlation with the maxima (p-value 0.06), and the correlation decreases and becomes negative as we progress toward the cluster of residues with high functional cost (p-value 0.99) (Figure S13). This was attempted by another statistical mechanics-based perturbation approach, which was able to predict the sense of mutations grouped as up and down mutations but with relatively low correlation coefficient of prediction with the fitness values.[33]

Analysis of the Data Set. The GNM-based perturbation analysis is applied to the remaining six proteins in the deep sequencing data set (Table 1), i.e., CcdB, GAL4 DNA binding region, PAB1-RRM2 domain, ubiquitin, TEM1 β-lactamase,

https://doi.org/10.1021/acs.jpcb.1c02511
J. Phys. Chem. B XXXX, XXX, XXX−XXX

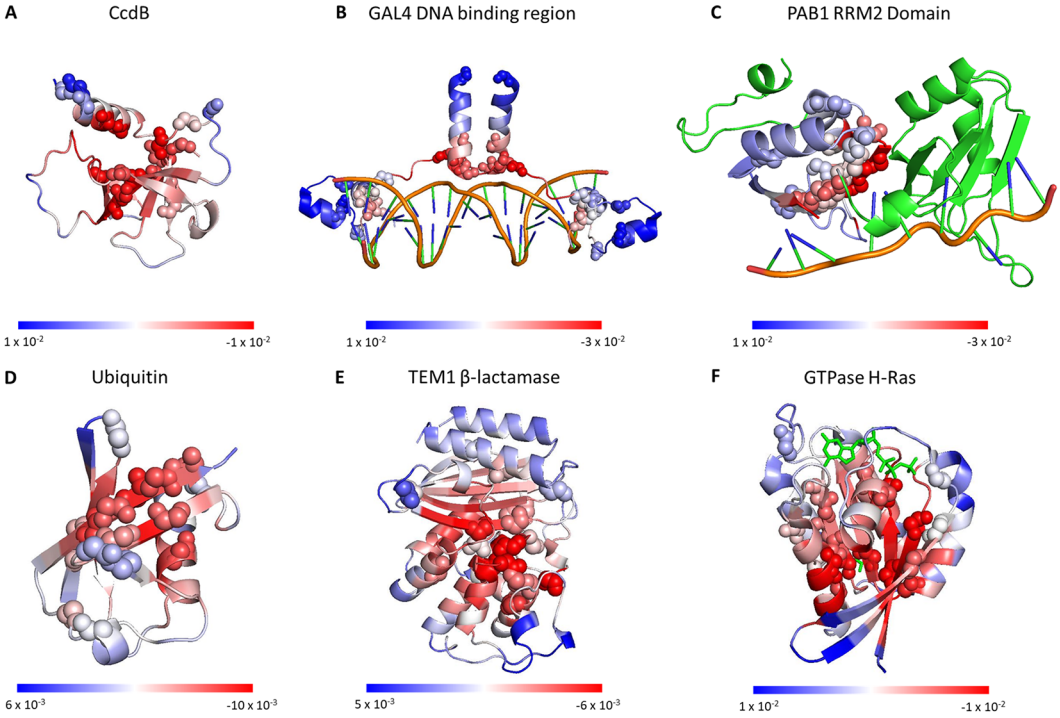


Figure 5. Cumulative residue fluctuation differences, calculated on the basis of the three slowest modes, colorcoded on the structure using the blue-white-red palettes. Mutation-sensitive positions, detected using the deep sequencing data, are indicated using space-filling models. (A) CcdB. (B) GAL4 DNA binding region bound to its DNA ligand. (C) PAB1-RRM2 domain within the context of its biological complex. (D) Ubiquitin. (E) TEM1 β-lactamase. (F) H-Ras GTPase in complex with its substrate.

Table 2. Mutation-Sensitive Amino Acid Positions and Positions Located at the Minima of the Cumulative Residue Fluctuation Profiles in the Three Slowest Modes[a]

structure	mutation-sensitive positions	minimum positions (three slowest modes)
CcdB	**2** 3 **5** 6 18 **20** 22 24 34 **35** 36 **63** **68** 83 **90** 94 98 99 100	**2** **5** 14 16 **20** 25 27 **35** 50 **63** 66 68 71 **90** 93 95 97
GAL4 - DNA binding region	10 11 12 14 **15** 17 **18** 22 26 28 38 41 **46** 47 50 **51** **54** 64	9 **15** **18** 21 40 43 **46** 48 **51** **54**
PAB1-RRM2 domain	**100** 101 **102** 103 104 106 110 115 122 123 125 127 **129** 141 **142** 143 **144** 145 146 155	**100** **102** 112 116 **129** 132 **142** **144**
ubiquitin	3 4 8 27 36 42 44 47 48 59 **61** **67** 68 **69** 70 **71**	6 13 15 17 23 26 30 38 43 50 52 **61** **67** **69** **71**
TEM1 (β-lactamase)	45 66 **70** **73** **130** 131 132 **134** **157** **166** **179** **181** 207 229 234 236 241 244 251 282	44 46 51 59 62 **70** **73** 76 81 90 103 108 119 123 127 **130** **134** 137 139 142 144 148 **157** 160 162 **166** 170 **179** **181** 183 187 190 192 208 212 214 217 221 225 233 235 245 247 249 261 276 279
GTPase H-Ras	**6** **10** 15 **16** 17 35 38 **40** 57 75 **78** **111** 112 113 **114** 115 119 138 141 156 159	4 **6** **10** **16** 19 24 **40** 42 46 53 55 60 68 71 76 **78** 81 89 93 96 100 109 **111** **114** 137 139 142 152 157 160

[a]Overlapping residues are marked as bold and underlined. Here, only exact overlap is considered. Many other positions are in close proximity to the minima.

and H-Ras GTPase. Mutation-sensitive positions are considered based on their functional costs shown in Figures S2−S7.

As with the PSD95-PDZ domain described above, the cumulative fluctuation difference profiles obtained for each structure based on the three slowest, the five slowest, the ten slowest, and all modes of motions are very similar to each other (Figures S14−19). We therefore focus on the results obtained using modes 1-through-3. For completeness we also append the cumulative residue fluctuation difference profiles of the individual modes (Figure S20). Mapping of the cumulative fluctuation difference profiles on the amino acid sequence (Figure 4) and structure (Figure 5) manifests that mutation-sensitive positions are commonly found at, or in the vicinity of, local/global minima (except in ubiquitin). The list of mutation-sensitive positions and the minima of the cumulative residue fluctuation difference profiles of the six proteins in the data set are given in Table 2. More detailed description and implications of the analysis for each case is provided in the Supporting Information.

Statistical Significance. To examine the significance of the correlation between mutation-sensitive positions and the minima of cumulative residue fluctuation profiles of the perturbation analysis, we applied the statistical analysis described above for the PSD95-PDZ domain also for the rest of the proteins in our data set. *p*-values for the correlation between the number of mutation-sensitive positions, deduced from the deep sequencing data, which overlap with minima in the cumulative residue fluctuation difference profile are shown

https://doi.org/10.1021/acs.jpcb.1c02511
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 3. Statistical Significance Analysis Results for the Correlation between the Number of Mutation-Sensitive Positions, Deduced from the Deep Sequencing Data, That Overlap with Minima in the Cumulative Residue Fluctuation Difference Profile, Calculated Here, Based on the Three Slowest Modes

structure	number of mutation-sensitive positions overlapping with local minima	mean number of random positions overlapping with local minima (random sampling)	standard deviation	*p*-value
PSD95-PDZ domain	11	4.32	1.61	1.70×10^{-5}
CcdB	7	3.19	1.47	0.005
GAL4−DNA binding region	5	3.13	1.34	0.081
PAB1-RRM2 domain	5	2.15	1.20	0.009
ubiquitin	4	3.15	1.42	0.274
TEM1 β-lactamase	7	3.57	1.63	4.27×10^{-4}
GTPase H-Ras	7	3.80	1.66	0.027

in Table 3, and *p*-values for the mean distance between these positions and the minima are listed in Table 4. Except for ubiquitin, the results show highly significant correlations. Furthermore, we extend the analysis and consider various selections of mutation-sensitive positions depending on the tiers of their functional costs (Table S3). The statistically significant correlation observed between mutation-sensitive positions and the minima in the cumulative residue fluctuation profiles is retained.

Table 4. Statistical Significance of the Mean Distance between the Mutation-Sensitive Positions, Deduced from the Deep Sequencing Data, and the Minima in the Cumulative Residue Fluctuation Difference Profile, Calculated Here, Based on the Three Slowest Modes

structure	mean distance (Å) of mutation-sensitive positions to local minima	mean distance (Å) of randomly selected positions to local minima	standard deviation (Å)	*p*-value
PSD95-PDZ domain	1.06 Å	2.16 Å	0.362 Å	0.0012
CcdB	1.39 Å	3.14 Å	0.562 Å	9.41×10^{-4}
GAL4−DNA binding region	2.17 Å	3.48 Å	0.703 Å	0.0305
PAB1-RRM2 domain	2.20 Å	3.87 Å	0.481 Å	2.60×10^{-4}
ubiquitin	2.11 Å	2.61 Å	0.524 Å	0.1730
TEM1 β-lactamase	1.52 Å	2.79 Å	0.518 Å	0.0071
GTPase H-Ras	1.63 Å	2.94 Å	0.560 Å	0.0098

Ubiquitin is an outlier, with weaker, and less significant, correlation between mutation-sensitive positions and the minima of cumulative residue fluctuation difference profiles. This could be due to the low-resolution nature and the absence of specific interactions (acidic, basic, polar, and nonpolar) in our model. To partially account for side chain specific interactions, we incorporate side chains into the GNM analysis of ubiquitin. The importance of lysine residues (especially K48 and K63) and side chain interactions for ubiquitin is well-known for ubiquitin.[84,85] When we perform the GNM-based perturbation analysis based on the Cβ atoms or side chain centroids instead of the Cα atoms (Figure S21), more mutation-sensitive positions reside at the minima of the cumulative residue fluctuation profile. This is reflected in lower *p*-values, which decreased from 0.1730 (for the Cα-based model) to 0.0285 and 0.0362 for the correlation of the mutation-sensitive positions with the minima when representing the residues on the basis of only their Cβ atoms and only side chain centroids, respectively. The corresponding mean distances of the mutation-sensitive positions to the minima decrease from 2.11 to 0.86 Å (for both Cβ and side chain centroids). Reassuringly, the *p*-values calculated using the same procedure for the remaining proteins in the deep sequencing data set were similar to the values obtained using the Cα-based model (Table S4).

CONCLUSION

Perturbations on mutation-sensitive positions here do not change the native state of the protein but rather modify equilibrium residue fluctuations, i.e., the features of dynamic modes, around this state. The shift in the eigenvalues leads to the most pronounced effect on the global fluctuations of the structure, facilitating allosteric interactions of functional importance. For this reason, even the highly approximate GNM description of the dynamics was sufficient to reveal many of the mutation-sensitive positions that emerge from the deep sequencing data.

Bearing that in mind, it is easy to rationalize that mutation-sensitive positions, observed in the deep mutation data, reside in (or near) global minima of the cumulative residue fluctuation difference profile. However, that some of the mutation-sensitive positions correspond to local minima of this profile might appear puzzling at first. This less trivial result is because the cumulative residue fluctuation difference profile is an average; a certain position could be of high mechanistic importance in one mode but not in another. Thus, a hinge position that repeats in more than one mode or belongs to the slowest mode is likely to appear as a global minimum with the stronger response upon perturbation. On the other hand, a hinge that belongs to a specific mode (other than the slowest) may appear as a local minimum that leads to a relatively weak response upon perturbation in the average behavior. It does not make the local minimum less important as the particular mode where this position is a global minimum could be critical for a specific function. In this respect, the local and global minima in the cumulative residue fluctuation difference profile are equally important within the context of this study.

Overall, the correspondence between local/global minima of the cumulative residue fluctuation difference profile and mutation-sensitive amino acid positions is yet another manifestation of the key role of allostery in protein function.

https://doi.org/10.1021/acs.jpcb.1c02511
J. Phys. Chem. B XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c02511.

Plots of the average functional cost of all the mutations per amino acid position, cumulative residue fluctuation difference profiles, cumulative eigenvalue difference, squared fluctuations of residues, statistical significance of the correlation between various groups of mutation-sensitive positions, tables of statistical significance of the correlation between the mutation-sensitive positions and of *p*-values, and text describing mutation-sensitive sites disclose intrinsic functional dynamic behavior (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Turkan Haliloglu − *Department of Chemical Engineering and Polymer Research Center, Bogazici University, Bebek, Istanbul 34342, Turkey;* orcid.org/0000-0002-1279-5803; Email: halilogt@boun.edu.tr

Authors

Yiğit Kutlu − *Department of Chemical Engineering and Polymer Research Center, Bogazici University, Bebek, Istanbul 34342, Turkey;* orcid.org/0000-0001-5917-2550

Nir Ben-Tal − *Department of Biochemistry and Molecular Biology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 6997801, Israel*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c02511

Notes

The authors declare no competing financial interest.

MATLAB codes for the GNM based mode perturbation method and *p*-value analysis are shared together with used PDB files on the following link https://github.com/yigitkutlu/GNM-Perturbation.

■ ACKNOWLEDGMENTS

Over the years, Prof. Ruth Nussinov has made continuous and very insightful contributions to our understanding of the interplay between dynamical conformational ensembles and allostery. These have inspired us over the years and in this study. This work was supported by grants from NATO Science for Peace and Security Program (SPS project G5568) to N.B.T. and T.H., The Scientific and Technological Research Council of Turkey (TUBITAK project 119F392), and Bogazici University Research Fund (BAP project 13460) to Y.K. and T.H. N.B.T.'s research is supported in part by the Abraham E. Kazan Chair in Structural Biology, Tel Aviv University.

■ REFERENCES

(1) Yue, P.; Li, Z.; Moult, J. Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease. *J. Mol. Biol.* **2005**, *353* (2), 459−473.

(2) Studer, R. A.; Dessailly, B. H.; Orengo, C. A. Residue Mutations and Their Impact on Protein Structure and Function: Detecting Beneficial and Pathogenic Changes. *Biochem. J.* **2013**, *449* (3), 581−594.

(3) Gao, M.; Zhou, H.; Skolnick, J. Insights into Disease-Associated Mutations in the Human Proteome through Protein Structural Analysis. *Structure* **2015**, *23* (7), 1362−1369.

(4) Nesse, R. M.; Bergstrom, C. T.; Ellison, P. T.; Flier, J. S.; Gluckman, P.; Govindaraju, D. R.; Niethammer, D.; Omenn, G. S.; Perlman, R. L.; Schwartz, M. D.; et al. Evolution in Health and Medicine Sackler Colloquium: Making Evolutionary Biology a Basic Science for Medicine. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (suppl_1), 1800−1807.

(5) Montanucci, L.; Martelli, P. L.; Ben-Tal, N.; Fariselli, P. A Natural Upper Bound to the Accuracy of Predicting Protein Stability Changes upon Mutations. *Bioinformatics* **2019**, *35* (9), 1513−1517.

(6) Rodrigues, C. H.; Pires, D. E.; Ascher, D. B. DynaMut: Predicting the Impact of Mutations on Protein Conformation, Flexibility and Stability. *Nucleic Acids Res.* **2018**, *46* (W1), W350−W355.

(7) Sanavia, T.; Birolo, G.; Montanucci, L.; Turina, P.; Capriotti, E.; Fariselli, P. Limitations and Challenges in Protein Stability Prediction upon Genome Variations: Towards Future Applications in Precision Medicine. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1968−1979.

(8) Pandurangan, A. P.; Blundell, T. L. Prediction of Impacts of Mutations on Protein Structure and Interactions: SDM, a Statistical Approach, and MCSM, Using Machine Learning. *Protein Sci.* **2020**, *29* (1), 247−257.

(9) Araya, C. L.; Fowler, D. M. Deep Mutational Scanning: Assessing Protein Function on a Massive Scale. *Trends Biotechnol.* **2011**, *29* (9), 435−442.

(10) Fowler, D. M.; Fields, S. Deep Mutational Scanning: A New Style of Protein Science. *Nat. Methods* **2014**, *11* (8), 801−807.

(11) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (12), 5463−5467.

(12) Goldman, D.; Domschke, K. Making Sense of Deep Sequencing. *Int. J. Neuropsychopharmacol.* **2014**, *17* (10), 1717−1725.

(13) Kessel, A.; Ben-Tal, N. *Introduction to Proteins: Structure, Function, and Motion*, 2nd ed.; CRC Press: London, England, 2018.

(14) Celniker, G.; Nimrod, G.; Ashkenazy, H.; Glaser, F.; Martz, E.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr. J. Chem.* **2013**, *53* (3−4), 199−206.

(15) Choi, Y.; Chan, A. P. PROVEAN Web Server: A Tool to Predict the Functional Effect of Amino Acid Substitutions and Indels. *Bioinformatics* **2015**, *31* (16), 2745−2747.

(16) Ng, P. C.; Henikoff, S. SIFT: Predicting Amino Acid Changes That Affect Protein Function. *Nucleic Acids Res.* **2003**, *31* (13), 3812−3814.

(17) Baugh, E. H.; Simmons-Edler, R.; Müller, C. L.; Alford, R. F.; Volfovsky, N.; Lash, A. E.; Bonneau, R. Robust Classification of Protein Variation Using Structural Modelling and Large-Scale Data Integration. *Nucleic Acids Res.* **2016**, *44* (6), 2501−2513.

(18) Adzhubei, I. A.; Schmidt, S.; Peshkin, L.; Ramensky, V. E.; Gerasimova, A.; Bork, P.; Kondrashov, A. S.; Sunyaev, S. R. A Method and Server for Predicting Damaging Missense Mutations. *Nat. Methods* **2010**, *7* (4), 248−249.

(19) Bromberg, Y.; Rost, B. SNAP: Predict Effect of Non-Synonymous Polymorphisms on Function. *Nucleic Acids Res.* **2007**, *35* (11), 3823−3835.

(20) Wainreb, G.; Ashkenazy, H.; Bromberg, Y.; Starovolsky-Shitrit, A.; Haliloglu, T.; Ruppin, E.; Avraham, K. B.; Rost, B.; Ben-Tal, N. MuD: An Interactive Web Server for the Prediction of Non-Neutral Substitutions Using Protein Structural Data. *Nucleic Acids Res.* **2010**, *38*, W523−8.

(21) Wainreb, G.; Wolf, L.; Ashkenazy, H.; Dehouck, Y.; Ben-Tal, N. Protein Stability: A Single Recorded Mutation Aids in Predicting the Effects of Other Mutations in the Same Amino Acid Site. *Bioinformatics* **2011**, *27* (23), 3286−3292.

(22) Raimondi, D.; Gazzo, A. M.; Rooman, M.; Lenaerts, T.; Vranken, W. F. Multilevel Biological Characterization of Exomic Variants at the Protein Level Significantly Improves the Identification of Their Deleterious Effects. *Bioinformatics* **2016**, *32* (12), 1797−1804.

(23) Yates, C. M.; Filippis, I.; Kelley, L. A.; Sternberg, M. J. E. SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features. *J. Mol. Biol.* **2014**, *426* (14), 2692−2701.
(24) Hopf, T. A.; Ingraham, J. B.; Poelwijk, F. J.; Schärfe, C. P. I.; Springer, M.; Sander, C.; Marks, D. S. Mutation Effects Predicted from Sequence Co-Variation. *Nat. Biotechnol.* **2017**, *35* (2), 128−135.
(25) Riesselman, A. J.; Ingraham, J. B.; Marks, D. S. Deep Generative Models of Genetic Variation Capture the Effects of Mutations. *Nat. Methods* **2018**, *15* (10), 816−822.
(26) Gray, V. E.; Hause, R. J.; Luebeck, J.; Shendure, J.; Fowler, D. M. Quantitative Missense Variant Effect Prediction Using Large-Scale Mutagenesis Data. *Cell Syst.* **2018**, *6* (1), 116−124.
(27) Kalescky, R.; Zhou, H.; Liu, J.; Tao, P. Rigid Residue Scan Simulations Systematically Reveal Residue Entropic Roles in Protein Allostery. *PLoS Comput. Biol.* **2016**, *12* (4), No. e1004893.
(28) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX Web Server: An Online Force Field. *Nucleic Acids Res.* **2005**, *33*, W382.
(29) Ettayapuram Ramaprasad, A. S.; Uddin, S.; Casas-Finet, J.; Jacobs, D. J. Decomposing Dynamical Couplings in Mutated ScFv Antibody Fragments into Stabilizing and Destabilizing Effects. *J. Am. Chem. Soc.* **2017**, *139* (48), 17508−17517.
(30) Gerek, Z. N.; Ozkan, S. B. Change in Allosteric Network Affects Binding Affinities of PDZ Domains: Analysis through Perturbation Response Scanning. *PLoS Comput. Biol.* **2011**, *7* (10), No. e1002154.
(31) Greener, J. G.; Sternberg, M. J. E. AlloPred: Prediction of Allosteric Pockets on Proteins Using Normal Mode Perturbation Analysis. *BMC Bioinf.* **2015**, *16* (1), 335.
(32) Modi, T.; Ozkan, S. B. Mutations Utilize Dynamic Allostery to Confer Resistance in TEM-1 β-Lactamase. *Int. J. Mol. Sci.* **2018**, *19* (12), 3808.
(33) Guarnera, E.; Berezovsky, I. N. Toward Comprehensive Allosteric Control over Protein Activity. *Structure* **2019**, *27* (5), 866−878.
(34) Campitelli, P.; Modi, T.; Kumar, S.; Ozkan, S. B. The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biophys.* **2020**, *49* (1), 267−288.
(35) Liu, J.; Nussinov, R. Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLoS Comput. Biol.* **2016**, *12* (6), No. e1004966.
(36) Dokholyan, N. V. Controlling Allosteric Networks in Proteins. *Chem. Rev.* **2016**, *116* (11), 6463−6487.
(37) Wodak, S. J.; Paci, E.; Dokholyan, N. V.; Berezovsky, I. N.; Horovitz, A.; Li, J.; Hilser, V. J.; Bahar, I.; Karanicolas, J.; Stock, G.; et al. Allostery in Its Many Disguises: From Theory to Applications. *Structure* **2019**, *27* (4), 566−578.
(38) Gunasekaran, K.; Ma, B.; Nussinov, R. Is Allostery an Intrinsic Property of All Dynamic Proteins? *Proteins: Struct., Funct., Genet.* **2004**, *57* (3), 433−443.
(39) Tsai, C.-J.; del Sol, A.; Nussinov, R. Allostery: Absence of a Change in Shape Does Not Imply That Allostery Is Not at Play. *J. Mol. Biol.* **2008**, *378* (1), 1−11.
(40) del Sol, A.; Tsai, C.-J.; Ma, B.; Nussinov, R. The Origin of Allosteric Functional Modulation: Multiple Pre-Existing Pathways. *Structure* **2009**, *17* (8), 1042−1050.
(41) Ma, B.; Tsai, C.-J.; Haliloğlu, T.; Nussinov, R. Dynamic Allostery: Linkers Are Not Merely Flexible. *Structure* **2011**, *19* (7), 907−917.
(42) Nussinov, R.; Tsai, C.-J. Allostery in Disease and in Drug Discovery. *Cell* **2013**, *153* (2), 293−305.
(43) Tsai, C.-J.; Nussinov, R. A Unified View of "How Allostery Works. *PLoS Comput. Biol.* **2014**, *10* (2), No. e1003394.
(44) Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **2016**, *116* (11), 6391−6423.
(45) Liu, Y.; Bahar, I. Sequence Evolution Correlates with Structural Dynamics. *Mol. Biol. Evol.* **2012**, *29* (9), 2253−2263.
(46) Haliloglu, T.; Bahar, I. Adaptability of Protein Structures to Enable Functional Interactions and Evolutionary Implications. *Curr. Opin. Struct. Biol.* **2015**, *35*, 17−23.
(47) Townsend, P. D.; Rodgers, T. L.; Pohl, E.; Wilson, M. R.; McLeish, T. C. B.; Cann, M. J. Global Low-Frequency Motions in Protein Allostery: CAP as a Model System. *Biophys. Rev.* **2015**, *7* (2), 175−182.
(48) Guarnera, E.; Berezovsky, I. N. Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. *PLoS Comput. Biol.* **2016**, *12* (3), No. e1004678.
(49) Lee, A. L. Contrasting Roles of Dynamics in Protein Allostery: NMR and Structural Studies of CheY and the Third PDZ Domain from PSD-95. *Biophys. Rev.* **2015**, *7* (2), 217−226.
(50) McLeish, T. C. B.; Cann, M. J.; Rodgers, T. L. Dynamic Transmission of Protein Allostery without Structural Change: Spatial Pathways or Global Modes? *Biophys. J.* **2015**, *109* (6), 1240−1250.
(51) Townsend, P. D.; Rodgers, T. L.; Glover, L. C.; Korhonen, H. J.; Richards, S. A.; Colwell, L. J.; Pohl, E.; Wilson, M. R.; Hodgson, D. R. W.; McLeish, T. C. B.; et al. The Role of Protein-Ligand Contacts in Allosteric Regulation of the Escherichia Coli Catabolite Activator Protein. *J. Biol. Chem.* **2015**, *290* (36), 22225−22235.
(52) Law, A. B.; Sapienza, P. J.; Zhang, J.; Zuo, X.; Petit, C. M. Native State Volume Fluctuations in Proteins as a Mechanism for Dynamic Allostery. *J. Am. Chem. Soc.* **2017**, *139* (10), 3599−3602.
(53) Adkar, B. V.; Tripathi, A.; Sahoo, A.; Bajaj, K.; Goswami, D.; Chakrabarti, P.; Swarnkar, M. K.; Gokhale, R. S.; Varadarajan, R. Protein Model Discrimination Using Mutational Sensitivity Derived from Deep Sequencing. *Structure* **2012**, *20* (2), 371−381.
(54) McLaughlin, R. N., Jr; Poelwijk, F. J.; Raman, A.; Gosal, W. S.; Ranganathan, R. The Spatial Architecture of Protein Function and Adaptation. *Nature* **2012**, *491* (7422), 138−142.
(55) Melamed, D.; Young, D. L.; Gamble, C. E.; Miller, C. R.; Fields, S. Deep Mutational Scanning of an RRM Domain of the Saccharomyces Cerevisiae Poly(A)-Binding Protein. *RNA* **2013**, *19* (11), 1537−1551.
(56) Firnberg, E.; Labonte, J. W.; Gray, J. J.; Ostermeier, M. A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape. *Mol. Biol. Evol.* **2014**, *31* (6), 1581−1592.
(57) Kitzman, J. O.; Starita, L. M.; Lo, R. S.; Fields, S.; Shendure, J. Massively Parallel Single-Amino-Acid Mutagenesis. *Nat. Methods* **2015**, *12* (3), 203−206.
(58) Mavor, D.; Barlow, K.; Thompson, S.; Barad, B. A.; Bonny, A. R.; Cario, C. L.; Gaskins, G.; Liu, Z.; Deming, L.; Axen, S. D. Determination of Ubiquitin Fitness Landscapes under Different Chemical Stresses in a Classroom Setting. *eLife* **2016**, *5*, e15802.
(59) Bandaru, P.; Shah, N. H.; Bhattacharyya, M.; Barton, J. P.; Kondo, Y.; Cofsky, J. C.; Gee, C. L.; Chakraborty, A. K.; Kortemme, T.; Ranganathan, R. Deconstruction of the Ras Switching Cycle through Saturation Mutagenesis. *eLife* **2017**, *6*, e27810.
(60) Bahar, I.; Atilgan, A. R.; Erman, B. Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential. *Folding Des.* **1997**, *2* (3), 173−181.
(61) Haliloglu, T.; Bahar, I.; Erman, B. Gaussian Dynamics of Folded Proteins. *Phys. Rev. Lett.* **1997**, *79* (16), 3090−3093.
(62) Yogurtcu, O. N.; Gur, M.; Erman, B. Statistical Thermodynamics of Residue Fluctuations in Native Proteins. *J. Chem. Phys.* **2009**, *130* (9), 095103.
(63) Hacisuleyman, A.; Erkip, A.; Erman, B.; Erman, B. Synchronous and Asynchronous Response in Dynamically Perturbed Proteins. *J. Phys. Chem. B* **2021**, *125* (3), 729−739.
(64) Emekli, U.; Schneidman-Duhovny, D.; Wolfson, H. J.; Nussinov, R.; Haliloglu, T. HingeProt: Automated Prediction of Hinges in Protein Structures. *Proteins: Struct., Funct., Genet.* **2008**, *70* (4), 1219−1227.
(65) Karplus, M.; Kushick, J. N. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* **1981**, *14* (2), 325−332.

https://doi.org/10.1021/acs.jpcb.1c02511
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(66) Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability. *Phys. Rev. Lett.* **1998**, *80* (12), 2733−2736.
(67) Laskowski, R. A.; Jabłońska, J.; Pravda, L.; Vařeková, R. S.; Thornton, J. M. PDBsum: Structural Summaries of PDB Entries. *Protein Sci.* **2018**, *27* (1), 129−134.
(68) Kaynak, B. T.; Doruker, P. Protein-Ligand Complexes as Constrained Dynamical Systems. *J. Chem. Inf. Model.* **2019**, *59* (5), 2352−2358.
(69) Tee, W.-V.; Guarnera, E.; Berezovsky, I. N. Disorder Driven Allosteric Control of Protein Activity. *Current Research in Structural Biology* **2020**, *2*, 191−203.
(70) De Los Rios, P.; Cecconi, F.; Pretre, A.; Dietler, G.; Michielin, O.; Piazza, F.; Juanico, B. Functional Dynamics of PDZ Binding Domains: A Normal-Mode Analysis. *Biophys. J.* **2005**, *89* (1), 14−21.
(71) Zheng, W.; Tekpinar, M. Large-Scale Evaluation of Dynamically Important Residues in Proteins Predicted by the Perturbation Analysis of a Coarse-Grained Elastic Model. *BMC Struct. Biol.* **2009**, *9* (1), 45.
(72) Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. Global Dynamics of Proteins: Bridging between Structure and Function. *Annu. Rev. Biophys.* **2010**, *39* (1), 23−42.
(73) Marcos, E.; Crehuet, R.; Bahar, I. On the Conservation of the Slow Conformational Dynamics within the Amino Acid Kinase Family: NAGK the Paradigm. *PLoS Comput. Biol.* **2010**, *6* (4), No. e1000738.
(74) AykaçFas, B.; Tutar, Y.; Haliloğlu, T. Dynamic Fluctuations Provide the Basis of a Conformational Switch Mechanism in Apo Cyclic AMP Receptor Protein. *PLoS Comput. Biol.* **2013**, *9* (7), No. e1003141.
(75) Sumbul, F.; Acuner-Ozbabacan, S. E.; Haliloglu, T. Allosteric Dynamic Control of Binding. *Biophys. J.* **2015**, *109* (6), 1190−1201.
(76) Sayılgan, J. F.; Haliloğlu, T.; Gönen, M. Protein Dynamics Analysis Reveals That Missense Mutations in Cancer-Related Genes Appear Frequently on Hinge-Neighboring Residues. *Proteins: Struct., Funct., Genet.* **2019**, *87* (6), 512−519.
(77) Acar, B.; Rose, J.; Aykac Fas, B.; Ben-Tal, N.; Lewinson, O.; Haliloglu, T. Distinct Allosteric Networks Underlie Mechanistic Speciation of ABC Transporters. *Structure* **2020**, *28* (6), 651−663.
(78) Yang, L.-W.; Bahar, I. Coupling between Catalytic Site and Collective Dynamics: A Requirement for Mechanochemical Activity of Enzymes. *Structure* **2005**, *13* (6), 893−904.
(79) Gautier, C.; Laursen, L.; Jemth, P.; Gianni, S. Seeking Allosteric Networks in PDZ Domains. *Protein Eng., Des. Sel.* **2018**, *31* (10), 367−373.
(80) Ota, N.; Agard, D. A. Intramolecular Signaling Pathways Revealed by Modeling Anisotropic Thermal Diffusion. *J. Mol. Biol.* **2005**, *351* (2), 345−354.
(81) Lockless, S. W.; Ranganathan, R. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science* **1999**, *286* (5438), 295−299.
(82) Gianni, S.; Walma, T.; Arcovito, A.; Calosci, N.; Bellelli, A.; Engström, A.; Travaglini-Allocatelli, C.; Brunori, M.; Jemth, P.; Vuister, G. W. Demonstration of Long-Range Interactions in a PDZ Domain by NMR, Kinetics, and Protein Engineering. *Structure* **2006**, *14* (12), 1801−1809.
(83) Kaya, C.; Armutlulu, A.; Ekesan, S.; Haliloglu, T. MCPath: Monte Carlo Path Generation Approach to Predict Likely Allosteric Pathways and Functional Residues. *Nucleic Acids Res.* **2013**, *41*, W249−55.
(84) Pickart, C. M.; Eddins, M. J. Ubiquitin: Structures, Functions, Mechanisms. *Biochim. Biophys. Acta, Mol. Cell Res.* **2004**, *1695* (1−3), 55−72.
(85) Komander, D.; Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **2012**, *81* (1), 203−229.

https://doi.org/10.1021/acs.jpcb.1c02511
J. Phys. Chem. B XXXX, XXX, XXX−XXX