

Relating protein conformational changes to packing efficiency and disorder

Nitin Bhardwaj¹ and Mark Gerstein^{1,2,3*}

¹Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06520

²Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520

³Department of Computer Science, Yale University, New Haven, Connecticut 06520

Received 10 August 2008; Revised 23 March 2009; Accepted 30 March 2009

DOI: 10.1002/pro.132

Published online 16 April 2009 proteinscience.org

Abstract: Changes in protein conformation play key roles in facilitating various biochemical processes, ranging from signaling and phosphorylation to transport and catalysis. While various factors that drive these motions such as environmental changes and binding of small molecules are well understood, specific causative effects on the structural features of the protein due to these conformational changes have not been studied on a large scale. Here, we study protein conformational changes in relation to two key structural metrics: packing efficiency and disorder. Packing has been shown to be crucial for protein stability and function by many protein design and engineering studies. We study changes in packing efficiency during conformational changes, thus extending the analysis from a static context to a dynamic perspective and report some interesting observations. First, we study various proteins that adopt alternate conformations and find that tendencies to show motion and change in packing efficiency are correlated: residues that change their packing efficiency show larger motions. Second, our results suggest that residues that show higher changes in packing during motion are located on the *changing interfaces* which are formed during these conformational changes. These changing interfaces are slightly different from shear or static interfaces that have been analyzed in previous studies. Third, analysis of packing efficiency changes in the context of secondary structure shows that, as expected, residues buried in helices show the least change in packing efficiency, whereas those embedded in bends are most likely to change packing. Finally, by relating protein disorder to motions, we show that *marginally disordered* residues which are ordered enough to be crystallized but have sequence patterns indicative of disorder show higher dislocation and a higher change in packing than ordered ones and are located mostly on the changing interfaces. Overall, our results demonstrate that between the two conformations, the cores of the proteins remain mostly intact, whereas the interfaces display the most elasticity, both in terms of disorder and change in packing efficiency. By doing a variety of tests, we also show that our observations are robust to the solvation state of the proteins.

Keywords: protein structure; packing efficiency; molecular motions; conformation changes; disorder; changing interfaces; protein cores

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Mark Gerstein, NB: 266 Whitney Ave., Room 426, New Haven, CT, 06511 or MG: 266 Whitney Ave., Room 432, New Haven, CT, 06511. E-mail: mark.gerstein@yale.edu

Introduction

Protein structures are not static; many of the bonds in proteins can rotate and flex, and structural segments of the protein can move on a variety of timescales.¹ The types and timescales of motions that the proteins experience play significant roles dictating the way they function. There may be local motions like small allosteric changes or large scale motions like domain

motions.² While a majority of these conformational changes take place upon the event of binding of proteins to lipids, ions, ligands and/or small molecules, some motions occur due to environmental changes such as varying pH, ion concentrations or temperature. Many biophysical techniques such as crystallography, NMR, atomic force microscopy and FRET can be used to study macromolecular conformational changes.^{3–5} Some computational methods have also been applied to quantitatively and qualitatively describe conformational changes such as molecular dynamics,^{6,7} normal modes analysis⁸ and elastic network models.⁹

Conformational changes prove that sequence-structure-function paradigm is not as straightforward; a sequence does not always fold into an exact fold/structure, which can further be altered to facilitate the protein's function. Both protein motions and disorder are good evidence of this complicated relationship. To perform their specific functions, proteins often adopt alternative conformations displaying considerable plasticity and fluidity. Similarly, natively disordered proteins exist as dynamic ensembles of different folds. While many of these conformational changes have been studied individually in a context-specific manner, it would be useful to study the general effects of these changes on structural properties on a large scale which would answer many questions. What are the regions on the molecular surface and inside the core that are affected? Where do these regions lie with respect to the parts that show motion? Do certain amino acids display a higher tendency to move during these conformational changes? Here, we relate protein motions to two features of protein structural organization: packing efficiency and disorder.

Atomic packing has been recognized as an important metric for characterizing protein structures since it was observed that average packing density for the protein cores is approximately the same as that for crystals of small organic molecules.¹⁰ Protein design and engineering studies suggest a crucial role for packing in protein stability and function,^{11–13} including exact complementarity of neighboring side chains.^{14–16} Even conservative mutations of single amino acids can lead to destabilizations.¹⁷ Additionally, the inclusion of an explicit packing term in protein design algorithms has significantly improved the accuracy of designed predictions,¹⁶ indicating that optimal packing is a crucial factor in protein structures.

Disorder in certain parts of the proteins results from a lack of relatively fixed structure.^{18,19} These proteins exist as inter-converting, dynamics ensembles of structures, instead of folding into one fixed structure. Conformational flexibility facilitates a number of post-translational modifications, such as phosphorylation²⁰ and ubiquitination.^{21,22} Not only can individual disordered proteins and regions bind to multiple partners,²³ but also multiple disordered sequences can each adapt

to fit one partner.²⁴ These partnering abilities of disordered proteins suggest their importance and common usage in protein interaction and signaling networks. Here, we relate protein disorder to conformational changes. It should be noted here that our definition of disorder is slightly different from the conventional terminology defining completely disordered residues. Here, since the protein structures are available, our focus is on '*marginally disordered*' residues (referred to as disordered residues at some places below): the ones that are sufficiently ordered to be crystallized in the protein structure but have signature sequence signals that indicate disorder in those regions.

Previous studies related to packing efficiency have looked at a static picture of protein structures. The aims of these studies were such that a single conformation of a protein was sufficient, such as development of novel methods to calculate packing density,^{10,25–28} analysis of protein-water interface,²⁹ its evaluation and relation to functional classification and size^{30,31} and its contribution in protein stability and folding.^{12,13,32,33} In this report, we analyze the dynamics of the proteins in relation to packing efficiency and disorder at a large-scale.

Results

Propensity to move and change packing efficiency

Beginning with the protein structures as available in the MolMovDB database³⁴ that show alternate conformations and after applying certain filters (see materials and methods), we obtained 40 pairs (80 structures). Next, we calculated packing efficiency using *packing_eff* program³⁵ that employs the radii-adjusted Voronoi polyhedra method.³⁶ To correlate protein conformational changes with changes in packing efficiency, we identified those residues that show a change in packing density. They were defined as the residues with magnitude change in packing efficiency larger than 0.1 between the two conformations (see Methods for details); others were classified as residues that maintain their packing density. For identification of residues that move, we used a more pair-specific threshold as the proteins in our set displayed motion over a large range and hence, classifying all these residues from different pairs into moving and stationary with one global threshold may not be appropriate (see Materials and Methods).

For every residue type, we calculated its overall propensity to move or change its packing efficiency. Propensity was defined the ratio of occurrence of that residue type that show a particular characteristic (such as change in packing efficiency) and its overall occurrence. E.G., Propensity of a residue type to move = (occurrence of that residue type that are involved in motions)/(total occurrence of that residue type). There are some reasonable observations from the trends in

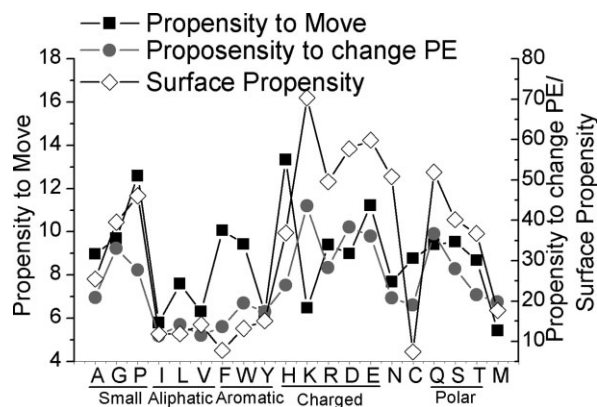


Figure 1. Propensity of each residue to move or change its packing efficiency. Propensity to move is plotted on the primary (left) Y-axis and, propensity to change packing and surface propensity are shown on secondary (right) Y-axis. Residues on the X-axis are grouped together based on their characteristics.

these propensities (see Fig. 1). First, aliphatic and aromatic residues (for e.g., Ile, Leu, Val, Phe and Trp) show little tendency to change their packing density where as charged residues (for e.g., His, Lys, Asp and Glu) show a high tendency to change packing. When put together, these two observations that hydrophobic residues which form the core of the proteins retain their packing efficiency and charged residues which are mostly located on the surface show a change in packing suggests that packing of the core of the proteins is maintained during the conformational changes and most of the packing efficiency changes occur on the molecular surface. Augmenting this observation,

hydrophilic residues, for e.g., Gly, Pro, Gln and Ser, show a relatively higher propensity to change packing efficiency. Finally, both these propensities are parallel to each other (correlation coefficient = 0.76) indicating that residues that have a higher tendency to move also have a higher propensity to change their packing efficiency between alternative conformations. We also plot the propensity to be found on the surface (defined as above). There are many amino acids that show differences in propensity to move and surface propensity (Leu, Phe, Trp, Lys, Arg, Asp and Cys) suggesting that former is not simply identifying residues with high surface propensity (see Fig. 1). As expected, we also find that surface propensity and propensity to change PE are more correlated (PCC = 0.89).

We also analyzed the maintenance of H-bonds between the two conformations for our set and found that most of the protein pairs preserve the number of H-bonds (Supporting Figure S1a). This suggests that although these proteins undergo conformational changes, they rearrange themselves in a way that maintains the H-bonds. The only protein in our set that changes the number of H-bonds drastically is Ran protein undergoes a substantial rearrangement of a helix arm (Supporting Figure S1b). This observation is in agreement with previous studies that suggest that upon ligand-binding, although H-bonds mediated by water molecules are lost/gained, most of the intra-protein H-bonds are preserved.^{37,38}

Next, we wanted to determine where these changes in packing occur on the molecular surface. One of the regions of interest during these conformational changes are the interfaces as most of these

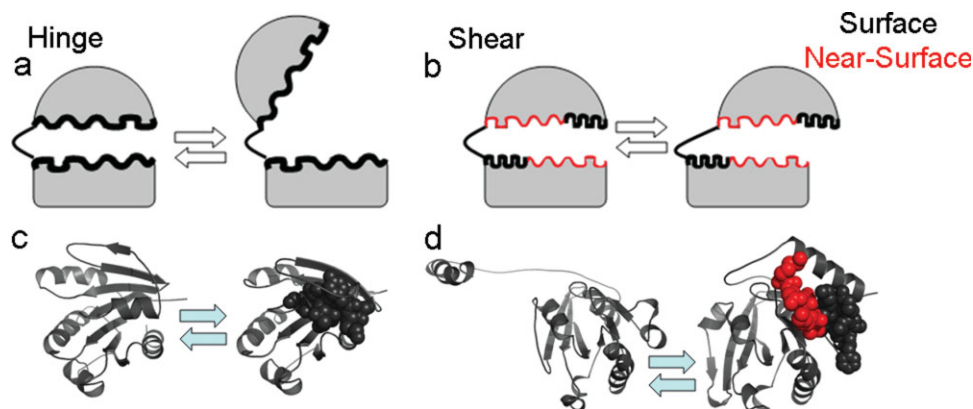


Figure 2. Protein conformational changes and changing interfaces in hinge motion (a and c) and shear motion (b and d). Changing interfaces are formed by those buried/exposed residues that become exposed/buried in the other conformation (shown as thick black zigzags) and are identified as those residues that have a large change in their exposed surface area between the two conformations. Interfaces are shown as zigzags (both thick and thin). Note that with the above definition, not all motion interfacial residues will be identified (for e.g., those shown as thin red zigzags that do not change their exposed SA that are called as near-surface interfaces). Examples of the protein motions (c and d). In the first case (c), Arf6-GDP (PDB: 1E0S), the top flexible region (including the small β -sheet) folds in and forms a changing interface (PDB: 2J5X). The interfacial residues are shown in vdW representation and display greater than 30% difference in the exposed SA. Similarly, in the second case (d), Ran protein (PDB: 1RRP), the flexible region on the far left (with the α -helix) slides in (PDB: 1BYU) forming a changing interface (shown in vdW representation). Surface and near-surface interfaces are colored in black and red respectively.

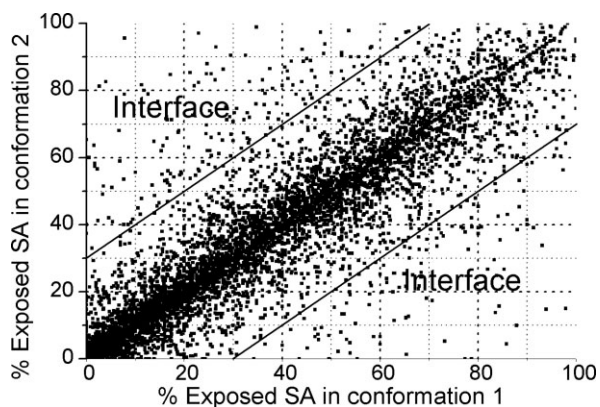


Figure 3. Distinguishing between interfacial and non-interfacial residues. The two solid lines indicate the region where the difference between the % exposed SA in the two conformations is more than 30%: the residues beyond the two lines are interfacial residues and the ones falling between the lines are non-interfacial atoms.

motions involve formation of new interfaces between the two states which we call as ‘changing interfaces’. These changing interfaces are formed by residues that are exposed/buried in one conformation and get buried/exposed in the other conformation (see Fig. 2). These changing interfaces are slightly different from the conventional *shear* or *static* interfaces as not all shear/static interfacial regions involve a change in exposed surface area; some of them remain buried during the motions [shown as red zigzags in Fig. 2(b)]. Schematic in Figure 2 shows the interfacial residues that may not be captured by this definition. We identified interfacial residues as those whose percentage exposed surface area differs by more than 30% between the two conformations while all others are non-interfacial. This cutoff was chosen as it gave the sharpest division with the highest interclass distance (after 10%, which is too small a change) of the residues into two classes (Supporting Figure S2). As examples, Figure 2(c,d) shows two such proteins with

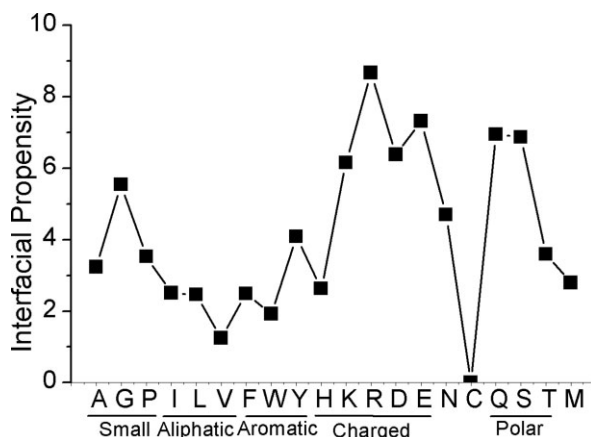


Figure 4. Propensity to be located on the changing interface for each residue type.

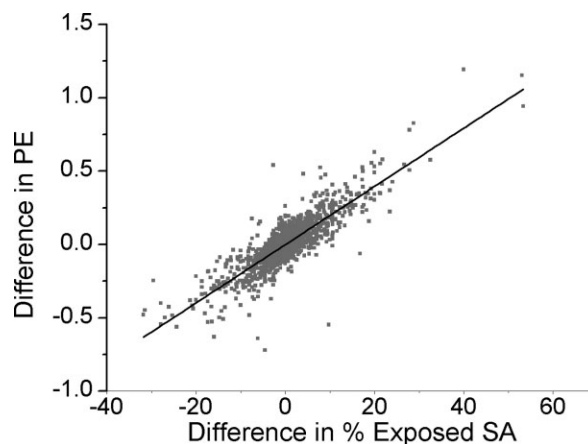


Figure 5. Difference in packing efficiency as a function of difference in % exposed SA between alternative conformations. The solid line is the linear fit between the two quantities.

residues with difference in % exposed SA $>30\%$ and, in fact, these residues are the interfacial residues. Figure 3 shows this differentiation between interfacial and non-interfacial residues using the % exposed SA in the two conformations.

For each residue type, we calculated the interfacial propensity defined in a similar way as above (see Fig. 4). Interestingly, interfacial propensity follows similar trends as propensity to change packing: charged and polar residues display higher propensities where as aliphatic and aromatic residues display lower propensities. The fact that residues that show high (low) propensities to be located on the interface also show high (low) propensities to change packing suggests that most of the packing changes between alternative conformations occur at the interfaces. This observation is further corroborated by investigating the relationship between difference in packing and difference in % exposed SA (see Fig. 5): residues having a larger difference between the % exposed SA in the two conformations (and hence having a higher propensity to be

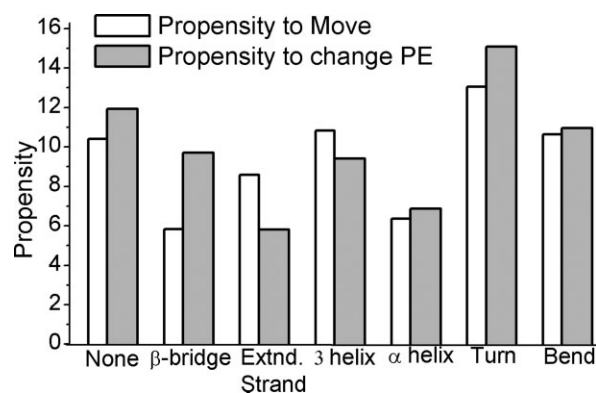


Figure 6. Propensities to move or change packing efficiency of residues located on different secondary structures.

Table I. *Atoms with the Highest Propensity to Change Packing*

GLU_OE1(46,63)	GLN_OE1(35,51)	LYS_NZ(18,29)	GLU_OE2(35,61)
ASN_OD1(37,73)	ARG_NH2(26,52)	GLN_NE2(26,54)	LYS_CE(30,64)
ASP_OD2(38,83)	HIS_NE2(22,50)	LYS_CD(54,123)	ARG_NE(40,100)
ASP_OD1(44,115)	GLU_CG(61,160)	HIS_CE1(17,49)	PRO_N(1,3)
TYR_OH(38,114)	LYS_CG(65,198)	TRP_NE1(17,52)	ARG_NH1(22,69)
ASN_ND2(21,68)	ARG_CD(32,118)	LYS_CB(75,277)	HIS_ND1(18,68)
THR_OG1(50,193)			

Atoms from charged and polar residues (these residues show a high propensity to change packing efficiency) are highlighted in bold. Also indicated are the number of occurrences when each atom type changes PE (the first number in parentheses) and their total occurrences (the second number). The atom types are listed in the decreasing order of this fraction (ratio of occurrences when they change PE and their total occurrences).

located on the interfaces) show a higher difference in packing efficiency.

We also compared the similarities between different amino acids regarding different propensities (propensities to move, change PE and lie on the interface) to their proximity with respect to the Blosum score. For a simple comparison between the trends, we build hierarchical clustering dendrograms based on these two properties (Blosum score and the three propensities, Supporting Figure S3). After dividing the set into four different clusters and comparing corresponding clusters (see Figure legend for details), we find that a majority of amino acids fall in the same clusters between the two dendrograms suggesting that similarities in these propensities might be due to inherent chemical similarities between the amino acids.

Relation to secondary structure

We next investigated if there is any relationship between a residue's propensity to show motion or change packing efficiency and its location in a specific secondary structure. Here, motion refers to the movement of the secondary structure as a more or less rigid body and not any internal deformation. We assigned secondary structure to every residue, using DSSP³⁹ and calculated the propensities to move or change packing efficiency (defined similarly as above) of residues that are embedded in different secondary structures (Fig. 6). Not surprisingly, highest propensities to move and change packing are observed by residues

that are located on flexible regions of the proteins like turns and bends or the ones that do not lie on any of the defined secondary structures. In contrast, lowest values of both these propensities are displayed by residues on α -helices suggesting that helices are the most robust to movement and changes in packing efficiency. Surprisingly, residues buried in β -bridges show a low propensity to move but a relatively higher propensity to change packing efficiency. This is reasonable in light of the fact that β -bridges are usually formed by hydrogen bonds between the neighboring sheets and a dislocation of the residues might disrupt this arrangement where as changes in packing efficiency may be compatible.

Propensities at the atomic level

We also performed similar analysis at a finer level of atoms. As above, we calculated the propensity of different kinds of atoms to change their packing. Table I lists the top 25 atom types with the highest propensities. As expected, the list is populated by atoms from the residues that show high propensity to change packing (charged and polar, 19/25, highlighted in bold). More specifically, a large majority of these atoms is formed by the side chain Oxygen and Nitrogen (14/19).

Next, we defined and calculated the interfacial propensities of atom types as above. Table II lists the first 25 atom types with the highest interfacial propensities. Interestingly, not only is the list dominated by

Table II. *Atoms with the Highest Interfacial Propensity*

ARG_NE(59,161)	ASN_OD1(43,176)	ARG_NH1(43,198)	LYS_CG(78,365)
LYS_CD(68,345)	GLU_OE1(54,281)	GLU_OE2(50,268)	GLU_CG(58,342)
ASP_OD1(50,296)	LYS_CE(52,339)	THR_CB(49,324)	VAL_CB(72,478)
LEU_CG(87,582)	GLU_CA(46,324)	SER_OG(43,325)	LYS_CB(58,442)
THR_OG1(42,343)	<i>ASP_N(42,346)</i>	<i>GLU_N(47,396)</i>	GLU_CB(47,403)
LEU_CA(60,583)	<i>LYS_N(47,466)</i>	<i>GLU_O(43,432)</i>	LYS_O(44,485)
GLY_O(42,488)			

Atoms which also appear in the list of atoms with high propensity to change packing efficiency (Table I) are highlighted in bold. Atoms that do not appear in table I but are similar to one of them are shown in italics. Also indicated are the number of occurrences when each atom type is found on the interface (the first number in parentheses) and their total occurrences (the second number). The atom types are listed in the decreasing order of this fraction (ratio of cases when they occur on the interface and their total number).

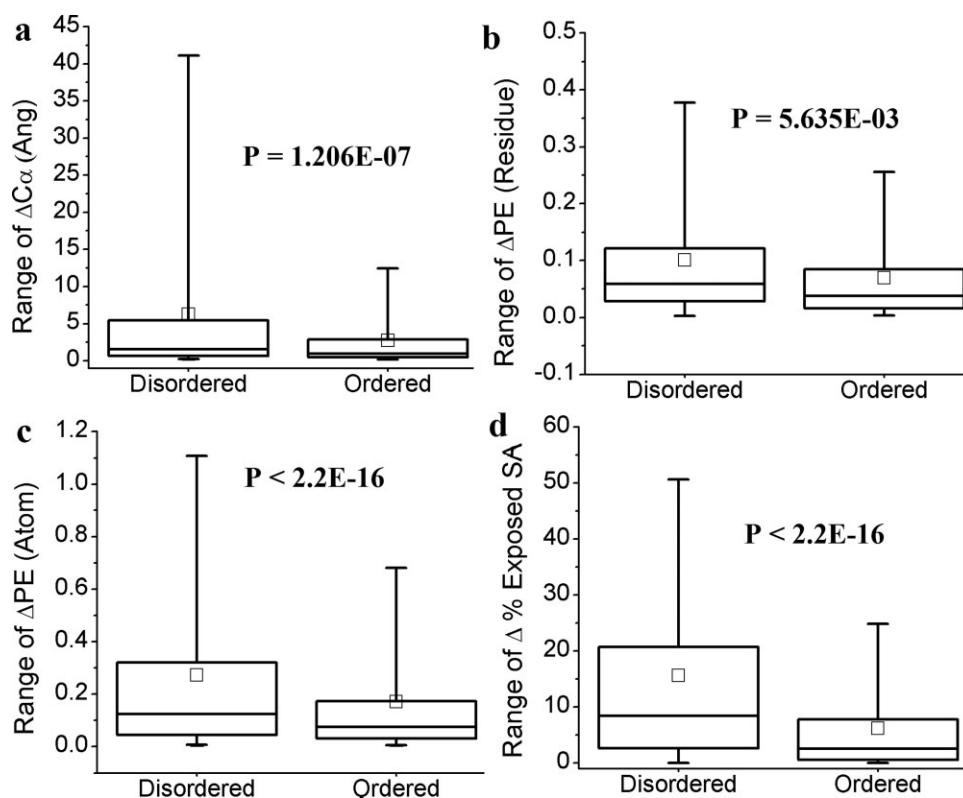


Figure 7. Comparing various motion-related parameters for ordered and marginally disordered residues: (a) Displacement of the $C\alpha$ atom during protein motion. (b) Residue-level and (c) atom-level change in packing efficiency. (d) Difference in % exposed SA for ordered and disordered residues between the two conformations. P -values indicated were calculated using the Kolmogorov-Smirnov test with the null-hypothesis that the two distributions of values (for disordered and ordered) are coming from the same underlying distribution. So, a lower P -value means that the two distributions are more dissimilar.

atoms from charged and polar residues, a majority of them are the same atoms that appear in list of atoms with high tendency (shown in bold and italics). This observation that atoms with high interfacial propensity also show a high inclination to change packing efficiency reinforces our demonstration above that most of the packing efficiency changes occur at the interfaces.

Relating protein motion to disorder

To relate protein disorder to protein motions on a large-scale, we compared motion-related parameters of marginally disordered residues to those of ordered ones. We identified marginally disordered residues using DISOPRED²⁴⁰ and obtained some interesting observations. First, marginally disordered residues displayed a much higher displacement of the $C\alpha$ atoms as compared to the ordered ones between these conformations [Fig. 7(a), $P = 1.2E-07$] indicating their larger contribution in protein motions. Second, disordered residues showed much higher changes in the packing efficiency both at the residue and atomic level [Fig. 7(b,c), $P = 5.6E-03$ and $P < 2.2E-16$, respectively]. Finally, disordered residues display a much higher difference in % exposed SA than ordered ones [Fig. 7(d), $P < 2.2E-16$] suggesting that disordered residues are

more likely to be located on the interfaces. This is also established by differentiating between disordered and ordered residues that lie at the interfaces (Fig. 8, see legend). These three observations, combined together, imply that the interfaces of the proteins are more disordered than the core of the proteins or other surface regions and, agreeably, show a much higher change in packing efficiency.

Robustness of results

All the results above were obtained by calculating the packing efficiency of protein structures in vacuum using the Voronoi method. This method constructs planes between neighboring atoms and the intersection of these planes forms a convex polyhedron which defines the Voronoi volume. This method is most useful on atoms for which these volumes are well defined or atoms that are at least partially buried. So, it may be argued that the above results that relate to packing efficiency might not hold true when the protein is solvated. To see if the results above were affected by the solvation states of the protein, we carried out the same calculations in the presence of explicit solvent by placing a water-box around each protein and also after simulating this water box to obtain a relaxed state of the solvent (see Materials and Methods). The idea of

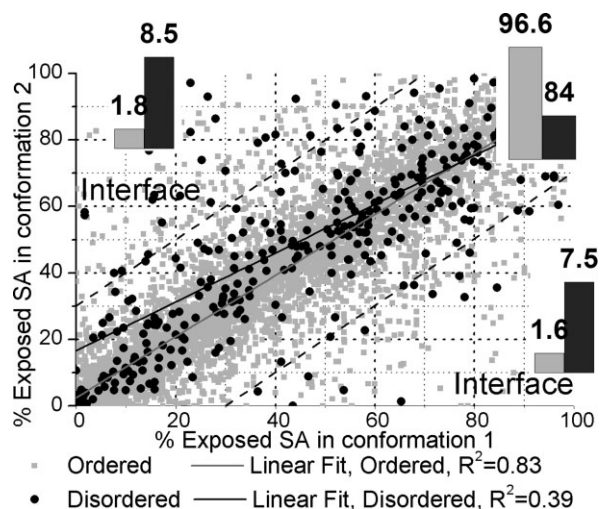


Figure 8. Differentiating between ordered and disordered interfacial and non-interfacial residues. The three bar graphs show the fraction of ordered and disordered residues lying in the three regions (two interfacial and one core region). A larger fraction (16%) of disordered residues fall on the interfaces than ordered ones (3.5%). This is also suggested by a much lower value of R^2 (which indicates a higher deviation from the linear fit and more proximity to the *interface* triangles on the graph) for disordered residues than ordered residues.

solvating the protein in a relaxed water box is to allow the Voronoi polyhedra to be closed and evaluate if the key results regarding packing efficiency still hold as they did for the proteins in vacuum.

We regenerated all key results related to packing efficiency above and found that they are reproducible with good agreement after solvating the proteins and after minimizing/simulating the water box around them (Supporting Figures S4–S6 and Table S1). We only reproduce results pertaining packing efficiency as only they are likely to be changed; the ones related to propensity to move or interfacial propensity will remain the same after solvating a protein. First, the trend of residue-wise propensities to change packing is well maintained (PCC = 0.87 between propensity set in vacuum and when solvated, PCC = 0.61 between propensity set in vacuum and after simulation, Figure S4). There are some differences in the absolute values of these propensities in the three states, which is justified as the protein is now surrounded by water molecules. Second, at the atomic level, the list of atoms showing a high propensity to change packing is dominated by atoms from charged and polar residues, as observed pre-solvation (Table S1). Third, in relation to secondary structures, identical behaviors are observed after solvation and simulation for all secondary structures (PCC = 0.92 between pre-solvation and post-solvation, PCC = 0.89 between pre-solvation and post-simulation; Figure S5). Finally, in relation to disorder, we observe that post-solvation and post-simulation, disordered residues are more likely to change their

packing as compared to ordered ones (Figure S6). This shows that the results reported above do not just apply to protein structure in vacuum but are independent of the solvation states of the proteins.

It has been shown previously that hydrogen atoms play an important role in determining the packing efficiency of proteins.⁴¹ For our calculations above, hydrogen atoms were removed (if present) from the structures. To see if the results presented in this study are affected by inclusion of Hydrogen atoms, we repeated the analysis and generated key results related to packing efficiency after adding explicit Hydrogen atoms to the protein structures. Encouragingly, we find that all our key results after addition of H atoms agree well with those without hydrogen atoms suggesting the robustness of our results to inclusion/exclusion of explicit H atoms (see Supporting Materials and Figures S7 through S10).

Discussion

In this study, we have studied the interrelationship of three characteristics of protein structures: motion, packing and disorder. Previous studies related to packing efficiency have looked at the static picture of protein structures with various aims that only required a single conformation of a protein. In this study, however, we extend the analysis to a dynamic perspective by studying the changes in packing efficiency of various proteins that show motion and adopt alternate conformations. Similarly, by identifying the marginally disordered residues in these proteins, we have also related protein disorder to both protein dynamics and changes in packing efficiency.

By investigating the relationship, we have obtained some seemingly-intuitive but interesting results. First, propensities of different types of residues to show motion or change their packing show a strong correlation. Charged and polar residues show high propensities where as aromatic and aliphatic residues have relatively lower propensities. Second, the interfacial propensities of residues also follow similar trends, i.e., the residues that show a higher a change in packing efficiency also have a higher tendency to be found on the changing interfaces. These observations, when put together, indicate that while most of the changes in packing efficiency occur at the interfaces, the core of the proteins remain well packed during even large scale motions. Next, in relation to secondary structures, α -helices are more robust in terms of low propensities to either move or change packing where as bends are more elastic displaying higher propensities. Interestingly, residues located on β -sheets have a low tendency to move but a high propensity to change packing. In relation to disorder, our results suggest that disordered residues display a higher movement. They also have a higher propensity to change their packing during these motions than ordered ones and are much more likely to be found on the changing interfaces.

We also show the consistency of these results at two levels and in various conditions. First, by performing similar analyses at the atomic level, we obtain similar results suggesting that our observations hold at the atomic level. Second, by solvating the proteins and simulating the solvent-box, we also show that our results are independent of the solvation states of the proteins. Finally, in this study we have used Voronoi-based definition of packing density while there are other ways to define packing efficiency such as that based in occluded surfaces.²⁸ However, the underlying method for both these definitions is based on the available volume around each atom. The difference arises for the surface atoms which don't have a definite Voronoi volume. To address this issue, we have solvated our proteins in simulated water boxes (this defines the Voronoi volume for every atom including the surface ones) and obtained similar results. Thus, we believe that our core observations will not be affected by using an alternative definition of packing density. Still, one of the potential limitations of this study arises from the fact that there are only a few proteins that show a large motion (Supporting Figure S11a). So, some of the findings presented here might not hold true for *all* kinds of large scale motions.

Having established some interesting correlations above between various structural metrics, a natural question that arises is: does loose/tight packing give any predictive indication about parts of the protein that move? This is similar to the question: do the packing efficiencies of regions that move differ significantly from those of the regions that do not. We investigated this and found that there are no significant differences between the two regions in terms of their packing efficiencies (Supporting Figure S12). This suggests that while *changes* in packing efficiency between the two conformations correlate well with protein motions, it would be difficult to predict which parts of the protein would show motion based purely on absolute packing efficiency in either one of the two conformations. On a related issue, it would also be useful to correlate protein motions to the entropic changes between different conformations giving some insight to the energetic requirements. However, currently there is no sufficient calorimetric data available for the protein set used in this study to make any statistically significant correlations (see Supporting Materials for details). At this point it can only be reasonably speculated that transition from unpacked and disordered states to tightly packed and ordered states would involve a high degree of entropic loss and this loss might help the protein in carrying out its function. It would also be interesting to study how the protein interaction networks reorganize due to the conformational changes of the participating proteins. Changes in conformations due to loose packing or other reasons are both the results and causes of formation of protein complexes. Loose packing, especially at the

binding interfaces, might indicate weak or unspecific binding between proteins. Similarly, specific properties of these interfaces allow certain interaction hubs to interact with multiple proteins using the same or different interfaces. It would be insightful to see how the interaction networks adjust to conformational changes of the participating proteins.

All the observations reported in this study suggest that, even during large scale movements involving whole domain motions, while most of the changes in structural characteristics take place on the molecular surfaces (more specifically at the changing interfaces of these motions), the cores of the proteins remain more or less intact. This is in alignment with previous studies that have suggested many important roles for core packing in proteins towards various purposes like protein stability, folding and design.^{12,16,32,42,43} The central role of core residues in determining protein conformation and stability may stem from simple features like presence of hydrophobic residues or favorable geometrical compatibility resulting in efficient packing. Any perturbation in this packing resulting from any phenomena like protein conformation change may be detrimental to its stability. So, it would only be natural for proteins to maintain the efficient packing in their cores while carrying out a myriad of functions which ensures a smooth functioning of the cell. Such insights into the correlations between various structural metrics in a dynamic perspective might be very useful for protein engineering and design studies.

Materials and Methods

Dataset collection

We began with the protein structures that show alternate conformations available in the MolMovDB database.³⁴ To carefully retain only genuine conformational changes, we only kept those pairs that had identical sequences in the two conformations. From this list, structures were removed that had missing coordinates for residues in the middle of the sequence. This allows a fair comparison of the packing efficiency between the two structures. With these filters, the final set had 40 pairs (80 structures). These structures came from different classes of motions as classified in MolMovDB such as shear motion, hinge motion and other motions that can not be classified either of the two (Table III), and had a wide range of sequence lengths. Although, when average RMSD is calculated over the entire structure there are not many proteins with high RMSD, there is a good fraction of residues that have relatively higher RMSD (Supporting Figure S11).

Calculation of packing efficiency

Many methods have been proposed to evaluate the packing densities in proteins, such as Voronoi method,^{10,25,26} occluded surface method^{28,30} and

Table III. Various Kinds of Motions Analyzed in this Study and the Corresponding Number of Protein Pairs

Protein region	Motion	Number of pairs
Fragment	Shear	6
	Hinge	7
	Not shear or hinge	3
	Unclassified	5
Domain	Shear	3
	Hinge	7
	Not shear or hinge	1
	Unclassified	1
	Refolding of tertiary structure	1
Subunits	Allosteric	4
	No allostery involved	2

sphere growth method 27.²⁷ For this study, packing efficiency was calculated using *packing_eff* program.³⁵ This program implements the traditional Voronoi polyhedra method³⁶ where a polyhedron is constructed by partitioning space such that all points within a polyhedron are closer to the atom defining the polyhedron than to any other atom. The Voronoi planes are shifted from the original equidistant planes to the modified set determined by the relative sizes of the van der Waals (vdW) radii of the atoms, i.e., bigger atoms take up more space in the Voronoi construct than smaller ones.⁴⁴ The packing density is then defined as Voronoi volume divided by vdW volume and measures how tightly each atom packs. Since it is a ratio of two quantities with the same unit, packing efficiency is unitless. Only atoms whose volumes were well-defined and were not unpacked were included.⁴⁴ Unpacked atoms usually consist of surface atoms or atoms near cavities and therefore do not have enough neighbors to pack tightly. The overall packing density for such surface-residues was measured by averaging over only those atoms that had a well-defined density in both the conformations. So, strictly speaking, these are “near-surface” atoms.

Identification of marginally disordered residues and their assignment to secondary structures

For identification of marginally disordered residues, we used DISOPRED2 which is an SVM-based protocol for identification of disordered regions in protein sequences and has been shown to achieve two-state accuracy (Q_2) above 93% for a false hit rate of 0.05.⁴⁰ It has been built and trained on a large set of about 750 non-redundant sequences with high resolution X-ray structures and has been shown to outperform other methods such as FoldIndex and methods from Obradovic & Dunker groups. A sequence profile is generated for each protein using a PSI-BLAST search against a filtered sequence database. The input vector for each residue is constructed from the profiles of a symmetric window of fifteen positions. The main reason for DISOPRED2’s outperformance is that it is

trained directly on protein sequence rather than measures of amino acid composition, sequence complexity or biophysical properties such as mean hydrophobicity. This may allow the classifier to recognize sequence motifs that have been shown to be associated with disorder such as Pro-X-Pro-X-Pro or Lys-X-X-Lys-X-Lys. It has been shown in a series of papers that there are clear patterns that characterize disordered regions such as low sequence complexity, amino acid compositional bias (e.g., towards aromatic residues) and high flexibility, and that disorder can be predicted successfully from amino acid sequence.^{19,45–47} Although the B-factor (for X-ray structures) can also be used as a measure of atoms’ oscillations, we chose DISOPRED2 for a few reasons (see Supporting Materials). However, we confirmed that identification of disordered residues from DISOPRED2 agrees encouragingly with the trends in B-factors for these structures: disordered residues from DISOPRED2 have much higher B-factors than the ordered ones ($P < 2.2E-16$, Supporting Figure S13).

To assign secondary structure to every residue, we used DSSP that lists seven kinds of secondary structures: α -helix, isolated β -bridge, extended strand, 3-helix (3/10 helix), 5 helix (π helix), hydrogen-bonded turn and bend.³⁹

Definition of residues that move and change their packing efficiency

We defined residues as the ones that change their packing efficiency if their magnitude change in packing efficiency was larger than 0.1 between the two conformations. All others were classified as residues that maintain their packing density. This cutoff was chosen because as observed from the frequency histogram of the residues showing different changes in PE (with a step of 0.1), the residues get very sharply divided between the two classes; there is a huge drop in the frequency of residues at change in PE of ± 0.1 (Supporting Information Figure S14).

The proteins in our set displayed motion over a large range: small loop movements over a few Å to large domain motions over more than 50 Å. So, a global threshold to classify these residues from different pairs into moving and stationary might not be plausible. Thus, we classified residues using a more case-specific threshold. For each protein pair we calculated the displacement of the C α atoms between the two conformations and sorted these displacements in the descending order. The two consecutive values in this sorted list that had the highest difference between them were then identified and the higher of these values was the threshold: any residue with C α displacement more than this value was classified as moving, and the rest were classified as stationary. This approach sets the threshold for every protein pair separately and divides the residues into two sets with the highest inter-class distance.

Solvation and simulation of proteins in explicit solvent

To test the robustness of our results to solvation states, we solvated the proteins and carried out same calculations in the presence of explicit solvent. We first solvated the proteins in our set using the *solvate* plugin implemented in VMD.⁴⁸ *Solvate* works by first enclosing the solute in a minimal convex volume and then defining an iso-surface of a density-function of the solvent. To obtain different orientations of water molecules around the protein, this process was iterated 10 times using different margins of solvent layers. The ensuing calculations were based on values that were averaged over all these 10 runs.

Further, to eliminate possible physically-irrelevant solvent-density or a constrained orientation of the water molecules, it may be needed to minimize the energy of the water box around the protein. This may be done using any of the energy minimization techniques followed by molecular dynamics simulations, which are time-consuming for reasonably large systems at an all-atom level. So, we performed the energy minimization steps followed by a small MD run on a smaller set of nine pairs (18 proteins, about 25% of the set). Steepest descent was used for minimization and NAMD⁴⁹ was used for MD simulations under standard conditions for 100 ps which is sufficient for the purpose of obtaining a reasonably relaxed state of water molecules (see Supporting Materials for details of the parameters used for simulation).

Acknowledgments

The authors thank the anonymous reviewers who provided useful suggestions and constructive criticisms that helped improve the quality of the article.

References

1. Gerstein M, Echols N (2004) Exploring the range of protein flexibility, from a structural proteomics perspective. *Curr Opin Chem Biol* 8: 14–19.
2. Gerstein M, Lesk AM, Chothia C (1994) Structural mechanisms for domain movements in proteins. *Biochemistry* 33: 6739–6749.
3. Marchant RE, Kang I, Sit PS, Zhou Y, Todd BA, Eppell SJ, Lee I (2002) Molecular views and measurements of hemostatic processes using atomic force microscopy. *Curr Protein Pept Sci* 3: 249–274.
4. Zuiderweg ER (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* 41: 1–7.
5. Kajihara D, Abe R, Iijima I, Komiyama C, Sisido M, Hohsaka T (2006) FRET analysis of protein conformational change through position-specific incorporation of fluorescent amino acids. *Nat Methods* 3: 923–929.
6. Yin Y, Jensen MO, Tajkhorshid E, Schulten K, (2006) Sugar binding and protein conformational changes in lactose permease. *Biophys J* 91: 3972–3985.
7. Holyoake J, Sansom MS (2007) Conformational change in an MFS protein: MD simulations of LacY. *Structure* 15: 873–884.
8. Hinsen K, Thomas A, Field MJ (1999) Analysis of domain motions in large proteins. *Proteins* 34: 369–382.
9. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 80: 505–515.
10. Richards FM (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. *J Mol Biol* 82: 1–14.
11. Dahiyat BI, Mayo S (1997a) De novo protein design: fully automated sequence selection. *Science* 278: 82–87.
12. Ventura S, Vega MC, Lacroix E, Angrand I, Spagnolo L, Serrano L (2002) Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nat Struct Biol* 9: 485–493.
13. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302: 1364–1368.
14. Kellis JT, Jr, Nyberg K, Fersht AR (1989) Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* 28: 4914–4922.
15. Desjarlais JR, Handel TM (1995) De novo design of the hydrophobic cores of proteins. *Protein Sci* 4: 2006–2018.
16. Dahiyat BI, Mayo SL (1997b) Probing the role of packing specificity in protein design. *Proc Natl Acad Sci USA* 94: 10172–10177.
17. Chen J, Stites WE (2001) Packing is a key selection factor in the evolution of protein hydrophobic cores. *Biochemistry* 40: 15280–15289.
18. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293: 321–331.
19. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41: 6573–6582.
20. Ruzza P, Donella-Deana A, Calderan A, Filippi B, Cesaro L, Pinna LA, Borin G (1996) An exploration of the effects of constraints on the phosphorylation of synthetic protein tyrosine kinase peptide substrates. *J Pept Sci* 2: 325–338.
21. Cox CJ, Dutta K, Petri ET, Hwang WC, Lin Y, Pascal SM, Basavappa R (2002) The regions of securin and cyclin B proteins recognized by the ubiquitination machinery are natively unfolded. *FEBS Lett* 527: 303–308.
22. Yoshida Y, Adachi E, Fukiya K, Iwai K, Tanaka K (2005) Glycoprotein-specific ubiquitin ligases recognize N-glycans in unfolded substrates. *EMBO Rep* 6: 239–244.
23. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE (1996) Structural studies of p21^{Waf1/Cip1/Sdi1} in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci USA* 93: 11504–11509.
24. Bustos DM, Iglesias AA (2006) Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins. *Proteins* 63: 35–42.
25. Finney JL (1975) Volume occupation, environment and accessibility in proteins. The problem of the protein surface. *J Mol Biol* 96: 721–732.
26. Richards FM (1977) Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* 6: 151–176.
27. Gregoret LM, Cohen FE (1990) Novel method for the rapid evaluation of packing in protein structures. *J Mol Biol* 211: 959–974.
28. Pattabiraman N, Ward KB, Fleming PJ (1995) Occluded molecular surface: analysis of protein packing. *J Mol Recognit* 8: 334–344.
29. Gerstein M, Chothia C (1996) Packing at the protein-water interface. *Proc Natl Acad Sci USA* 93: 10167–10172.

30. Fleming PJ, Richards FM (2000) Protein packing: dependence on protein size, secondary structure and amino acid composition. *J Mol Biol* 299: 487–498.
31. Seeliger D, de Groot, BL (2007) Atomic contacts in protein structures. A detailed analysis of atomic radii, packing, and overlaps. *Proteins* 68: 595–601.
32. Lim WA, Sauer RT (1991) The role of internal packing interactions in determining the structure and stability of a protein. *J Mol Biol* 219: 359–376.
33. Pattabiraman N (2003) Role of residue packing in protein folding. *Trends Anal Chem* 22: 554–560.
34. Flores S, Echols N, Milburn D, Hespenheide B, Keating K, Lu J, Wells S, Yu EZ, Thorpe M, Gerstein M (2006) The Database of Macromolecular Motions: new features added at the decade mark. *Nucleic Acids Res* 34: D296–D301.
35. Voss NR, Gerstein M (2005) Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J Mol Biol* 346: 477–492.
36. Voronoi GF (1908) Nouvelles applications des paramètres continus à la théorie de formes quadratiques. *J Reine Angew Math* 134: 198–287.
37. Arora, S (2005) Optimizing side-chain interactions in protein-ligand interfaces. Kuhn LA, Punch WF, Eds., In: *Department of Computer Science*. Ann Arbor: Michigan State University.
38. Cozzini P, Kellogg GE, Spyarakis F, Abraham DJ, Costantino G, Emerson A, Fanelli A, Gohlke H, Kuhn LA, Morris GM, Orozco M, Pertinhez TA, Rizzi M, Sotriffer CA (2008) Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem* 51: 6237–6255.
39. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577–2637.
40. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337: 635–645.
41. Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC (1999) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J Mol Biol* 285: 1711–1733.
42. Johnson EC, Lazar GA, Desjarlais JR, Handel TM (1999) Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure* 7: 967–976.
43. Farber PJ, Mittermaier A (2008) Side chain burial and hydrophobic core packing in protein folding transition states. *Protein Sci* 17: 644–651.
44. Tsai J, Voss N, Gerstein M (2001) Determining the minimum number of types necessary to represent the sizes of protein atoms. *Bioinformatics* 17: 949–956.
45. Li X, Romero P, Rani M, Dunker AK, Obradovic Z (1999) Predicting protein disorder for N-, C-, and internal regions. *Genome Inform Ser Workshop Genome Inform* 10: 30–40.
46. Dunker AK, Obradovic Z (2001) The protein trinity—linking function and disorder. *Nat Biotechnol* 19: 805–806.
47. Romero P, Obradovic Z, Li X, Garner E, Brown C, Dunker AK (2001) Sequence complexity and disordered proteins. *Proteins: Struct Funct Genet* 42: 38–48.
48. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33–38:27–38.
49. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–1802.