Structure
Technical Advance

Efficient and Robust Analysis of Biomacromolecular Flexibility Using Ensembles of Network Topologies Based on Fuzzy Noncovalent Constraints

Christopher Pfleger[1] and Holger Gohlke[1,*]
[1]Mathematisch-Naturwissenschaftliche Fakultät, Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany
*Correspondence: gohlke@uni-duesseldorf.de
http://dx.doi.org/10.1016/j.str.2013.07.012

SUMMARY

We describe an approach (ENT^{FNC}) for performing rigidity analyses of biomacromolecules on ensembles of network topologies (ENT) generated from a single input structure. The ENT is based on fuzzy noncovalent constraints, which considers thermal fluctuations of biomacromolecules without actually sampling conformations. Definitions for fuzzy noncovalent constraints were derived from persistency data from molecular dynamics (MD) simulations. A very good agreement between local flexibility and rigidity characteristics from ENT^{FNC} and MD simulations-generated ensembles is found. Regarding global characteristics, convincing results were obtained when relative thermostabilities of citrate synthase and lipase A structures were computed. The ENT^{FNC} approach significantly improves the robustness of rigidity analyses, is highly efficient, and does not require a protein-specific parameterization. Its low computational demand makes it especially valuable for the analysis of large data sets, e.g., for data-driven protein engineering.

INTRODUCTION

Biomacromolecules are composed of flexible and rigid regions. This stability heterogeneity allows biomacromolecules to fulfill their diverse functional roles (Teague, 2003). Hence, a precise knowledge of flexibility and rigidity characteristics of a biomacromolecule is a prerequisite for understanding its function and valuable information for rational protein engineering and structure-based ligand design. Flexible and rigid regions in biomacromolecules can be determined with experimental methods (Henzler-Wildman and Kern, 2007; Cozzini et al., 2008; Sterner and Brunner, 2008; Bernadó, 2010; Kleckner and Foster, 2011; Hammel, 2012) and computational approaches, including molecular dynamics (MD) simulations (Young et al., 2001; Dodson and Verma, 2006; Cozzini et al., 2008) and approaches based on connectivity networks (Heringa and Argos, 1991; Dokholyan et al., 2002; Halle, 2002; Vendruscolo et al., 2002; Greene and Higman, 2003; Böde et al., 2007).

In a different approach, protein structures are modeled as constraint networks, which are analyzed by applying concepts based on rigidity theory. Here, atoms are represented as bodies and covalent and noncovalent interactions (hydrogen bonds, salt bridges, and hydrophobic tethers) as sets of bars (constraints; Jacobs et al., 2001; Rader et al., 2002). Initially, each body has six degrees of freedom (Whiteley, 2005). However, potential motions are constrained by the bars connecting the bodies. Once the network is constructed, the pebble game algorithm (Jacobs and Thorpe, 1995), implemented in the FIRST software, identifies flexible and rigid regions from the number and spatial distribution of the degrees of freedom ("constraint counting"). This analysis only takes seconds for a protein of ~300 residues. The theory underlying this approach is rigorous (Katoh and Tanigawa, 2011). Results from rigidity analyses have been successfully compared to results from experiments and those from other computational approaches (Hespenheide et al., 2002; Jacobs et al., 2003; Gohlke et al., 2004; Rader and Bahar, 2004; Livesay and Jacobs, 2006; Radestock and Gohlke, 2008; Fulle and Gohlke, 2009a).

Rather than analyzing "static" networks, several studies analyzed "perturbed" networks in which hydrogen bond constraints are sequentially removed, that way simulating the thermal unfolding of a biomacromolecule (Rader et al., 2002; Radestock and Gohlke, 2008, 2011; Rader, 2009; Rathi et al., 2012). Hydrogen bonds are removed in the order of increasing strength according to a hydrogen bond energy E_{HB} (see Supplemental Experimental Procedures available online; Dahiyat et al., 1997). During the thermal unfolding simulation, a phase transition occurs at which the network loses the ability to carry stress; the transition point is referred to as rigidity percolation threshold (Thorpe, 1983) and has been related to the thermostability of proteins (Radestock and Gohlke, 2008, 2011; Rathi et al., 2012).

Rigidity analyses are sensitive with respect to the input structural information (Gohlke et al., 2004; Mamonova et al., 2005). Two reasons account for this: (1) biomacromolecules have a soft matter-like character (Zaccai, 2000), i.e., noncovalent interactions frequently break and (re-)form ("flickering") such that the number and distribution of constraints in the networks vary; and (2) biomacromolecules are generally marginally stable (Taverna and Goldstein, 2002), i.e., their network state is close to the rigidity percolation threshold. As an overall consequence, a few constraints more or less can result in a network either being largely rigid or flexible.

The sensitivity problem can be overcome by performing rigidity analysis on an ensemble of conformations, e.g., generated by MD simulations (Gohlke et al., 2004; Rathi et al., 2012). However, this compromises the efficiency of the rigidity analysis. To overcome this drawback, an ensemble of network topologies (ENT) can be generated from a single input structure by simulating the flickering of noncovalent constraints rather than the motions of the atoms. This idea has been pioneered in the distance constraint model (DCM; Jacobs et al., 2003; Livesay et al., 2004; Jacobs and Dallakyan, 2005) and the virtual pebble game (VPG) approach (González et al., 2012). While conceptually appealing, a downside of the DCM approach is that it requires experimental data for a protein-specific parameterization of the model. A downside of the VPG approach is that it is less accurate at the rigidity percolation threshold; analyzing network states of biomacromolecules around this threshold is particularly interesting, however.

In this study, we present an approach that performs rigidity analyses on an ENT generated from a single input structure (ENT^{FNC}) by using fuzzy noncovalent constraints (FNC). As such, the number and distribution of noncovalent constraints are modulated by random components within certain ranges, thus simulating thermal fluctuations of a biomacromolecule. The approach significantly improves the robustness of rigidity analyses, is highly efficient, and does not require a protein-specific parameterization. We analyzed MD simulations of hen egg white lysozyme (HEWL) structures as to the persistence of noncovalent bonds. From the analysis, we developed definitions for fuzzy hydrogen bond, salt bridge, and hydrophobic constraints. We validated the approach by comparison to rigidity analyses performed on single network topologies (SNT) of HEWL structures as well as on ensembles of HEWL conformations generated by MD simulations (ENT^{MD}). Furthermore, we demonstrate that our ENT^{FNC} approach is transferable to other protein systems. The ENT^{FNC} approach has been implemented into the CNA software (Pfleger et al., 2013a) and is available via the CNA web server (Krüger et al., 2013).

Theory

The FNC model consists of two parts related to the modeling of hydrogen bonds (including salt bridges) and hydrophobic tethers. Parameters of the model are derived from ENT^{MD} of HEWL. Values of the parameters are reported in the Results; here, we detail the theory underlying the FNC model.

Part Ia

To account for the flickering of hydrogen bonds (Zaccai, 2000), we determined the persistence characteristics of these interactions along MD trajectories. We did so for hydrogen bonds and salt bridges separately, and we distinguished hydrogen bonds in different secondary structure elements (α helices, 3_{10} helices, β sheets, and loop regions) following previous work (Stickle et al., 1992; Mamonova et al., 2005; Kieseritzky et al., 2006; Almond et al., 2007). From this, we derived the probability $p(HB,t)$ with which a hydrogen bond (salt bridge) of type t found in the input structure will be present across the ensemble of generated network topologies.

Part Ib

We next addressed that the hydrogen bond energy E_{HB} determines the order with which hydrogen bonds are removed in a thermal unfolding simulation. This order may strongly determine the computed global and local stability characteristics. E_{HB} is computed by a simplified energy function (see Supplemental Experimental Procedures; Dahiyat et al., 1997). When analyzing a SNT, thermal motions of atoms are neglected that may influence E_{HB} and, hence, the order of hydrogen bond removal. To account for this effect, Gaussian white noise is added to $E_{HB,i}^{initial}$ computed for a hydrogen bond i from the single input structure (Equation 1).

$$E_{HB,i} = E_{HB,i}^{initial} + \mathcal{N}(0, SD_{HB,t}) \quad \text{(Equation 1)}$$

The white noise is sampled from a Gaussian distribution $\mathcal{N}(0, SD_{HB,t})$ with a mean of zero and a standard deviation $(SD_{HB,t})$ that depends on the type t of the hydrogen bond; $SD_{HB,t}$ is determined from analyzing hydrogen bond energies in ENT^{MD}. Sampling from a Gaussian distribution follows the rationale that the shape of the energy well of a hydrogen bond can be fitted by a harmonic approximation (Leach, 2001) and that fluctuations about the minimum of a quadratic function show a Gaussian distribution (Levy and Karplus, 1979).

Part II

Regarding hydrophobic interactions, we wanted to model that these interactions are less specific than polar ones (Rose and Wolfenden, 1993). To do so, we developed a fuzzy constraint representation in which tethers between closer atoms are included with a higher probability in a network topology than those between atoms further apart; more specifically: (1) tethers between atoms that are in van der Waals contact d_{vdW} (van der Waals radii: C: 1.7 Å, S: 1.8 Å) are always included; (2) tethers between atoms that are further apart than $d_{vdW} + D_{max}$ are never included; here, D_{max} = 1.5 Å because this value equates to half of the distance between the contact minimum and the solvent-separated minimum in a potential of mean force of hydrophobic solutes (Pratt and Chandler, 1977); and (3) for distances in between, the probability for a tether to be included is computed from Equation 2, which approaches these two extremes.

$$p(d_{ij}) = e^{-\frac{1}{2}\left(\frac{\left(d_{ij}-d_{vdW}\right)^2}{D_{cut}^2}\right)^2} \quad \text{(Equation 2)}$$

$p(d_{ij})$ is a Gaussian with a squared distance dependency, d_{ij} is the distance between two atoms i and j, and D_{cut} determines the full width at half maximum of the Gaussian. Gaussians have been applied successfully for modeling the strength of pairwise interactions between hydrophobic atoms (Crivelli et al., 2002; Huey et al., 2007; Forli and Olson, 2012). Preliminary tests have shown that it is advantageous to favor hydrophobic tethers at shorter distances; this is accounted for by the squared distance dependency.

With this FNC model, in five steps, the ENT^{FNC} is generated from a single input structure, and global and local stability characteristics are analyzed (Figure 1):

(1) An initial network topology is generated from the input structure.
(2) Information about noncovalent constraints is extracted. Hydrogen bonds (including salt bridges) with

$$E_{HB}^{initial} < 0 \text{ kcal/mol}$$

and hydrophobic tethers between C and/or S atoms with a distance < $d_{vdW} + D_{max}$ are identified. The secondary

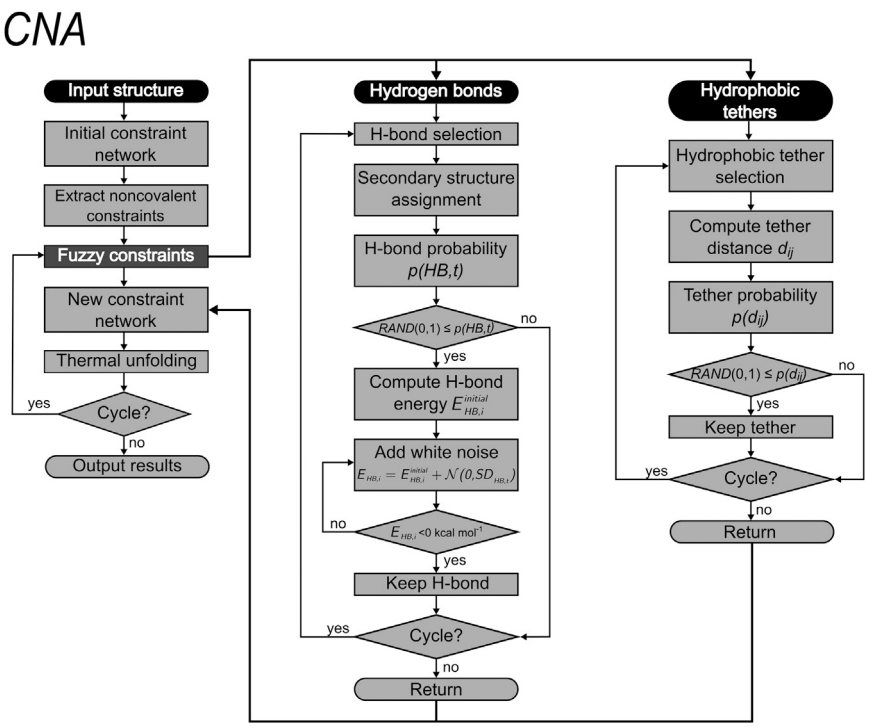


Figure 1. Work Flow of the ENT^{FNC} Approach

The ENT^{FNC} approach has been integrated into the CNA software package. *RAND*(0,1) draws a random number with equal probability from the range [0,1]. See Theory for further details.

structure a hydrogen bond or salt bridge is involved in is assigned from the input structure using DSSP (Joosten et al., 2011).

(3) The number and distribution of noncovalent constraints is modified according to the definitions of FNC.

(4) A network topology is built.

(5) Global and local stability characteristics are computed by the CNA software.

Steps 3–5 are repeated until a user-specified number of networks is generated over which the global and local stability characteristics are averaged.

RESULTS

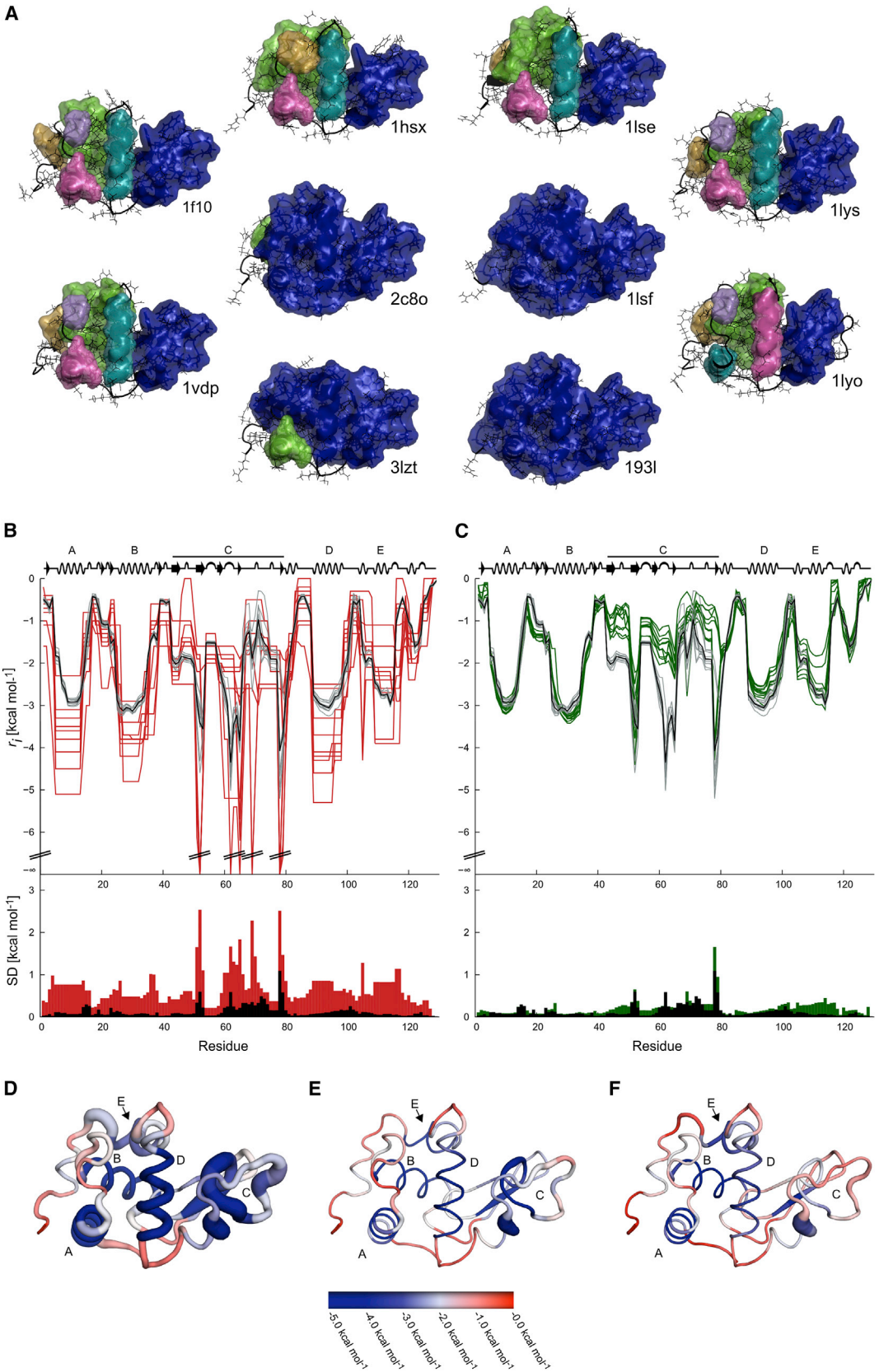
Rigidity Analyses Are Highly Sensitive with Respect to the Input Structure

To illustrate to what extent the results of a rigidity analysis depend on the chosen input structure, we computed rigid cluster decompositions (RCD) using FIRST for network topologies derived from ten HEWL crystal structures (SNT approach; Figure 2A). All of these structures (see Supplemental Experimental Procedures) have been resolved from 0.93 Å to 1.93 Å. Structures of a similar quality have been used in FIRST analyses before (Gohlke et al., 2004; Radestock and Gohlke, 2008; Fulle and Gohlke, 2009b; Rathi et al., 2012). The structures have been selected from a set of 38 HEWL structures with the aim to avoid structural redundancy (see Supplemental Experimental Procedures; Table S1). This resulted in mutual root-mean-square deviations (rmsd) of the C_α atoms (of all atoms) of at most 0.86 Å (1.85 Å). The RCD results fall into two classes: (1) between 43% and 50% of all atoms are part of a rigid cluster with the remaining ones being located in flexible regions (Protein Data Bank [PDB] IDs: 1LYO, 1VDP, 1F10, 1HSX, 1LSE, and 1LYS); and (2) a single rigid cluster dominates the system and contains between 57% and 72% of all atoms (PDB IDs: 2C8O, 1LSF, 3LZT, and 193L). These differences in the RCD results originate from differences in the number and distribution of noncovalent constraints in the network topologies: Network topologies with more constraints result in the "single rigid cluster" RCD, and the largest difference in the number of constraints across all ten networks is 41 (22%; Table S2).

To quantify the results of the rigidity analysis, the C_α atom-based rigidity index r_i was computed by CNA for each HEWL structure (Figures 2B and 2D). r_i maps flexibility and rigidity characteristics within a network topology by monitoring when a covalent bond segregates from a rigid cluster during a thermal unfolding simulation (Pfleger et al., 2013b). Figure 2B reveals that the local stability characteristics are only moderately consistent in helices A, B, D, and E (SD ~0.8 kcal/mol) and vary strongly in the beta sheet region C (SD up to 2.5 kcal/mol). Particularly large SDs are observed for the "spikes" at residues D52, W62, N65, T69, and I78, which reveal regions that are highly stabilized by interactions to hydrophobic atoms (Figure S4). Accordingly, only 11% of the HEWL residues show r_i values that differ across all ten curves by < 1.0 kcal/mol; the maximal difference is found for residue M105 (6.7 kcal/mol) in helix E. Note that differences of 1.0 kcal/mol can already lead to misinterpretations of results from rigidity analysis considering that significant differences in the (relative) thermostability of proteins have been mapped to energy differences of the same magnitude (Radestock and Gohlke, 2008, 2011; Rathi et al., 2012). The overall picture does not change when energy-minimized structures are used as input (Figure S1), although the SDs drop to values of ~0.5 (0.9) kcal/mol in the case of helices A, B, and D (helix E and region C). Two "spikes" still prevail. Accordingly, only 30% of the HEWL residues show r_i values that differ across all ten curves by <1.0 kcal/mol. Overall, these results demonstrate a high sensitivity of the rigidity analysis with respect to the input structure. This is remarkable because the structural deviation of the ten HEWL structures is only slightly larger than the uncertainty in the structure determination.

Results of Rigidity Analyses Averaged over ENT^{MD} Are Starting Structure Independent

Previous studies have pointed to the benefit of averaging results from rigidity analysis over ensembles of conformations from MD simulations in terms of a much decreased sensitivity with



(legend on next page)

respect to the input structure (Gohlke et al., 2004; Rathi et al., 2012). To provide benchmark results for subsequent analyses, we generated MD trajectories of 300 ns length starting from each of the ten HEWL structures (Table S1). Our simulation setup and a simulation length of that order provide an accurate representation of HEWL dynamics (Koller et al., 2008). We then repeated the rigidity analyses by CNA, averaging over 1,500 conformations extracted from each of the trajectories (ENT^{MD} approach). As the main outcome, the averaged local stability characteristics are much more consistent (Figures 2B and 2E) than if SNT were analyzed, with the SD of all MD averages < 0.3 kcal/mol in general and < 0.6 kcal/mol in region C. The sole exception concerns residue I78 with SD = 1.1 kcal/mol. Accordingly, 90% of the HEWL residues show r_i values that differ across all 10 curves by < 1.0 kcal/mol; the maximal difference is found at residue I78 in region C (3.3 kcal/mol). Additionally, the stable regions identified in all ensembles for residues Y53, W62–N65, and I78 are in very good agreement with protection factors determined by H/D experiments for HEWL (Radford et al., 1992). In contrast, the stable regions are only identified in five of the ten HEWL structures when using the SNT approach (Figure S4). Additionally, analyzing structures simulated at 300 K leads to larger r_i values (indicating a higher flexibility) for most of the regions of HEWL than in the case of the HEWL crystal structures (Figures 2B and 2D). This finding is important if results from rigidity analysis on biomacromolecules are to be compared to experimental data obtained at room temperature. In summary, averaging over ENT^{MD} leads to robust, i.e., starting structure-independent, results of rigidity analyses.

Parameterizing Fuzzy Noncovalent Constraints Using Data on Breaking and (Re-)Forming of Noncovalent Interactions from MD Simulations

For parameterizing the FNC model, we first analyzed the persistence characteristics of noncovalent bonds during the MD simulations of the ten HEWL structures. Using definitions of noncovalent constraints as in FIRST, we monitored hydrogen bonds (including salt bridges) and hydrophobic tethers. The results show a bimodal distribution of the persistence characteristics of hydrogen bonds, with about 77% persisting for < 60 ns (i.e., 20% of the trajectory length) and 7% being stable almost across the entire simulation (persistence time > 90% of the trajectory length; Figure S2A). In contrast, the majority of hydrophobic tethers (89%) persist only for < 10% of the trajectory length (Figure S2B).

We furthermore analyzed whether hydrogen bonds are more persistent in secondary structure elements such as α helices, 3_{10} helices, or β sheets than in loop regions. Secondary structure elements were identified by DSSP (Joosten et al., 2011) in each conformation extracted from the MD trajectories. The analysis reveals that >85% of the backbone hydrogen bonds in α helices, 3_{10} helices, and β sheets persist in > 80% of the trajectory length (data not shown), in agreement with previous studies (Stickle et al., 1992; Kieseritzky et al., 2006; Almond et al., 2007). In contrast, hydrogen bonds between charged groups (salt bridges) are only present in about 20% of the extracted conformations (data not shown), again in agreement with previous findings (Mamonova et al., 2005).

Table 1. Type-Dependent Probabilities and Standard Deviations of Hydrogen Bond Energies

Type of Hydrogen Bond	$p(HB,t)$ [a,b]	$SD_{HB,t}$ [c]
α Helix (1→5)	0.8	2.0 ± 0.02
3_{10} Helix (1→4)[d]	0.6	1.2 ± 0.03
β Sheet	0.8	1.5 ± 0.04
sp^2-sp^2	0.4	1.6 ± 0.02
sp^2-sp^3	0.3	2.0 ± 0.05
sp^3-sp^2	0.5	1.6 ± 0.04
sp^3-sp^3	0.5	1.5 ± 0.03
Salt bridge	0.8	0.7 ± 0.04

[a]Hydrogen bond energies were computed from geometric parameters (Dahiyat et al., 1997).
[b]Probability with which a hydrogen bond (salt bridge) will be present in generated network topologies.
[c]SD and SEM in kcal/mol.
[d]Including hydrogen bonds in β turns.

To determine to what extent the energy E_{HB} of a hydrogen bond fluctuates, hydrogen bonds with a persistence of >10% of the trajectory length were analyzed. We distinguished between backbone hydrogen bonds of α helices, 3_{10} helices (including β turns), and β sheets. For all other polar interactions, we distinguished hydrogen bonds from salt bridges and further classified hydrogen bonds with respect to the hybridization state of the donor and acceptor atoms (sp^2-sp^2, sp^2-sp^3, sp^3-sp^2, and sp^3-sp^3). $SD_{HB,t}$ (Equation 1) was then calculated from the energies of all hydrogen bonds of type t found in all conformations extracted from the ten independent MD trajectories. Table 1 shows that $SD_{HB,t}$ is type-dependent. The largest fluctuations are found for backbone hydrogen bonds in α helices and hydrogen bonds involving sp^2-sp^3 hybridized donor and acceptor atoms. The lowest fluctuation is found for salt bridges (0.7 kcal/mol). The standard error of the mean (SEM) of $SD_{HB,t}$ is estimated from ten fluctuation values originating from each of the ten independent MD trajectories. The SEM is

Figure 2. Local Stability Characteristics of HEWL

(A) Rigid cluster decompositions using the FIRST program (Jacobs et al., 2001) obtained with a cutoff of the hydrogen bond energy $E_{HB} = -1.0$ kcal/mol and a hydrophobic tether distance cutoff $D_{cut} = 0.25$ Å. Rigid clusters are depicted as uniformly colored bodies with the largest rigid cluster in blue.
(B) Rigidity index r_i for the SNT analyses of the ten HEWL structures (red), r_i curves for the ten ENT^{MD} analyses of HEWL (gray), and the average over all ENT^{MD} analyses (black). The histogram below shows the standard deviation of the r_is across the crystal structures and the MD ensembles, respectively.
(C) Rigidity index r_i for the ENT^{FNC} analyses of the ten HEWL structures (green). The histogram below shows the standard deviation of the r_is. For comparison, the results from the ENT^{MD} analyses are depicted again.
(D–F) The mean r_i values and standard deviations from the SNT analyses (D), ENT^{MD} analyses (E), and ENT^{FNC} analyses (F) are mapped onto a HEWL structure. The colors show the r_i values and the diameter of the putty plot the standard deviation at each residue position. The diameter is scaled with respect to the maximum SD of all three analyses.
See also Table S2 and Figures S1 and S4.

$\leq$ 0.05 kcal/mol for each hydrogen bond type, i.e., it is < 10% even in the case of the lowest fluctuation found for salt bridges.

We next used these results for parameterizing the flickering of hydrogen bonds (including salt bridges) in the FNC model (Part Ia in Theory). Some backbone hydrogen bonds are particularly important for the stability of secondary structures and have a pronounced persistence along a MD trajectory. As such, we required backbone hydrogen bonds in α helices ($1\rightarrow5$) and β sheets to be present in 80% of the generated network topologies (i.e., $p(HB,t)$ = 0.8). Preliminary tests showed that backbone hydrogen bonds in 3_{10} helices ($1\rightarrow4$) should be included in 60% of the generated network topologies. Hydrogen bonds in β turns are treated as those in 3_{10} helices (Baker and Hubbard, 1984). All other sp^2-sp^2, sp^2-sp^3, sp^3-sp^2, and sp^3-sp^3 hydrogen bonds have been found to be less persistent and, thus, are included in 40% (sp^2-sp^2), 30% (sp^2-sp^3), and 50% (sp^3-sp^2, sp^3-sp^3) of the generated network topologies. In total, this leads to average numbers of hydrogen bonds in the generated network topologies that differ by < 4% from the average values of the MD ensembles (Figure S3A). Although hydrogen bonds between charged groups (salt bridges) reveal a low persistence along the MD trajectories, we required that these interactions be present in 80% of the generated network topologies. This still leads to generally lower numbers of salt bridges in the generated network topologies compared to MD results (Figure S3A). However, the difference in the absolute numbers amounts to only two to three salt bridges (i.e., ~5% with respect to all polar interactions) for the HEWL system.

Second, we addressed the effect of thermal motions on computed hydrogen bond energies E_{HB} (Part Ib in Theory) by applying the $SD_{HB,t}$ (Table 1) in Equation 1. Only hydrogen bonds with $E_{HB} < 0$ kcal/mol are included in a new network. Varying E_{HB} that way yields rearranged orders in which hydrogen bonds are removed during a thermal unfolding simulation. Kendall's τ coefficient reveals that the orders in 1,500 networks generated that way are independent from the order in the underlying crystal structure (SNT versus ENT^{FNC} in Table S3; the mean ± SEM over all ten cases is $\tau = 0.029 \pm 0.013$). The same result is obtained if the order of hydrogen bonds is compared between networks generated from 1,500 conformations extracted from MD trajectories and the underlying crystal structure, respectively (SNT versus ENT^{MD} in Table S3; $\tau = 0.019 \pm 0.007$). Finally, the pairwise comparison of the orders of hydrogen bonds in 1,500 networks from ENT^{FNC} versus 1,500 networks from ENT^{MD} also reveals nonexisting correlations on average (Table S3; $\tau = 0.033 \pm 0.013$). However, for all ten independent MD simulations, for > 97% of the networks extracted from the MD trajectory at least one network from ENT^{FNC} is found where the orders of hydrogen bonds significantly ($p < 0.01$) correlate; in these cases $\tau > 0.15$ (Figure S3B). This observation is notable in that it already suggests that ensembles generated from either sampling scheme should lead to similar results in thermal unfolding simulations.

Third, we addressed the less specific character of hydrophobic tethers by favoring tethers at shorter distances over those at longer distances (Part II in Theory). In Equation 2, D_{cut} was set to 0.25 Å as used in the SNT and ENT^{MD} approaches. With these settings, the average numbers of hydrophobic tethers in networks generated by the ENT^{FNC} approach differ by < 14% from those found in networks generated from MD trajectories (Figure S3A).

In summary, our definitions of FNCs for polar interactions and hydrophobic tethers yield noncovalent constraints in ENT^{FNC} derived from single crystal structures that agree very well in terms of ensemble properties with noncovalent bonds identified in structures from MD simulations.

Averaged Results from ENT^{FNC} Agree Almost Perfectly with Those from ENT^{MD}

For validation, we applied the ENT^{FNC} approach on each of the ten energy minimized HEWL structures. Results from CNA were averaged over 1,500 network topologies each. The results of the ENT^{FNC} analyses (Figures 2C and 2F) are considerably more consistent than if a SNT was analyzed, with the SD of all ensemble averages < 1.0 kcal/mol except for residue I78 with SD = 1.9 kcal/mol. Accordingly, 88% of the HEWL residues show r_i values that differ across all ten curves by <1.0 kcal/mol, which is in remarkable agreement with ENT^{MD} results. The largest differences are observed in region C and helix E. For testing the sensitivity of these results on $SD_{HB,t}$ in Equation 1, we repeated the above calculations, once setting the SD to $SD_{HB,t} + SEM_t$ and once to $SD_{HB,t} - SEM_t$ (Table 1). In each case, the results are within the uncertainty of the calculations obtained with $SD_{HB,t}$ across all ten HEWL systems (data not shown). This demonstrates that the ENT^{FNC} results are robust with respect to variations in $SD_{HB,t}$. Compared to the ENT^{MD} results, the identification of known stable regions (Y53, W62–N65, and I78) is less pronounced; still, these regions reveal the highest stability characteristics in region C. Additionally, analyzing ENT^{FNC} topologies leads to larger r_i values (indicating a higher flexibility) for most of the regions of HEWL than in the case of the SNT from the HEWL structures (Figures 2B and 2D). Except for region C, these r_i values are nearly identical to those derived from ENT^{MD} topologies. In summary, averaging over an ENT^{FNC} leads to robust, i.e., less starting structure-dependent rigidity analyses, as observed for the ENT^{MD} approach. The local flexibility and rigidity characteristics also agree almost perfectly with those from ENT^{MD} analyses in terms of the magnitudes of the r_i values.

Validation of the ENT^{FNC} Approach on External Data Sets

We next applied the ENT^{FNC} approach on an external data set not used for the parameterization of the FNC. We investigated five citrate synthase (CS) structures from different organisms with respect to their global stability characteristics by means of thermal unfolding simulations (Rathi et al., 2012). The organisms differ in their optimal growth temperatures (T_{og}), which range from 310.2 K to 373.2 K (Table S4). Four of the CS structures are crystal structures; the fifth (TsCS) has been generated by homology modeling. We generated ENT^{FNC} with 1,500 topologies for each CS structure. As a reference, ENT^{MD} with 1,500 conformations were extracted from MD trajectories of 30 ns length. Additionally, we analyzed either the crystal structures/homology model (SNT) or these structures after energy minimization (SNT^{min}). Phase transitions temperatures T_p were computed from the change of the cluster configuration entropy H_{type2} along the thermal unfolding simulation (Table S4; Pfleger et al., 2013b), making use of a linear relationship parameterized on melting temperatures of pairs of homologs from meso- and thermophilic organisms (Radestock and Gohlke, 2011). Often, melting temperatures are estimated from T_{og} values by assuming

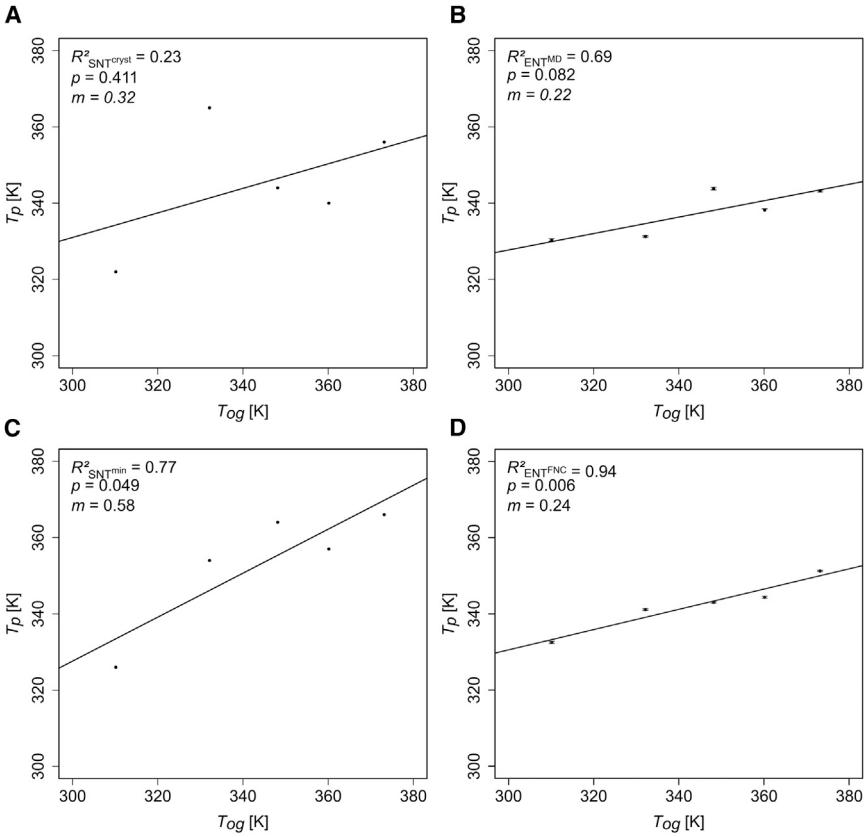


Figure 3. Correlations between Predicted T_p and Optimal Growth Temperatures T_{og}

(A–D) Correlations between predicted T_p and optimal growth temperatures T_{og} from SNT (A), ENT^{MD} (B), SNT^{min} (C), and ENT^{FNC} (D) analyses of five different CS structures. Error bars in (B) and (D) show the SEM. Least squares fit lines have been added.

See also Table S4.

that they are ~25 K higher (Dehouck et al., 2008; Radestock and Gohlke, 2008, 2011); that way, they also provide a relationship between T_p and T_{og}. In the case of analyzing ENT^{FNC} and ENT^{MD}, T_p is averaged over ensembles of 1,500 networks; in both cases, the SEM of T_p is < 0.5 K (Table S4).

No correlation (R^2 = 0.23, p = 0.411) between T_p and T_{og} values is found if SNT were used (Figure 3A); the T_p of the homology model is largely over predicted. If CNA is performed on ENT^{MD}, a fair correlation is found (R^2 = 0.69, p = 0.082; Figure 3B), demonstrating a successful structural refinement of the homology model. However, the CS structures are not correctly ranked with respect to their T_{og} values. The SNT^{min} yield a good correlation (R^2 = 0.77, p = 0.049; Figure 3C), again largely due to a better prediction for the homology model. Still, the ranking of the CS structures is not perfect. The best correlation is obtained if the ENT^{FNC} approach is pursued (R^2 = 0.94, p = 0.006; Figure 3D). Now, all CS structures are correctly ranked with respect to their T_{og} values.

The comparison of the results from the SNT, SNT^{min}, ENT^{MD}, and ENT^{FNC} approaches reveals an influence of the treatment of the CS structures on the slope of the correlation lines. The ensemble-based approaches lead to lower slopes ($m_{ENT^{MD}}$ = 0.22, $m_{ENT^{FNC}}$ = 0.24) than if the crystal structures/homology model are analyzed (m_{SNT} = 0.32). "Heating" structures to 300 K thus seems to obliterate differences in the thermostability. Apparently, the ENT^{FNC} approach implicitly captures part of this temperature effect. The opposite is observed if the structures are "cooled" as in the case of SNT^{min}, resulting in the largest slope $m_{SNT^{min}}$ = 0.58. In view of this, the phase transition temperatures determined from thermal unfolding simulations should be considered relative values only (Radestock and Gohlke, 2011). Still, the temperatures are very helpful, e.g., when it comes to comparing the thermostability of two or more homologous proteins or the stability of a wild-type with its mutant (Radestock and Gohlke, 2008, 2011; Rathi et al., 2012).

All CS are structurally highly similar (rmsd of the C_α atoms: 1.22–2.32 Å) but differ strongly in the pairwise sequence identities (~20% to ~60%). In contrast, for another data set currently under investigation in our group (P.C. Rathi, H.G., unpublished data), mutants with improved thermostability have been generated from a wild-type lipase A (Ahmad et al., 2008; Ahmad and Rao, 2009), resulting in high pairwise sequence identities of the 14 structures (> 93%). This allows us to cite some additional, yet preliminary results. When considering 12 of the 14 structures (the structures with the lowest and highest thermostabilities are treated as outliers), T_p computed by the ENT^{FNC} approach correlate fairly with experimentally determined melting temperatures (R^2 = 0.52, p = 0.008, $m_{ENT^{single}}$ = 0.28). In our opinion, this is a remarkable result as to the predictive power of the ENT^{FNC} approach because the sequences of some of the mutant pairs differ by only one amino acid.

DISCUSSION

We introduced the ENT^{FNC} approach with the aim to improve the robustness of rigidity analyses of biomacromolecules while preserving their computational efficiency. To this end, we provided definitions for FNC based on persistency data of noncovalent bonds derived from MD simulations. With the FNC, an ENT is generated from a single input structure over which results from rigidity analyses are averaged. Thus, by mimicking the flickering of noncovalent bonds, the ENT^{FNC} approach allows performing rigidity analyses on ensembles of network topologies rather than on ensembles of conformations.

The approach was validated at several levels. As to the definitions of fuzzy hydrogen bonds and salt bridges, the network topologies generated with FNC differ by at most 5% from those generated from MD ensembles in terms of the average number of polar interactions. The definition of fuzzy hydrophobic

tethers leads to a related difference of at most 14%. The higher difference in the latter case may reflect that hydrophobic interactions are less specific than polar ones (Mamonova et al., 2005). Furthermore, for almost all networks extracted from an MD trajectory at least one network from ENT^{FNC} is found where the sequences with which hydrogen bonds are removed from a network during a thermal unfolding simulation significantly correlate.

At the next level, we compared results from rigidity analyses on ENT^{FNC} to those on ENT^{MD}; in both cases, the same ten HEWL structures were used as input for the ENT^{FNC} approach or as starting structures for the MD simulations. Remarkably, averaging over an ENT^{FNC} leads to robust, almost starting structure-independent rigidity analyses, as found for the ENT^{MD} approach. Furthermore, local flexibility and rigidity characteristics determined for the ENT^{FNC} agree almost perfectly with those from the ENT^{MD} approach in terms of the magnitudes of the r_i. These findings indirectly confirm the appropriateness of the FNC definitions. Furthermore, they suggest that the ENT^{FNC} approach is viable for overcoming the problem of the sensitivity of rigidity analyses with respect to the input structure.

Finally, we tested the ENT^{FNC} approach for computing relative thermostabilities on data sets of CS and lipase A structures. In both cases, convincing results were obtained. These results are encouraging, for two reasons: (1) both protein systems were not used in the course of parameterizing the FNC. Hence, this demonstrates the transferability of the ENT^{FNC} approach; and (2) both protein systems strongly differ in terms of the extent of the sequence similarity among the members. Yet, even for the series of sequentially highly similar lipase A proteins, the ENT^{FNC} approach shows a high predictive power. This indicates that, while the ENT^{FNC} approach is largely insensitive with respect to small conformational changes of input structures, it remains sensitive enough to pick up effects on the thermostability due to small sequential variations.

A major advantage of the ENT^{FNC} approach over the ENT^{MD} approach is the computational efficiency. As such, the MD simulations for ENT^{MD} required ~16 (8) days of computing time per HEWL (CS) structure on a single NVIDIA Tesla M2070 GPU. Clearly, that way the computational efficiency of a rigidity analysis is compromised. In contrast, the ENT^{FNC} approach only required ~6 min (~2 hr) per HEWL (CS) structure for energy minimization and ~2 hr (~19 hr) per HEWL (CS) structure for the ENT^{FNC} analyses. This amounts to a speed up of a factor of ~4,000 (~100) in the case of HEWL (CS). With that, the ENT^{FNC} approach seems well suited for application in large-scale studies on proteins, e.g., for predicting the effects of site-saturation mutagenesis on thermostability.

As a downside, the ENT^{FNC} approach only mimics the flickering of noncovalent bonds for a given conformation of the biomacromolecule. In contrast, changes in the network due to conformational changes of the biomacromolecule—as additionally detected by the ENT^{MD} approach—will be missed. Thus, the ENT^{FNC} approach is prone to fail if such conformational changes have a determining influence on the biomacromolecule's stability or function. In turn, the ENT^{FNC} approach should be most suitable for comparing biomolecular systems where major conformational changes are not expected. This should be given when comparing homologous proteins (Radestock and Gohlke, 2008, 2011) or wild-type and mutant proteins, making the ENT^{FNC} approach well applicable for data-driven protein engineering. Likewise, the ENT^{FNC} approach could be applied for estimating the influence of ligand molecules on biomolecular stability (Gohlke et al., 2004) if the ligand binding is not accompanied by a large induced fit. In that respect, it is encouraging to note that HEWL structures within one MD trajectory differed by up to 3.3 Å C_α rmsd, yet, local flexibility and rigidity characteristics determined by the ENT^{FNC} approach agree almost perfectly with those from the ENT^{MD} approach.

A few approaches exist that are similar in spirit to the ENT^{FNC} approach. The DCM approach generates an ENT by considering mean-field probabilities of hydrogen bond and torsion constraints in a Monte Carlo sampling. Average stability characteristics are then calculated by rigidity analyses on each topology in the ensemble. While conceptually appealing, a downside of the DCM approach is that it requires experimental data for a protein-specific parameterization of the model (Jacobs et al., 2003; Livesay et al., 2004; Jacobs and Dallakyan, 2005). Recently, the VPG has been introduced, which provides ensemble averaged descriptions of a biomacromolecule's flexibility and rigidity without having to sample multiple network topologies (González et al., 2012). While it is highly efficient, the VPG suppresses fluctuations of network rigidity and, hence, tends to be less accurate at the rigidity percolation threshold where most such fluctuations occur (Gonzalez et al., 2011). This is a drawback when analyzing biomacromolecules considering that they are generally marginally stable (Taverna and Goldstein, 2002), i.e., their network state is close to the rigidity percolation threshold. As a further development, the VPG-x approach improves the accuracy of the description of network rigidity by combining the original VPG with a statistical sampling approach albeit at the cost of losing VPG's efficiency. As all approaches sample over an ENT either directly or indirectly, DCM, VPG-x, and ENT^{FNC} belong to the same computational complexity class. Regarding the representation of noncovalent constraints in these approaches, we see it as an advantage that the FNC defined in this study have been parameterized based on data from state-of-the-art MD simulations. Thus, the definitions should implicitly include solvation and temperature effects. Furthermore, no protein-specific information was used; rather, the definitions are based on hybridization states, atom types, and secondary structure and thus are transferable to other protein systems.

In summary, the ENT^{FNC} approach introduced here has been demonstrated to be a viable approximation to the ENT^{MD} approach for performing ensemble-based rigidity analyses on biomacromolecules in a computationally efficient manner. Our results position the ENT^{FNC} approach for linking biomolecular structure, flexibility, (thermo-)stability, and/or function for large-scale data sets of systems where only limited conformational changes occur. The ENT^{FNC} approach should thus be a valuable complement to the existing approaches for biomolecular rigidity analysis.

EXPERIMENTAL PROCEDURES

Details on the structure preparation of the HEWL and CS systems, the setup and execution of MD simulations of the HEWL and CS systems, and the computation of global and local stability characteristics in the case of the

SNT and ENT^{MD} approaches are given in the Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2013.07.012.

ACKNOWLEDGMENTS

We thank Doris L. Klein, Nadine Homeyer, Daniel Mulnaes, and Britta Nisius for fruitful discussions. We thank Prakash Chandra Rathi for providing the dataset of lipase A structures. H.G. acknowledges insightful conversations with Ileana Streinu, Meera Sitharam, and Mike Thorpe.

Received: April 27, 2013
Revised: July 4, 2013
Accepted: July 17, 2013
Published: August 29, 2013

REFERENCES

Ahmad, S., and Rao, N.M. (2009). Thermally denatured state determines refolding in lipase: mutational analysis. Protein Sci. *18*, 1183–1196.

Ahmad, S., Kamal, M.Z., Sankaranarayanan, R., and Rao, N.M. (2008). Thermostable Bacillus subtilis lipases: in vitro evolution and structural insight. J. Mol. Biol. *381*, 324–340.

Almond, A., Blundell, C.D., Higman, V.A., MacKerell, A.D., and Day, A.J. (2007). Using molecular dynamics simulations to provide new insights into protein structure on the nanosecond timescale: Comparison with experimental data and biological inferences for the hyaluronan-binding link module of TSG-6. J. Chem. Theory Comput. *3*, 1–16.

Baker, E.N., and Hubbard, R.E. (1984). Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol. *44*, 97–179.

Bernadó, P. (2010). Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. Eur. Biophys. J. *39*, 769–780.

Böde, C., Kovács, I.A., Szalay, M.S., Palotai, R., Korcsmáros, T., and Csermely, P. (2007). Network analysis of protein dynamics. FEBS Lett. *581*, 2776–2782.

Cozzini, P., Kellogg, G.E., Spyrakis, F., Abraham, D.J., Costantino, G., Emerson, A., Fanelli, F., Gohlke, H., Kuhn, L.A., Morris, G.M., et al. (2008). Target flexibility: an emerging consideration in drug discovery and design. J. Med. Chem. *51*, 6237–6255.

Crivelli, S., Eskow, E., Bader, B., Lamberti, V., Byrd, R., Schnabel, R., and Head-Gordon, T. (2002). A physical approach to protein structure prediction. Biophys. J. *82*, 36–49.

Dahiyat, B.I., Gordon, D.B., and Mayo, S.L. (1997). Automated design of the surface positions of protein helices. Protein Sci. *6*, 1333–1337.

Dehouck, Y., Folch, B., and Rooman, M. (2008). Revisiting the correlation between proteins' thermoresistance and organisms' thermophilicity. Protein Eng. Des. Sel. *21*, 275–278.

Dodson, G., and Verma, C.S. (2006). Protein flexibility: its role in structure and mechanism revealed by molecular simulations. Cell. Mol. Life Sci. *63*, 207–219.

Dokholyan, N.V., Li, L., Ding, F., and Shakhnovich, E.I. (2002). Topological determinants of protein folding. Proc. Natl. Acad. Sci. USA *99*, 8637–8641.

Forli, S., and Olson, A.J. (2012). A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. J. Med. Chem. *55*, 623–638.

Fulle, S., and Gohlke, H. (2009a). Statics of the ribosomal exit tunnel: implications for cotranslational peptide folding, elongation regulation, and antibiotics binding. J. Mol. Biol. *387*, 502–517.

Fulle, S., and Gohlke, H. (2009b). Constraint counting on RNA structures: linking flexibility and function. Methods *49*, 181–188.

Gohlke, H., Kuhn, L.A., and Case, D.A. (2004). Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach. Proteins *56*, 322–337.

Gonzalez, L.C., Livesay, D.R., and Jacobs, D.J. (2011). Improving protein flexibility predictions by combining statistical sampling with a mean-field virtual pebble game. Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine, 294–298.

González, L.C., Wang, H., Livesay, D.R., and Jacobs, D.J. (2012). Calculating ensemble averaged descriptions of protein rigidity without sampling. PLoS ONE *7*, e29176.

Greene, L.H., and Higman, V.A. (2003). Uncovering network systems within protein structures. J. Mol. Biol. *334*, 781–791.

Halle, B. (2002). Flexibility and packing in proteins. Proc. Natl. Acad. Sci. USA *99*, 1274–1279.

Hammel, M. (2012). Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). Eur. Biophys. J. *41*, 789–799.

Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. Nature *450*, 964–972.

Heringa, J., and Argos, P. (1991). Side-chain clusters in protein structures and their role in protein folding. J. Mol. Biol. *220*, 151–171.

Hespenheide, B.M., Rader, A.J., Thorpe, M.F., and Kuhn, L.A. (2002). Identifying protein folding cores from the evolution of flexible regions during unfolding. J. Mol. Graph. Model. *21*, 195–207.

Huey, R., Morris, G.M., Olson, A.J., and Goodsell, D.S. (2007). A semiempirical free energy force field with charge-based desolvation. J. Comput. Chem. *28*, 1145–1152.

Jacobs, D.J., and Thorpe, M.F. (1995). Generic rigidity percolation: The pebble game. Phys. Rev. Lett. *75*, 4051–4054.

Jacobs, D.J., and Dallakyan, S. (2005). Elucidating protein thermodynamics from the three-dimensional structure of the native state using network rigidity. Biophys. J. *88*, 903–915.

Jacobs, D.J., Rader, A.J., Kuhn, L.A., and Thorpe, M.F. (2001). Protein flexibility predictions using graph theory. Proteins *44*, 150–165.

Jacobs, D.J., Dallakyan, S., Wood, G.G., and Heckathorne, A. (2003). Network rigidity at finite temperature: Relationships between thermodynamic stability, the nonadditivity of entropy, and cooperativity in molecular systems. Phys. Rev. E. Stat. Nonlin. Soft Matter Phys. *68*.

Joosten, R.P., te Beek, T.A., Krieger, E., Hekkelman, M.L., Hooft, R.W., Schneider, R., Sander, C., and Vriend, G. (2011). A series of PDB related databases for everyday needs. Nucleic Acids Res. *39*(Database issue), D411–D419.

Katoh, N., and Tanigawa, S. (2011). A proof of the molecular conjecture. Discrete Comput. Geom. *45*, 647–700.

Kieseritzky, G., Morra, G., and Knapp, E.W. (2006). Stability and fluctuations of amide hydrogen bonds in a bacterial cytochrome c: a molecular dynamics study. J. Biol. Inorg. Chem. *11*, 26–40.

Kleckner, I.R., and Foster, M.P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. Biochim. Biophys. Acta *1814*, 942–968.

Koller, A.N., Schwalbe, H., and Gohlke, H. (2008). Starting structure dependence of NMR order parameters derived from MD simulations: implications for judging force-field quality. Biophys. J. *95*, L04–L06.

Krüger, D.M., Rathi, P.C., Pfleger, C., and Gohlke, H. (2013). CNA web server: rigidity theory-based thermal unfolding simulations of proteins for linking structure, (thermo)stability, and function. Nucleic Acids Res. *41*, W340–W348.

Leach, A.R. (2001). Molecular Modelling: Principles and Applications, Second Edition (Harlow: Prentice Hall).

Levy, R.M., and Karplus, M. (1979). Vibrational approach to the dynamics of an alpha-helix. Biopolymers *18*, 2465–2495.

Livesay, D.R., and Jacobs, D.J. (2006). Conserved quantitative stability/flexibility relationships (QSFR) in an orthologous RNase H pair. Proteins *62*, 130–143.

Livesay, D.R., Dallakyan, S., Wood, G.G., and Jacobs, D.J. (2004). A flexible approach for understanding protein stability. FEBS Lett. *576*, 468–476.

Mamonova, T., Hespenheide, B., Straub, R., Thorpe, M.F., and Kurnikova, M. (2005). Protein flexibility using constraints from molecular dynamics simulations. Phys. Biol. *2*, S137–S147.

Pfleger, C., Rathi, P.C., Klein, D.L., Radestock, S., and Gohlke, H. (2013a). Constraint Network Analysis (CNA): a Python software package for efficiently linking biomacromolecular structure, flexibility, (thermo-)stability, and function. J. Chem. Inf. Model. *53*, 1007–1015.

Pfleger, C., Radestock, S., Schmidt, E., and Gohlke, H. (2013b). Global and local indices for characterizing biomolecular flexibility and rigidity. J. Comput. Chem. *34*, 220–233.

Pratt, L.R., and Chandler, D. (1977). Theory of hydrophobic effect. J. Chem. Phys. *67*, 3683–3704.

Rader, A.J. (2009). Thermostability in rubredoxin and its relationship to mechanical rigidity. Phys. Biol. *7*, 16002.

Rader, A.J., and Bahar, I. (2004). Folding core predictions from network models of proteins. Polymer (Guildf.) *45*, 659–668.

Rader, A.J., Hespenheide, B.M., Kuhn, L.A., and Thorpe, M.F. (2002). Protein unfolding: rigidity lost. Proc. Natl. Acad. Sci. USA *99*, 3540–3545.

Radestock, S., and Gohlke, H. (2008). Exploiting the link between protein rigidity and thermostability for data-driven protein engineering. Eng. Life Sci. *8*, 507–522.

Radestock, S., and Gohlke, H. (2011). Protein rigidity and thermophilic adaptation. Proteins *79*, 1089–1108.

Radford, S.E., Buck, M., Topping, K.D., Dobson, C.M., and Evans, P.A. (1992). Hydrogen exchange in native and denatured states of hen egg-white lysozyme. Proteins *14*, 237–248.

Rathi, P.C., Radestock, S., and Gohlke, H. (2012). Thermostabilizing mutations preferentially occur at structural weak spots with a high mutation ratio. J. Biotechnol. *159*, 135–144.

Rose, G.D., and Wolfenden, R. (1993). Hydrogen bonding, hydrophobicity, packing, and protein folding. Annu. Rev. Biophys. Biomol. Struct. *22*, 381–415.

Sterner, R., and Brunner, E. (2008). The relationship between catalytic activity, structural flexibility and conformational stability as deduced from the analysis of mesophilic-thermophilic enzyme pairs and protein engineering studies. In Thermophiles: Biology and Technology at High Temperatures, F. Robb, G. Antranikian, D. Grogan, and A. Driessen, eds. (London, New York: CRC Press), pp. 25–38.

Stickle, D.F., Presta, L.G., Dill, K.A., and Rose, G.D. (1992). Hydrogen bonding in globular proteins. J. Mol. Biol. *226*, 1143–1159.

Taverna, D.M., and Goldstein, R.A. (2002). Why are proteins marginally stable? Proteins *46*, 105–109.

Teague, S.J. (2003). Implications of protein flexibility for drug discovery. Nat. Rev. Drug Discov. *2*, 527–541.

Thorpe, M.F. (1983). Continuous deformations in random networks. J. Non-Cryst. Solids *57*, 355–370.

Vendruscolo, M., Dokholyan, N.V., Paci, E., and Karplus, M. (2002). Small-world view of the amino acids that play a key role in protein folding. Phys. Rev. E: Stat., Nonlinear. Soft Matter Physiol. *65*, 1–4.

Whiteley, W. (2005). Counting out to the flexibility of molecules. Phys. Biol. *2*, S116–S126.

Young, M.A., Gonfloni, S., Superti-Furga, G., Roux, B., and Kuriyan, J. (2001). Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. Cell *105*, 115–126.

Zaccai, G. (2000). How soft is a protein? A protein dynamics force constant measured by neutron scattering. Science *288*, 1604–1607.

