doi:10.1016/j.jmb.2009.05.068 J. Mol. Biol. (2009) 391, 484–497

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Calculation of Proteins' Total Side-Chain Torsional Entropy and Its Influence on Protein–Ligand Interactions

Kateri H. DuBay and Phillip L. Geissler*

Department of Chemistry, University of California at Berkeley, Berkeley, CA 94720, USA

Chemical Sciences Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA

Physical Biosciences Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA

Received 12 January 2009; received in revised form 20 May 2009; accepted 22 May 2009
Available online 28 May 2009

Edited by M. Levitt

Despite the high density within a typical protein fold, the ensemble of sterically permissible side-chain repackings is vast. Here, we examine the extent of this variability that survives energetic biases due to van der Waals interactions, hydrogen bonding, salt bridges, and solvation. Monte Carlo simulations of an atomistic model exhibit thermal fluctuations among a diverse set of side-chain arrangements, even with the peptide backbone fixed in its crystallographic conformation. We have quantified the torsional entropy of this native-state ensemble, relative to that of a noninteracting reference system, for 12 small proteins. The reduction in entropy per rotatable bond due to each kind of interaction is remarkably consistent across this set of molecules. To assess the biophysical importance of these fluctuations, we have estimated side-chain entropy contributions to the binding affinity of several peptide ligands with calmodulin. Calculations for our fixed-backbone model correlate very well with experimentally determined binding entropies over a range spanning more than 80 kJ/(mol·308 K).

Keywords: side-chain entropy; configurational entropy; side-chain fluctuations; protein–ligand binding; protein thermodynamics

Introduction

Native protein conformations are extremely dense, with packing fractions comparable to those of organic crystals.[1] This observation motivated in early studies of protein structure and dynamics a jigsaw-puzzle notion, in which amino acid side chains of a folded structure become fixed in a unique spatial arrangement by steric interactions with their neighbors. Still today, many computational procedures that explore side-chain packing strive to identify a single native configuration.[2]

The three-dimensional structure of a protein, however, can fluctuate considerably. Large-scale motions involve partial or full unfolding and backbone hinge motions, but subtle structural variation on smaller scales has been highlighted by several experimental measurements. NMR relaxation techniques in particular resolve fluctuations at the level of single bond vectors in both the backbones and side-chains of folded proteins.[3–5] The Lipari–Szabo order parameters[6] they determine, which increase from $S^2_{\mathrm{axis}}=0$ to $S^2_{\mathrm{axis}}=1$ as rotational motion becomes restricted, report on the range of picosecond to nanosecond dynamics for backbone amide and side-chain methyl groups. Computational studies suggest that order parameters lower than 0.8 point to transitions between multiple rotameric states in addition to the inevitable vibrations about optimal torsional angles.[7,8]

Side-chain methyl group order parameters often lie in the range $0.2<S^2_{\mathrm{axis}}<0.8$,[3] indicating extensive exploration of different rotameric states. These results are corroborated by dipolar coupling measurements, suggesting that side chains substantially populate different rotameric states within the ensemble of folded configurations.[9,10] Evidence for alternative side-chain conformations has even been found in electron density maps from crystallography experiments.[11] The data accumulating from such studies paint a consistent picture: Residual side-chain fluctuations in the native-state ensemble are distributed heterogeneously throughout the protein; side-chain bond vectors fluctuate more significantly than do those along the backbone; and the entropy associated

*Corresponding author. E-mail address: geissler@berkeley.edu.

Abbreviations used: MC, Monte Carlo; MD, molecular dynamics; LJ, Lennard–Jones; SB, salt bridge; HB, hydrogen bond; IS, implicit solvent; SASA, solvent-accessible surface area; CaM, calmodulin; PYP, photoactive yellow protein; WL, Wang–Landau.

with such fluctuations is likely to be a significant player in protein thermodynamics.[4,12]

Computational studies focusing on geometric aspects of side-chain packing have reconciled the evidence for significant torsional fluctuations with constraints due to steric interactions in a dense environment.[13] Much as in a dense liquid, volume exclusion reduces the diversity of accessible configurations greatly, but by no means completely. Nearly 10^{20} distinct side-chain conformations were determined to satisfy hard-core constraints in a 125-residue protein with native backbone structure.[13] To what degree non-steric interactions further reduce this variability is not at all clear *a priori*. Populating even a very small fraction of the geometrically acceptable arrangements would be sufficient to allow for significant contributions to free energies of folding and ligand binding.

Mean field theories,[14] various interpretations of molecular dynamics (MD) simulations,[15–18] and several Monte Carlo (MC) approaches[13,19–21] have all been used to estimate the residual entropy of side-chain rotations in folded proteins. Each of these approaches, however, is limited by underlying approximations or formidable practical challenges. Mean field approaches, by definition, do not account for a complete range of thermal fluctuations; straightforward MD simulations can explore only rearrangements that occur on computationally accessible time scales. MC methods are similarly hindered by sampling difficulties intrinsic to such tightly packed systems. As a compromise, entropies are sometimes calculated separately for single residues or small groups of neighboring residues while keeping other residues fixed.[19] Studies that do confront the full combinatorial problem, allowing all side chains to rotate simultaneously, have neglected potentially important contributions from intra-rotameric motions[20] or have considered geometric effects independent of non-steric interactions.[13,21]

In this article, we present a new approach for estimating side-chain torsional entropy. Building on algorithms developed by Kussell *et al.*,[13] our calculations are enabled by enhanced MC methods and a schematic treatment of forces due to sterics, van der Waals interactions, hydrogen bonding, salt bridges (SBs), and solvation. Through this combination, we achieve thorough sampling of thermal fluctuations, incorporate fully coupled rotations of all residues, and address a comprehensive set of physical interactions. Model outlines our approach and the physical perspectives underlying it. Results and Discussion describes applications to a series of small globular proteins, quantifying and comparing the ways in which various forces act to limit rotational freedom.

Within the model we have developed, substantial freedom remains in the packing of side-chains, even in the presence of strong, anisotropic attractions such as hydrogen bonding. The corresponding entropy can, therefore, in principle, strongly influence the thermodynamics of folding, protein–protein binding, and protein–ligand interactions. Indeed, it now appears from calorimetric data that, in several systems, entropy changes figure prominently in tuning protein binding affinities.[22,23] In the case of stromelysin 1 binding to the N-terminal domain of tissue inhibitor of metalloproteinases 1, they even overcome a substantially unfavorable enthalpy of binding.[24] Implicating the involvement of side chains in these phenomena, entropies inferred from NMR order parameters correlate strongly with calorimetrically determined binding entropies for calmodulin (CaM) and several peptide ligands.[12] We find even better agreement between binding entropy measurements and calculated values based on the methods we have developed. This comparison is discussed in detail in Results and Discussion.

Model

In developing a theoretical approach, we are guided by the notion that side-chain rearrangements within a protein's native state are not strongly mediated by motions of the peptide backbone. Physically, we expect that once the molecule has folded, it is subject to global constraints of high packing fraction that vary little with small-amplitude backbone fluctuations. Empirically, we note that correlations observed between backbone NMR order parameters, S^2, and their associated side-chain parameters, S^2_{axis}, are weak.[25] Following Kussell *et al.*,[13] we thus adopt a model in which the peptide backbone is fixed in its crystallographically determined conformation. As a result, applications of our methods are limited to proteins whose native structures have been determined with high resolution.

The sole degrees of freedom in our calculations are dihedral angles χ for rotatable side-chain bonds with heavy-atom (i.e., non-hydrogen) substituents. Other variables are known to influence side-chain entropy,[15] but torsional entropy alone is thought to provide a good approximation.[19] Natural amino acids possess no more than a handful of such dihedral degrees of freedom. Alanine, for example, has none, while lysine and arginine possess the largest number (four). As in a simple molecule such as propane, local bonding energetics bias such angles to lie in one of typically three ranges. For classification purposes, we consider these ranges as discrete rotameric states, each with an ideal angular value θ. We do, however, permit deviations from these ideal angles, $\phi = \chi - \theta$. We and others have found them to be essential for accommodating tightly packed rearrangements.[13,26] The intrinsic energetic penalty $E_{\mathrm{dihedrals}}$ limiting such fluctuations in our model is quadratic in ϕ, except for dihedrals between sp^2 and sp^3 hybridized carbons, where $E_{\mathrm{dihedrals}} = 0$ and is therefore χ-independent. Correspondingly, these bonds possess a single discrete rotamer state.

It is well known from studies of microscopic structure in liquids[27] and polymeric materials[28] that the most essential feature of non-covalent interactions in dense environments is the harsh repulsion between overlapping moieties. Energetic models that discard constraints of volume exclusion in favor of slowly varying potentials for computational convenience[29]

are therefore not suitable for our purpose of quantifying side-chain entropy. Nonetheless, the precise dependence of steric interactions on inter-atomic distances is likely unimportant,[13,30] provided that penetration becomes prohibitively costly at the appropriate length scale. We employ a Lennard–Jones (LJ) potential between all pairs of heavy atoms separated by at least three bonds, which describes van der Waals attractions in addition to imposing steric constraints.[31,32] This interaction is truncated at both small and large distances: For separations larger than twice the LJ diameter, we set the potential energy to zero (and shift the entire potential to maintain continuity at the cutoff); separations smaller than 3/4 the van der Waals contact distance are assigned infinite energy and thus disallowed entirely. This latter modification, introduced for practical reasons, has no physical consequences at reasonable temperatures and densities.

The pairwise interactions we expect to exert the largest influence on side-chain packing are electrostatic in nature, namely, SBs and hydrogen bonds (HBs). We model these energetics based on previous coarse-grained approaches.[33,34] Although we make no effort to represent electrostatic forces between residues in great detail, their strength and anisotropy should be appropriate to the chemical variety of natural amino acids.

Finally, we treat hydrophobic effects in terms of the relative amounts of polar and non-polar surface area exposed to solvent. This simplistic implicit solvent (IS) description does not address the sensitivity of aqueous solvation to the spatial distribution of hydrophobic and hydrophilic moieties at the protein surface,[35,36] but it does roughly account for the many-body nature of such effects. For this purpose, we utilized an inexpensive but faithful approximation to standard procedures for determining solvent-accessible surface area (SASA).[37,38] See Methods for details.

The full potential energy function governing our model sums these various interactions,

$$E(\Theta,\Phi) = E_{\text{dihedrals}} + E_{\text{non-bonded}} + E_{\text{implicit solvent}}. \quad (1)$$

It depends on the set of N torsional angles for all rotatable bonds described above, which we specify through the nearest ideal values, $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$, and deviations about them, $\Phi = \{\phi_1, \phi_2, \ldots, \phi_N\}$. Note that we have collected LJ, SB, and hydrogen-bonding contributions into a total potential $E_{\text{non-bonded}}$ for pairwise-additive, non-bonded interactions. Free parameters in the energies of Eq. (1) were tuned exclusively for the purpose of ensuring that side-chain packing in crystallographic configurations yields energies not much larger than those of alternative arrangements generated in the course of computer simulations. Their values lie well within the range of analogous parameters appearing in other models that attempt a similar level of resolution.

Because it represents steric constraints realistically, our model shares with many other approaches severe challenges to thorough sampling of thermal fluctuations. From typical configurations, it is difficult to rotate a side-chain bond through the ~120° needed to transit from the neighborhood of one ideal angle to another without introducing steric overlaps. In real systems, such an isolated rotation would incur great energetic cost; in our model, the price is often not even finite. We circumvent this problem with MC sampling procedures that preserve the Boltzmann distribution determined by Eq. (1). Specifically, we employ a modified energy function in which the singular hard core of our van der Waals potential is replaced by a finite constant energy $\varepsilon_{\text{tunnel}}$. Correcting exactly for the resulting bias is trivial, since the relative weights of sterically allowed configurations are unchanged. For many purposes, one need only discard sampled configurations that violate steric constraints (see Methods). The advantage of this artifice is an ability to "tunnel" through disallowed regions of configuration space. If $\varepsilon_{\text{tunnel}}$ does not greatly exceed the energy $k_B T$ of typical thermal excitations, simulations can move much more readily through the free-energy barriers that frustrate MD. An optimal value of $\varepsilon_{\text{tunnel}}$ must also ensure that the proportion of sterically inadmissible states generated by MC simulations is not overwhelmingly large. This procedure can enhance sampling efficiency considerably. Several of the calculations we present nonetheless additionally required adaptive umbrella sampling[39] and/or staging through multiple ensembles in which side-chain interactions are gradually introduced (see Methods) to obtain well-converged results with available computing resources.

The model energetics and Metropolis MC methods we have described provide a straightforward and computationally manageable way to characterize side-chain fluctuations quantitatively. By design, our sampling scheme is not dynamically realistic on the time scale of torsional vibrations. Individual trial moves that advance these simulations often switch directly between distinct rotameric states. In the course of natural dynamics, such transitions occur on time scales of picoseconds to milliseconds.[40,41] We have found that MC trajectories comprising 50,000 sweeps are sufficient (but not excessive) for sampling a representative set of side-chain rearrangements in small globular proteins (including on the order of 250 rotatable bonds). Exploring the same range of fluctuations using straightforward MD simulations of detailed atomistic models such as CHARMM or AMBER, which proceed in roughly femtosecond steps, would be extremely taxing if not unfeasible. Indeed, previous MD work suggests that the breadth of side-chain motions cannot be reliably gauged from nanosecond trajectories even for very small proteins.[42]

Results and Discussion

Entropy of side-chain configurations

Absolute entropies are not well defined for continuous classical variables. It is therefore necessary in

computing torsional entropy of a model such as ours to specify a standard state. For this purpose, we choose a noninteracting reference system where all dihedral angles are equally likely, $E^{(\mathrm{ref})}=0$. All entropies we report are given relative to this maximally flexible system, $\Delta^{(\mathrm{ref})}S=S_{\mathrm{config}}-S_{\mathrm{config}}^{(\mathrm{ref})}$, where S_{config} is the configurational entropy associated with fluctuations both within and between distinct rotameric wells. This choice of reference state has several merits. First, a state in which motions of one residue are independent from all others serves as a crude proxy for side-chain fluctuations of an unfolded protein. In other words, $\Delta^{(\mathrm{ref})}S$ could be thought of as a rough estimate for the change in torsional entropy upon folding. Second, by setting $E^{(\mathrm{ref})}$ equal to a constant, we remove all chemical details distinguishing between different rotating moieties. The reference state consequently has an entropy per rotatable bond, $s^{(\mathrm{ref})}$, that is consistent across proteins of arbitrary composition. Thus, while we can determine side-chain entropies only up to an additive constant, $Ns^{(\mathrm{ref})}$, where N is the total number of rotatable bonds, we ensure that $s^{(\mathrm{ref})}$ has the same value for all proteins we consider. Finally, a noninteracting standard state facilitates ligand affinity calculations based on the thermodynamic cycle shown in Fig. 1. Because non-translational free-energy contributions are invariant when two molecules A and B bind in their noninteracting reference states, association entropies can be computed via

$$\begin{aligned}\Delta^{(\mathrm{binding})}S &= S_{\mathrm{A}\cdot\cdot\mathrm{B}}-(S_{\mathrm{A}}+S_{\mathrm{B}})\\ &= \Delta^{(\mathrm{ref})}S_{\mathrm{A}\cdot\cdot\mathrm{B}}-\left(\Delta^{(\mathrm{ref})}S_{\mathrm{A}}+\Delta^{(\mathrm{ref})}S_{\mathrm{B}}\right).\end{aligned}\quad(2)$$

We compute these entropy differences using the corresponding changes in energy and partition function Q,

$$\Delta^{(\mathrm{ref})}S=k_{\mathrm{B}}\ln\left(\frac{Q_{\mathrm{config}}}{Q_{\mathrm{config}}^{(\mathrm{ref})}}\right)+\frac{1}{T}\left(\langle E\rangle-\langle E^{(\mathrm{ref})}\rangle^{(\mathrm{ref})}\right)\quad(3)$$

Angled brackets denote equilibrium averages over canonical ensembles at temperature T. (We perform most calculations at $T=300$ K.) Lacking superscripts, these brackets refer to the Boltzmann distributions determined by the full energy function of Eq. (1); the superscript "(ref)" refers to statistics of the noninteracting reference system. The ratio of partition functions in Eq. (3) could be evaluated using Zwanzig's formula,

$$\begin{aligned}\frac{Q_{\mathrm{config}}}{Q_{\mathrm{config}}^{(\mathrm{ref})}} &= \frac{\sum_{\Theta}\int d\Phi\exp(-\beta E(\Theta,\Phi))}{\sum_{\Theta}\int d\Phi\exp\left(-\beta E^{(\mathrm{ref})}\right)}\\ &= \left\langle\exp\left(-\beta\left[E(\Theta,\Phi)-E^{(\mathrm{ref})}\right]\right)\right\rangle^{(\mathrm{ref})},\end{aligned}\quad(4)$$

where $\beta^{-1}=k_{\mathrm{B}}T$. It is therefore necessary in principle only to sample configurations from the noninteracting system. This approach is not practical, however, since the ensembles defined by E and $E^{(\mathrm{ref})}$ overlap weakly. We overcome this problem with a staging protocol that introduces several intervening ensembles. In these intermediate states, IS and non-bonded interactions are scaled by a parameter $0<\lambda<1$ (see Methods).

Side-chain configurational entropy is commonly discussed in terms of separate contributions from vibrations within a rotameric state (S_{vib}) and from conformational transitions between discrete rotameric states (S_{conf}).[43] Many computational efforts focus exclusively on S_{conf}, even though recent theoretical studies highlight the importance of vibrational entropy changes in ligand binding.[18,44] That a large set of rotameric states becomes accessible only when such vibrations are allowed indicates that these motions are in fact strongly interdependent.[13,26] Our calculations of $\Delta^{(\mathrm{ref})}S$ make no attempt to treat vibrational and conformational contributions separately. We will, however, describe ways to quantify the variability of one motion, while fixing or integrating out the other.

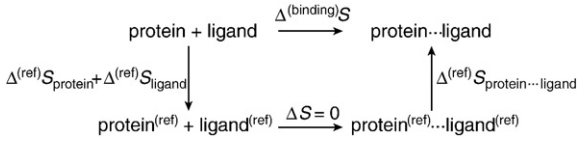

Fig. 1. Thermodynamic cycle relating the change in entropy upon protein–ligand binding, $\Delta^{(\mathrm{binding})}S$, to the entropic differences between the interacting and the noninteracting reference cases for the bound and unbound species. Note that $\Delta S=0$ for the binding of the ligand to the protein within the reference system, allowing $\Delta^{(\mathrm{binding})}S$ to be calculated as shown in Eq. (2).

Entropic losses due to side-chain interactions

We have applied the techniques outlined in the previous section to determine side-chain entropies of 12 small proteins, ranging in size from 46 to 143 residues and exhibiting a diverse set of secondary structures. For each molecule, we have also performed calculations with model energetics that include only a subset of the interaction types described by $E_{\text{non-bonded}}$ and $E_{\text{implicit solvent}}$. In this way, we quantify the extent to which different kinds of forces limit torsional freedom in the dense environment of a folded protein. Results for the entropy reduction per rotatable bond, $\Delta^{(\mathrm{ref})}S/N$, are shown in Fig. 2. For each variant of the model, the similarity of $\Delta^{(\mathrm{ref})}S/N$ values across the entire set of proteins is striking. Local energetic biases due to covalent bonding, described by $E_{\text{dihedrals}}$ and included in all of the interacting systems, result in a significant reduction in entropy. Although weakly dependent on the specific amino acid makeup of the protein, it is found to be quite consistent across the 12 proteins. Of the various interactions considered in isolation, electrostatics yields the largest entropy reduction in most cases. Sterics and van der Waals attractions effect changes similar in magnitude but typically somewhat smaller. Solvation forces, in effect acting only at the periphery of the molecule, contribute least, even though approximately 60% of the residues in

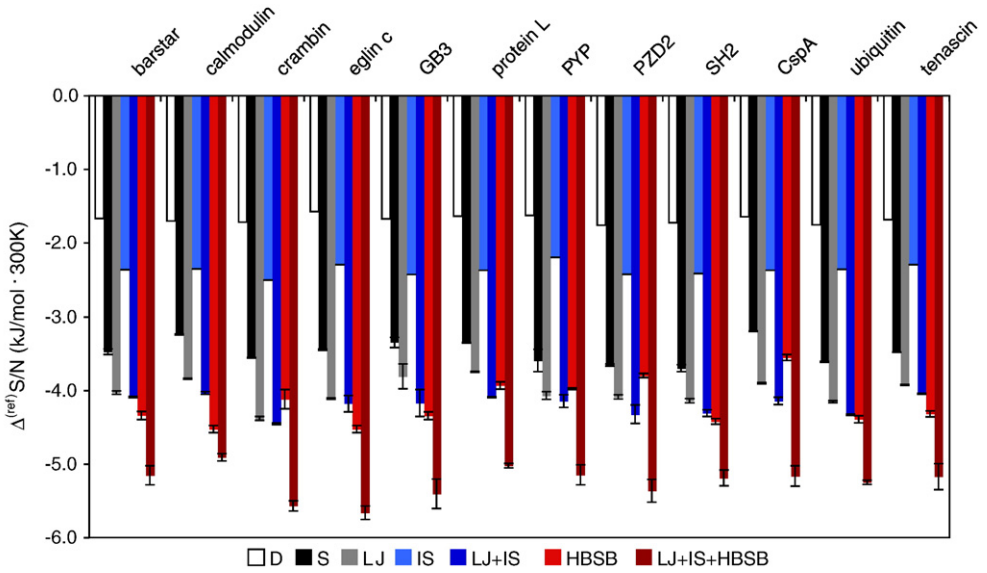

Fig. 2. Total side-chain dihedral entropy per rotatable bond of 12 small proteins, relative to that of a noninteracting reference system [see Eq. (3)]. Results are shown for various combinations of interaction types at 300 K. D refers to a noninteracting reference system that includes only the intrinsic dihedral energy $E_{\text{dihedrals}}$. All other cases include this dihedral potential together with subsets of side-chain interactions: S indicates steric energetics due to the repulsive part of the LJ potential; LJ indicates the full Lennard–Jones potential; IS indicates the implicit solvent; and HBSB indicates the hydrogen bonding and SB interactions. The proteins studied here are barstar (1a19[45]), calmodulin (3cln[46]), crambin (1cbn[47]), eglin c (1cse[48]), GB3 (1igd[49]), protein L (1hz6[50]), PYP (1f9i[51]), PZD2 (1r6j[52]), SH2 (1d1z[53]), CspA (1mjc[54]), ubiquitin (1ubq[55]), and tenascin (1ten[56]). These results were calculated using Metropolis MC.[57] Five trials starting from different randomly chosen side-chain configurations were run for 50,000 MC sweeps each when calculating $\langle E(\Theta, \Phi)\rangle$ and for 17 stages of three 20,000-sweep trials each when calculating Q/Q_{ref}. Error bars represent one standard deviation.

photoactive yellow protein (PYP) are considered solvent-exposed.[13] The CaM structure, consisting of two globular regions connected by an extended α-helix,[46] retains the most entropy. The steric contribution for CaM is among the smallest in this set of molecules, as might be expected from its relatively open structure, but the isolated effects of other interaction types are not at all atypically weak.

Averaging the entropy reduction per rotatable bond over this set of proteins yields $\Delta^{(\text{ref})}S/N = -5.2$ kJ/(mol·300 K). As we have noted, one might regard the noninteracting reference system as a schematic representation of the unfolded state, whose side-chain rotations should be considerably less restricted than in a native fold. A more faithful description of the unfolded state would include the local biases of $E_{\text{dihedrals}}$, which operate regardless of non-covalent structure. Accounting for the corresponding entropic reduction, averaged over the set of proteins we consider, of −1.8 kJ/(mol·300 K), we obtain a typical difference of $\Delta^{(\text{ref,dihedrals})}S/N = -3.6$ kJ/(mol·300 K) between the noninteracting state with restrained dihedrals, denoted by a superscript "(ref,dihedrals)," and the fully interacting state. The change in entropy upon folding due to side-chain conformational fluctuations has been estimated from several different approaches, leading to a consensus figure of ≈−2.1 kJ/(mol·300 K) per rotatable bond.[43] That $\Delta^{(\text{ref,dihedrals})}S/N$ exceeds this value in magnitude is not surprising. Viewed as an approximation of an unfolded protein, our reference state, even with dihedral restraints, certainly overestimates torsional freedom. Further, rigidity of the peptide backbone would likely cause our model to underestimate torsional freedom of the folded state. Despite these limitations, the two values are nonetheless well within k_B of one another. We consider this correspondence an assuring sign that our model captures the basic physical determinants of side-chain entropy correctly.

Over the set of fully globular proteins (which excludes CaM) we have studied, results for $\Delta^{(\text{ref})}S/N$ range from −5.02 kJ/(mol·300 K) for protein L to −5.66 kJ/(mol·300 K) for eglin C. Since the standard state is equivalent in all cases, the range of absolute torsional entropies per rotatable bond, S_{config}/N, is identical in breadth to the range in $\Delta^{(\text{ref})}S/N$. Judging from these 12 proteins, natural variations in native side-chain environments can easily shift torsional entropies by an amount $\delta S_{\text{config}}/N \approx 0.6$ kJ/(mol·300 K). [Note that when the values of $\Delta^{(\text{ref,dihedrals})}S/N$ for protein L and eglin C are compared, the difference is ≈0.7 kJ/(mol·300 K), indicating that this difference is not simply due to differing numbers of rotatable sp^3–sp^3 hybridized bonds.] In the context of protein binding thermodynamics, this result provides a rough gauge for the potential strength of entropic driving forces. If, for example, two globular proteins form a complex whose interface is comparable to the internal structure of typical native folds, overall side-chain entropy may nonetheless change by as much as N (0.6 kJ/(mol·300 K)). For a complex with 100 residues, this maximum change in total entropy would amount to a substantial 102 kJ/(mol·300 K).

Side-chain entropic contributions to CaM–ligand binding

Calorimetry provides unambiguous evidence for strong entropic contributions to protein binding equilibria.[12,24] CaM, for example, binds a series of peptides with similar affinities, but with widely varying entropies of association.[12] For the specific ligand CaMKKα(p), the contribution to the free energy of binding due to entropy alone is nearly 100 kJ/mol, but it is not clear how such entropic changes are distributed among the degrees of freedom associated with solvent, peptide backbone, and amino acid side-chains. The role of side-chain rotations in CaM binding thermodynamics has recently been explored by estimating torsional entropy from NMR order parameters.[12] Although the connection between S^2_{axis} and side-chain entropy is not precise, and although this estimate, of necessity, neglects correlated fluctuations of different residues and fluctuations that take place on time scales longer than those detected in the relaxation experiment, thermodynamic trends were successfully predicted.[12] Specifically, Frederick *et al.* found a linear correlation between calorimetric results for $\Delta^{(\mathrm{binding})}S$ and those computed from NMR data, with a slope of 0.51 and correlation coefficient $r = 0.88$. Side-chain contributions to CaM affinity thus appear considerable.

Our approach provides a way to estimate side-chain contributions to $\Delta^{(\mathrm{binding})}S$ without the assumptions inherent when inferring thermodynamic behavior from NMR order parameters. This CaM–peptide system thus serves as a test both of the methods we have developed and of the notion that torsional fluctuations can play an essential role in peptide binding. We focus on the four peptides considered in Ref. 12 for which calorimetric data[12,58,59] and high-resolution structures are available: CaMKK_{α} (1ckk[60]), smMLCK (1cdl[61]), CaMKI (1mxe[62]), and eNOS (1niw[63]). The thermodynamic cycle in Fig. 1 was used to calculate $\Delta^{(\mathrm{binding})}S$ from our $\Delta^{(\mathrm{ref})}S$ calculations of the bound and unbound CaM and ligand species. The entropy of unbound CaM was computed using the globular structure of Ref. 64 (1prw). The backbone conformation of each ligand when co-crystallized with CaM was used as well for the unbound peptide.

Binding entropies determined by our model match the trend of experimental data as well as their overall scale, as shown in Fig. 3. In particular, the experimental order $\Delta^{(\mathrm{binding})}S_{\mathrm{CaMKK\alpha(p)}} < \Delta^{(\mathrm{binding})}S_{\mathrm{smMLCK(p)}} < \Delta^{(\mathrm{binding})}S_{\mathrm{CaMKI(p)}} < \Delta^{(\mathrm{binding})}S_{\mathrm{eNOS(p)}}$ is correctly reproduced, although the difference between smMLCK(p) and CaMKKα(p) cannot be resolved within statistical errors. Correlation between computed values and experimental measurements exceeds reasonable expectations, given our exclusive focus on side-chain contributions and neglect of backbone fluctuations. We emphasize that model parameters were not adjusted to obtain this agreement. Neither is the correspondence a trivial consequence of peptide size and composition; the complexes we have studied possess similar numbers of rotatable bonds (between $N = 281$ and $N = 286$) and rank differently by N and by $\Delta^{(\mathrm{binding})}S$. Furthermore, calculations employing reduced sets of interaction types in many cases compare poorly with experiment. Our results thus bolster the conclusion of Ref. 12 that side-chain torsional rearrangements constitute a major, if not dominant, source of CaM binding entropy.

Heterogeneous distribution of side-chain entropy

Though ordered, a folded protein is structurally heterogeneous on all scales from atomic to macromolecular. One might expect that the rotational freedom of side chains is similarly nonuniform. Indeed, the fluctuation spectrum of a protein's interior has been likened to that of a solid, while the exposed surface is often considered fluid.[65] Our calculations reveal spatial patterns of torsional variability that are not nearly as simple as this conjecture would suggest.

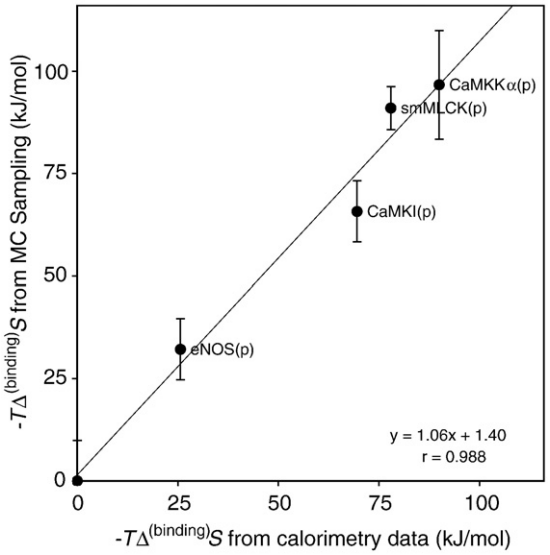

Fig. 3. Entropic contributions to the binding free energies $-T\Delta^{(\mathrm{binding})}S$ for four CaM–peptide complexes. Results of MC simulations are plotted against corresponding calorimetric measurements from Ref. 12. Unbound CaM is shown on the plot as the reference point at (0,0). For the CaM–peptide complexes, $\langle E(\Theta, \Phi)\rangle$ was calculated using a WL bias in six sets of 10 trials, each with 90,000–100,000 sweeps. Average values were calculated within each set and errors were calculated across the six sets. For unbound CaM and peptides, Metropolis MC was sufficient to calculate $\langle E(\Theta, \Phi)\rangle$, and 5 trials of at least 50,000 sweeps each were performed. A WL bias was also used to calculate $\langle \Delta(C)\rangle^{(\mathrm{tunnel})}$ and the first ratio of partition functions on the right-hand side of Eq. (20) in three sets of 10 trials for each of the CaM–peptide complexes. Metropolis MC was used to calculate the remaining 22 stages in the Q/Q_{ref} calculation of the CaM–peptide complexes, as well as the full Q/Q_{ref} calculation for unbound CaM (in 26 stages) and the unbound peptides (in 17 stages). Each stage included 3 trials of 20,000 sweeps. Averages and errors were calculated between the three independent calculations of Q/Q_{ref}. Error bars represent one standard deviation. Errors in the calorimetric measurements of $T\Delta^{(\mathrm{binding})}S$ are ≤ 1.0 kJ/mol.[12]

We do find, on average, that side chains of surface residues are less tightly constrained by native interactions than are those of the interior. However, there are many exceptions, and simple features of crystallographic structures such as secondary structure and packing density do not reliably foreshadow the extent of local side-chain fluctuations.[25]

As a measure of local torsional variability, we consider the Gibbs entropy $\overline{S}_i^{(\mathrm{res})}$ associated with a single residue's notionally discrete rotameric states,

$$\overline{S}_i^{(\mathrm{res})} = -k_{\mathrm{B}} \sum_{\Theta_i} p(\Theta_i) \ln p(\Theta_i), \qquad (5)$$

where $\Theta_i = \{\theta_1^{(i)}, \ldots, \theta_{N_i}^{(i)}\}$ denotes the set of ideal torsion angles for each of the N_i rotatable sp^3–sp^3 hybridized side-chain bonds belonging to residue i. The populations $p(\Theta_i)$ of these 3^{N_i} states are determined in simulations by constructing a histogram over sterically allowed configurations. Effectively integrating out torsional fluctuations $\Phi_i = \{\phi_1^{(i)}, \ldots, \phi_{N_i}^{(i)}\}$ within each ideal rotameric state, we focus on discrete degrees of freedom with a manageable set of possible realizations. As a result, we can calculate converged, absolute values of $\overline{S}_i^{(\mathrm{res})}$. This analysis focuses explicitly on "conformational" contributions to entropy. Others have calculated analogous quantities for different models,[43] some lacking vibrational fluctuations altogether.[21] In our calculations, coupling between conformational and vibrational motions, and between rearrangements of different residues, is implicit in the weights $p(\Theta_i)$.

We have computed $\overline{S}_i^{(\mathrm{res})}$ for all residues in each of the small proteins listed in Fig. 2, and for several subsets of the interaction types in Eq. (1). Here, we present and discuss in detail results only for PYP, whose behavior is typical of the entire set of molecules. Figure 4 illustrates the complex spatial distributions of rotational freedom generated by our model. It also demonstrates that different interaction types limit side-chain rearrangements in different ways.

Values of $\overline{S}_i^{(\mathrm{res})}$ are indicated in Fig. 4 by the coloring of residues within PYP's three-dimensional structure. Although side chains are shown in their crystallographic configurations, it is fluctuations away from this ideal packing that determine the local entropies depicted. The color scale varies by residue according to its maximum possible value of $\overline{S}_i^{(\mathrm{res})}$. Bright red corresponds to this maximum value $\left(\overline{S}_i^{(\mathrm{res})} = k_{\mathrm{B}} N_i \ln 3\right)$, while dark blue signifies an absence of rotamer variability $\left(\overline{S}_i^{(\mathrm{res})} = 0\right)$. Residues that possess no rotatable bonds are colored blue, though Eq. (5) is not well defined in this case.

Results for our noninteracting reference system are shown in Fig. 4a. Lacking any bias on side-chain configuration, all residues with rotatable bonds exhibit their maximum local entropy and are thus colored red. Figure 4b–e correspond to different subsets of interaction types, each including the basic local energetics $E_{\mathrm{dihedrals}}$ of torsional rotations. Figure 4f shows results for the full model potential of Eq. (1).

Of the interaction types we consider, the combination of steric constraints and van der Waals attractions effects local entropy in ways most similar to the solid/liquid caricatures of a protein's interior/exterior (see Fig. 4b). However, even in this case, the entropic distinction between exposed and buried

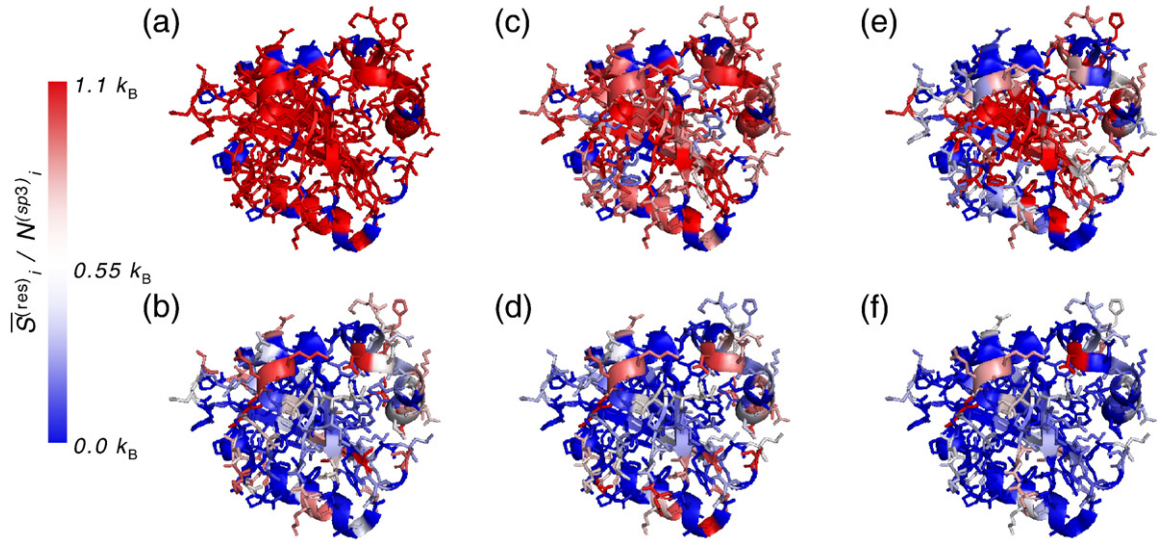

Fig. 4. Side-chain conformational entropy, $\overline{S}_i^{(\mathrm{res})}$ [see Eq. (5)], for all residues i in PYP (1f9i[51]) for various kinds of interactions. The side-chains are color coded according to each residue's value of $\overline{S}_i^{(\mathrm{res})}$, with red indicating the residue's maximum entropy and blue indicating its minimum. $\overline{S}_i^{(\mathrm{res})}$ values have been normalized by the number of sp^3–sp^3 hybridized rotatable bonds. (a) Noninteracting reference system (blue residues indicate amino acids without rotatable bonds), (b) LJ interactions, (c) IS interactions, (d) both LJ and IS interactions, (e) HB and SB interactions, and (f) all interactions. All interacting runs include the effects of $E_{\mathrm{dihedrals}}$. Results were calculated using Metropolis MC for five independent trials, each run for 50,000 MC sweeps. Images were made using MacPyMOL.[66]

residues is not clear-cut. By itself, the IS energy has an opposing effect, significantly limiting the motion of only those residues that can be readily accessed by solvating water molecules (see Fig. 4c). Electrostatic interactions exert a rather different influence on patterns of torsional freedom (see Fig. 4e), in isolation affecting only those residues that donate/accept HBs or participate in SBs. The directionality of these forces, as well as the fact that HB partners may reside on the peptide backbone, begets restrictions on side-chain motion that are, in general, much more localized and anisotropic than those due to other interaction types. The net effect of all these interactions, when operating simultaneously, is a local entropy much reduced from that of the reference state and distributed throughout the structure much less smoothly than would be expected from the notion that buried residues adopt unique rotamer states (see Fig. 4f).

These same interactions restrict torsional vibrations as well, whose variety is essentially overlooked by the local entropy of Eq. (5). This neglect is reasonable for assessing rotamer flexibility in qualitative terms but does not suffice for quantifying the magnitude of entropic driving forces. As an example, the total side-chain rotational entropy of a protein can be estimated by summing local entropies over all residues, $\overline{S} = \sum_i \overline{S}_i^{(\mathrm{res})}$. Discarding contributions from vibrational fluctuations in this manner diminishes computed CaM-peptide binding entropies by nearly 60%. Nevertheless, the reduction of $T\Delta^{(\mathrm{binding})}\overline{S}$ estimates is consistent in magnitude across the peptide ligands we have studied, so that correlation with experimental data remains strong, with a correlation coefficient of $r = 0.96$.

In a separate approach to quantifying the importance of torsional vibrations, we consider a new reference state, denoted by "(ref′,dihedrals)," in which rotamers do not interact but are nonetheless constrained to a single set of ideal dihedral angles Θ and are governed by the dihedral potential. We can thus estimate the loss of entropy $\Delta^{(\mathrm{ref}',\mathrm{dihedrals})}S_{\mathrm{config}}$ solely due to restrictions on vibrational motion resulting from interactions. These values closely mirror the results shown in Fig. 2, but on a scale smaller by roughly a quarter.

Entropies of the reference systems we have considered are simply related, $S_{\mathrm{config}}^{(\mathrm{ref})} - S_{\mathrm{config}}^{(\mathrm{ref}')} = k_{\mathrm{B}}N^{(sp^3)}\ln 3$, where $N^{(sp^3)}$ is the total number of rotatable sp^3–sp^3 hybridized bonds. A thermodynamic cycle can thus be used to connect interacting systems differing by constraints on ideal dihedral angles. In this way, we find that fixing Θ in the fully interacting model effects an entropy loss of ≈ 0.8 kJ/(mol·300 K) per rotatable sp^3–sp^3 hybridized bond.

Comparisons to experimentally determined side-chain fluctuations

Comparisons between these detailed local entropies and experimental data are ambiguous in several respects. As we have noted, the quantity $\overline{S}_i^{(\mathrm{res})}$ discards contributions of torsional vibrations we have found to be numerically significant. On the experimental side, currently feasible measurements at this level of resolution can only be related to thermodynamics in approximate ways. NMR order parameters, for example, are sensitive only to the range of rotational fluctuations that function on picosecond to nanosecond time scales.[3] Dipolar coupling and J-coupling experiments report on longer time scale motions, but for side-chains, they are generally only applied to the rotatable bond closest to the peptide backbone (whose dihedral angle is denoted χ_1).[67] Despite these limitations, we employ methyl order parameters and χ_1 rotamer populations as rough points of comparison for our computer simulations.

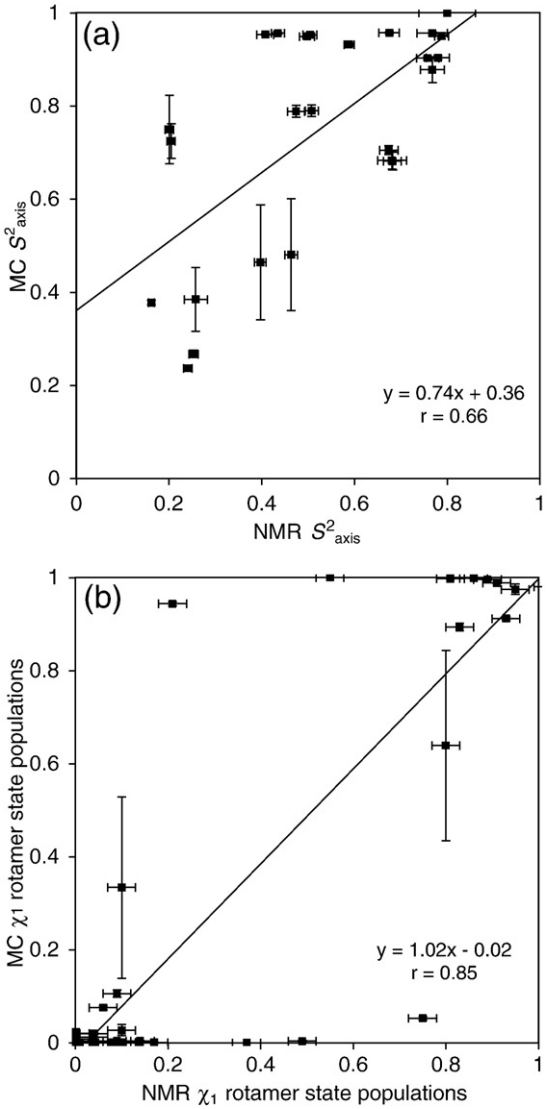

Fig. 5. Side-chain NMR order parameters, S^2_{axis}, and χ_1 rotameric populations for eglin c (1cse[48]). Results of MC sampling plotted against NMR-derived measurements. (a) Comparison of the MC and NMR[10] methyl group order parameters. (b) Comparison of χ_1 rotamer state populations determined from MC sampling and experimental three-bond J-coupling constants.[10] Five independent trials were run for 50,000 sweeps each using Metropolis MC. Error bars represent one standard deviation.

Figure 5 presents results for the specific protein eglin c, for which both methyl group order parameters and χ_1 rotamer populations have been experimentally determined.[10]

Values of S^2_{axis} derived from NMR data[10] and those calculated from MC simulations (using the approach described in Ref. 68), both shown in Fig. 5a, are modestly correlated ($r=0.66$). Previous simulation results obtained from 50-ns MD trajectories for a detailed model of calbindin have matched experimental measurements more closely, but not dramatically so ($r=0.8$).[42] Numerical calculations for eglin c that utilize a sampling procedure inconsistent with Boltzmann statistics generate still stronger correlation,[69] perhaps highlighting the sensitivity of S^2_{axis} to very sluggish rearrangements. The result of such comparisons indicates that side-chain fluctuations are overly restricted in our model, as might be expected from the neglect of backbone flexibility. Alanine orientation, for example, is completely fixed in our model, yielding $S^2_{axis}=1$ identically. The NMR result for alanine in eglin c, $S^2_{axis}=0.8$,[10] points to non-negligible effects of backbone motion, although such effects appear to correlate weakly with measured side-chain fluctuations.[25]

Populations of distinct χ_1 rotamer states inferred from experiment[10] also agree reasonably (but not strikingly) well with results from our simulations (see Fig. 5b). The dearth of probabilities between 0.1 and 0.9 indicate that these bonds are strongly biased toward one rotameric state. This fact should not, however, be taken as a sign of overall torsional rigidity. Bonds that are not proximal to the backbone show greater variability. Indeed, in a typical configuration of our model, roughly one-sixth of the rotatable sp^3–sp^3 hybridized side-chain bonds in eglin c adopt an ideal dihedral angle θ_i different from the most probable.

Importance of model interactions and thorough sampling

The high correlation between experimental and calculated $T\Delta^{(binding)}S$ values in the CaM–ligand system suggests that our model includes the interactions most essential for describing side-chain fluctuations within the folded protein. We emphasize the importance of considering energetics beyond those imposing steric constraints, despite the dense environment; when non-steric interactions are omitted, calculated entropies correlate only moderately with calorimetric measurements. Similarly, neglecting inter-residue correlations and intra-rotameric fluctuations substantially reduces the quantitative correspondence with experimental data. These results strongly recommend models of side-chain thermodynamics that include intra-rotameric fluctuations[18,44] and respect not only constraints of packing but also the diversity and broad energy spectrum of sterically allowed configurations.

Our MC sampling methods probe diverse side-chain configurations that may be difficult to access using more straightforward sampling methods. Notably, NMR order parameters, S^2_{axis}, estimated from a 5-ns MD trajectory of barstar resemble our MC results more closely than do those determined from only 250 ps of time evolution.[4] The diversity of side-chain packings we have identified suggests an important role for still slower fluctuations. Even with our MC sampling procedure, obtaining converged results for CaM–peptide $T\Delta^{(binding)}S$ values requires the implementation of advanced techniques such as staging and the use of Wang–Landau (WL) procedures. This necessity highlights the limitations associated with calculating entropies from MD simulations alone, as has been attempted previously.[15–18] In one study on protein–protein binding, several shorter MD trajectories were run in order to improve the convergence of calculated binding entropies, but the errors were still quite large.[70]

Combining MC and MD techniques might provide an optimal approach for exploring structural excursions broadly while preserving the dynamical character of short-time relaxation.[71] Capturing the time dependence of the slowest side-chain rearrangements, which in our model must navigate severe dynamical bottlenecks, will likely require importance sampling in trajectory space.[72]

Conclusions

We have examined spontaneous side-chain fluctuations in several folded proteins using computer simulations that sample all side-chain torsional degrees of freedom simultaneously. Overall, our MC method facilitates exploration of rearrangements that proceed sluggishly in the course of natural dynamics, and our model appears to successfully capture the physical character of these variations. Their extent is likely underestimated due to backbone constraints, rendering conclusions about their thermodynamic significance conservative.

We have assessed the impact of various interaction types in restricting the range of side-chain motions, by quantifying entropy reductions relative to a noninteracting reference system. The ability to probe these interactions separately is a strength of our computational approach that would be difficult to mimic experimentally. These reductions, normalized by the number of rotatable bonds, are remarkably consistent among the 12 proteins we have considered, despite significantly heterogeneous distributions of rotational freedom. Under the collected influence of steric, dispersive, and electrostatic forces, globular proteins in our model possess, on average, an entropy per rotatable bond of 5.2 ± 0.2 kJ/(mol·300 K) less than their noninteracting counterparts.

Our binding entropy calculations for CaM–peptide complexes, which correlate strongly with calorimetric measurements, underscore the thermodynamic importance of side-chain torsional freedom. They also hint at the possibility that correlated side-chain fluctuations could communicate structural change over significant distances. Indeed, NMR studies show that effects of side-chain mutation or

ligand binding on side-chain methyl dynamics can extend far from the site of perturbation.[10,73] The computational tools we have presented are well suited to explore this unconventional mechanism for protein allostery.

Methods

Model

The potential energy function governing side-chain fluctuations in our model is a sum of three physically distinct contributions: from the local torsional bias of covalent bonding ($E_{\text{dihedrals}}$), from direct interactions between non-bonded moieties ($E_{\text{non-bonded}}$), and from the free energy of aqueous solvation ($E_{\text{implicit solvent}}$).

The local dihedral energy $E_{\text{dihedral},i}$ of a rotatable bond i depends on its hybridized geometry. Since ideal angles are difficult to identify for sp^3–sp^2 hybridized rotatable bonds,[74] we impose no intrinsic bias on the corresponding rotations, that is, $E_{\text{dihedral},i}=0$ independent of χ_i for these bonds. For the more prevalent sp^3–sp^3 rotatable bonds, the dihedral energy function is constructed so that the Boltzmann weight $\exp(-\beta E_{\text{dihedral},i})$ is a sum of (unnormalized) Gaussian distributions centered at ideal rotamer angles θ_i,

$$E_{\text{dihedral},i} = -k_B T \ln\left(\sum_{\theta_i} \exp\left[-\frac{(\chi_i-\theta_i)^2}{2\sigma^2}\right]\right) \qquad (6)$$

We parameterize this function through the approximate width of empirical distributions of side-chain torsional rotations, $\sigma=12.7°$, as found in the rotamer library.[74] For each sp^3–sp^3 bond, three ideal values of θ_i are assigned using data from Ref. 74 (see Supplemental Material). Since the range of χ_i is unbounded in our simulations, each of these three ideal values is in fact repeated with a period of 2π; that is, Gaussian distributions in Eq. (6) are centered at θ_i, $\theta_i\pm2\pi$, $\theta_i\pm4\pi$,.... The strongest overlap among these distributions is between neighboring ideal angles; however, in practice, σ is sufficiently small compared to the spacing between ideal values that the overlap is extremely weak. Neglecting this overlap entirely, we could consider $E_{\text{dihedral},i}$ as a piecewise continuous superposition of quadratic functions centered at each ideal rotamer angle. With this approximation, a protein's total intrinsic dihedral energy can be written

$$E_{\text{dihedrals}}(\Phi) \approx k_B T \sum_i \frac{\phi_i^2}{2\sigma^2} h_i, \qquad (7)$$

where $\phi_i = \min_{\theta_i}(\chi_i-\theta_i)$ is the deviation of dihedral angle χ_i from its nearest ideal rotamer angle, θ_i. The indicator function h_i takes values of $h_i=1$ if bond i is sp^3–sp^3 hybridized and $h_i=0$ if bond i is sp^3–sp^2 hybridized. The exact function $E_{\text{dihedrals}}=\sum_i E_{\text{dihedral},i}$ is a slightly smoothed version of Eq. (7), more closely resembling the detailed dihedral potentials used in CHARMM and AMBER.

Our model includes non-bonded interactions due to sterics and van der Waals attractions, due to SBs, and due to hydrogen bonding:

$$E_{\text{non-bonded}}(\Theta,\Phi) = \sum_{i\neq j}\left[L_{ij}(r_{ij}) + \frac{q_i q_j}{K(r_{ij})r_{ij}} + H_{ij}(r_{ij},\Psi)\right]. \qquad (8)$$

We denote the distance between heavy atoms i and j as r_{ij}, their charges as q_i and q_j, and the set of angles describing their HB geometry as Ψ. The factor $K(r)=0.124r$ Åmol/kJ accounts empirically for the screening of ionic interactions in the heterogeneous environment of a protein's interior.[31,34]

We represent steric as well as dispersion interactions between heavy atoms using a modified LJ potential

$$L_{ij}(r_{ij}) = \begin{cases} \infty, & r_{ij}<r_{ij}^* \\ \epsilon_{ij}\left[\left(\frac{r_{ij}^{\min}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^{\min}}{r_{ij}}\right)^{6} + \alpha\right], & r_{ij}^*\le r_{ij}<2\sigma_{ij} \\ 0, & r_{ij}\ge 2\sigma_{ij}. \end{cases} \qquad (9)$$

We set the distance $r_{ij}^{\min}$ of minimum energy to be the sum of van der Waals radii for atoms i and j (taken from Ref. 32). Attraction strengths ε_{ij} are taken from Ref. 31. The smooth decay at long distances is truncated at $r_{ij}=2\sigma_{ij}$, where $\sigma_{ij}=(1/2)^{1/6}r_{ij}^{\min}$, and the entire potential is shifted by the constant $\alpha=0.0615$ so that the potential is continuous; that is, $\lim_{r_{ij}\to 2\sigma_{ij}^-} L_{ij}=0$. We describe the harsh repulsion at short distances ($r_{ij}^*=0.75r_{ij}^{\min}$) with a hard sphere potential (rather than the sharp but smooth r^{-12} of LJ) for sampling purposes as described below. Also, toward that end, we define a non-singular version of the steric interaction

$$L_{ij}^{(\text{tunnel})}(r_{ij}) = \begin{cases} \epsilon_{\text{tunnel}}, & r_{ij}<r_{ij}^* \\ L_{ij}(r_{ij}), & r_{ij}\ge r_{ij}^*. \end{cases} \qquad (10)$$

The superscript "(tunnel)" accompanying other quantities indicates usage of $L^{(\text{tunnel})}$ in place of L. Steric repulsions and dispersion attractions are only considered for atoms separated by at least three bonds.

Hydrogen bonding between the donors and acceptors specified in Table 1 of Ref. 33 is described by a potential adapted from Ref. 31,

$$H_{ij}(r_{ij},\Psi) = \begin{cases} D_0\left[5\left(\frac{r_{ij}^{\text{HBmin}}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij}^{\text{HBmin}}}{r_{ij}}\right)^{10} + \eta\right]F(\Psi), & r_{ij}^{\text{HB*}}\le r_{ij}<4.0\text{Å} \\ 0, & \text{otherwise.} \end{cases} \qquad (11)$$

The strength $D_0=18$ kJ/mol of a perfectly aligned HB was chosen such that the total energies of crystallographic structures lie within the energy range of typical repacked structures. Averaged over fluctuations in donor–acceptor geometry, the resulting dissociation energy amounts to roughly 11 kJ/mol when the full potential is considered for PYP. The donor–acceptor distance $r_{ij}^{\text{HBmin}}=2.75$ Å of minimum hydrogen-bonding energy was taken from Ref. 31. $r_{ij}^{\text{HB*}}$ is set to 2.52 Å, and the entire potential is again shifted by a constant $\eta=0.0858$ to preserve continuity. Orientation dependence of this model potential is determined by a set Ψ of three angles. In terms of the unit vectors $\hat{u}_{\text{DA}}$ pointing from the donor D to the acceptor A, $\hat{u}_{\text{DD}'}$ pointing from the donor to its nearest bonded heavy atom D′, and $\hat{u}_{\text{AA}'}$ pointing from the acceptor to its nearest bonded heavy atom A′, these angles are defined as $\psi_{\text{D}}=\cos^{-1}(\hat{u}_{\text{DA}}\cdot\hat{u}_{\text{DD}'})$ and $\psi_{\text{A}}=\cos^{-1}(\hat{u}_{\text{AD}}\cdot\hat{u}_{\text{AA}'})$. ψ_{n} is the angle between the normals of the planes defined by (D,D′, D″) and (A,A′,A″), where A″ and D″ are the next antecedent heavy atoms, bound either to the acceptor's or the donor's nearest bound neighbor or to the acceptor

or donor itself. See Ref. 33 and Fig. 1 for details.[31,33,34] We take their influence to be multiplicatively separable,

$$F(\psi_D, \psi_A, \psi_n) = f_D(\psi_D) f_A(\psi_A) g_n(\psi_n), \quad (12)$$

where

$$f_D(\psi_D) = \begin{cases} \cos^2(\psi_D - \psi_D^*) & 90^\circ \le \psi_D \le 180^\circ \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

For donors that are sp^2 hybridized, $\psi_D^* = 120°$, while for sp^3 donors, $\psi_D^* = 109.5°$. The function $f_A(\psi_A)$ differs from $f_D(\psi_D)$ only in the range $60° \le \psi_A \le 180°$ over which it is nonzero, and only for sp^3 hybridized acceptors. Finally,

$$g_n(\psi_n) = \begin{cases} 0, & \psi_n > 60^\circ \text{and the donor is } sp^2 \text{ hybridized} \\ 1, & \text{otherwise.} \end{cases} \quad (14)$$

Hydrogen bonding between protein and solvent is allowed when a side-chain donor or acceptor has not formed its maximum number of HBs[33] with other protein donors or acceptors. Contributions of these bonds to the pairwise interaction energy $E_{\text{non-bonded}}$ are small, favorable by exactly 2 kJ/mol in all cases, with no distance or angular component. More substantial effects of these bonds are subsumed in $E_{\text{implicit solvent}}$, whose strength is determined by the energy of sequestering non-polar atoms from solvent by exposing polar moieties instead.

We represent solvent–protein interactions primarily according to the composition of SASA

$$E_{\text{implicit solvent}}(\Theta, \Phi) = \gamma A_{\text{non-polar}}(\Theta, \Phi), \quad (15)$$

where $\gamma = 0.3$ kJ/molÅ^2 is the surface tension of a hydrocarbon–water interface.[37] The exposed non-polar area $A_{\text{non-polar}}(\Theta, \Phi)$ is calculated using a computationally inexpensive implementation of the Shrake–Rupley algorithm.[38] Fifty points are placed at uniform density on a sphere centered at each heavy atom, with a radius R equal to the sum of its van der Waals radius and that of a water molecule. We then determine the fraction x of such points that lie outside all spheres centered on neighboring atoms. A non-polar atom's contribution to SASA is computed as $4\pi R^2 x$. For the crystal structure of PYP, this estimate differs from values obtained with the more taxing but exact method GETAREA[75] by only 1.5% of the total surface area for typical heavy atoms and by 1.0% of the total surface area for the final value of $A_{\text{non-polar}}$.

Within the framework of this model, glycines, alanines, and prolines possess no degrees of freedom and therefore cannot contribute to the overall entropy. In addition, residues that participate in disulfide bridges, those residues binding to Ca^{2+} in CaM, and the residue attached to the chromophore in PYP are considered to have no rotatable bonds. All bond lengths and angles are taken directly from the Protein Data Bank structures for each protein, and in all cases when more than one structure is resolved, the first structure is always used. Within the crystalline unit, the most complete structures were used. In crystal structures where non-standard amino acids are used to assist in crystallization or phasing, we mutate those residues back to standard amino acids before sampling. Unresolved residues at the N- or C-termini were not included in the modeling, while unresolved side chains or atoms in among the resolved portion of each protein were arbitrarily assigned appropriate initial positions.

We developed this model using only PYP and protein L for testing and refining. No potential refinement was done to optimize results for eglin c, CaM, or CaM–ligand complexes.

Sampling

All computer simulations were performed in canonical ensembles permissive of steric overlaps, that is, according to the regularized potential $E^{(\text{tunnel})}$. Physical quantities of interest must be calculated for the full potential E, which, of course, precludes steric clashes. We have constructed these two potentials such that converting computed averages $\langle\cdot\rangle^{(\text{tunnel})}$ of an arbitrary observable $\cdot$ into physically realistic averages $\langle\cdot\rangle$ is a straightforward task. Let C be the number of hard steric overlaps (instances of $r_{ij} < r_{ij}^*$) in a given configuration. It is simple to show that

$$\langle\cdot\rangle = \frac{\langle\cdot\,\Delta(C)\rangle^{(\text{tunnel})}}{\langle\Delta(C)\rangle^{(\text{tunnel})}}, \quad (16)$$

where the indicator function, $\Delta(x) = 1$ for $x = 0$ and $\Delta(x) = 0$ otherwise, effectively imposes steric constraints. Similarly, partition functions for the two ensembles are related by $Q = Q^{(\text{tunnel})}\langle\Delta(C)\rangle^{(\text{tunnel})}$.

Poor overlap between the canonical ensemble of interest and that of the noninteracting reference state requires that the ratio of partition functions $Q/Q^{(\text{ref})}$ in Eq. (4) be computed in stages. To avoid performing simulations with hard steric constraints, we first make use of the above result for the regularized partition function,

$$\frac{Q}{Q^{(\text{ref})}} = \frac{Q}{Q^{(\text{tunnel})}} \frac{Q^{(\text{tunnel})}}{Q^{(\text{ref})}} = \langle\Delta(C)\rangle^{(\text{tunnel})} \frac{Q^{(\text{tunnel})}}{Q^{(\text{ref})}} \quad (17)$$

The statistical consequences of adding the dihedral potential $E_{\text{dihedrals}}$ to our reference system can be evaluated with little computational effort, since no coupling among different rotatable bonds is involved. We can even calculate the corresponding ratio of partition functions analytically, $Q^{(\text{ref,dihedrals})}/Q^{(\text{ref})} = (2\pi\sigma^2)^{N^{(sp^3)}/2}$. We introduce additional factors to exploit this simplicity,

$$\frac{Q}{Q^{(\text{ref})}} = \langle\Delta(C)\rangle^{(\text{tunnel})} \frac{Q^{(\text{tunnel})}}{Q^{(\text{ref, dihedrals})}} \frac{Q^{(\text{ref, dihedrals})}}{Q^{(\text{ref})}}. \quad (18)$$

Finally, we introduce non-bonded and IS interactions in a gradual way through the potential

$$E^{(\text{switch})}(\lambda) = E_{\text{dihedrals}} + \lambda\left[E_{\text{non-bonded}}^{(\text{tunnel})} + E_{\text{implicit solvent}}\right] \quad (19)$$

By varying the switching parameter λ between 0 and 1, we interpolate between ensembles; in particular, $E^{(\text{switch})}(0) = E^{(\text{ref,dihedrals})}$ and $E^{(\text{switch})}(1) = E^{(\text{tunnel})}$. The non-interacting reference ensemble with dihedral bias can then be transformed into the fully interacting ensemble in a series of M steps,

$$\frac{Q^{(\text{tunnel})}}{Q^{(\text{ref,dihedrals})}} = \frac{Q^{(\text{switch})}(\lambda_0)}{Q^{(\text{switch})}(\lambda_1)} \frac{Q^{(\text{switch})}(\lambda_1)}{Q^{(\text{switch})}(\lambda_2)} \cdots \frac{Q^{(\text{switch})}(\lambda_{M-1})}{Q^{(\text{switch})}(\lambda_M)}, \quad (20)$$

where $\lambda_i = 1 - i/M$. Partition function ratios are evaluated according to

$$\frac{Q^{(\text{switch})}(\lambda_{i-1})}{Q^{(\text{switch})}(\lambda_i)} = \left\langle \exp\left[-\frac{\beta}{M}\left(E_{\text{non-bonded}}^{(\text{tunnel})} + E_{\text{implicit solvent}}\right)\right]\right\rangle_{\lambda_i}^{(\text{switch})}, \quad (21)$$

where $\langle\cdot\rangle_{\lambda_i}^{(\text{switch})}$ denotes an average in the ensemble corresponding to energy function $E^{(\text{switch})}(\lambda_i)$. By making M large, the difference between consecutive ensembles can be

made arbitrarily small, ensuring convergence of numerical averages in reasonable time.

Our MC simulations proceed by steps that attempt to reassign the value of a randomly selected side-chain dihedral angle χ_i. Trial values $\chi_i^{(\mathrm{trial})}$ are generated from a distribution $p^{(\mathrm{gen})}$ proportional to $\exp(-\beta E_{\mathrm{dihedrals}})$, accounting for local dihedral biases. Specifically, for sp^3–sp^3 hybridized bonds,

$$p^{(\mathrm{gen})}\left(\chi_i^{(\mathrm{trial})}\right) = \frac{1}{3}\left(2\pi\sigma^2\right)^{-\frac{1}{2}}\exp\left[-\phi_i^{(\mathrm{trial})2}/2\sigma^2\right]. \quad (22)$$

In this case, two-thirds of the attempted MC moves include hopping to a different rotameric state. For sp^3–sp^2 hybridized bonds, which lack intrinsic torsional bias in our model, trial values are selected from a distribution uniform in $\chi_i^{(\mathrm{trial})}$. These trial moves are accepted with a Metropolis probability[57] $p^{(\mathrm{acc})}$ based on the Boltzmann distribution determined by $E^{(\mathrm{switch})}(\lambda)$:

$$p^{(\mathrm{acc})} = \min\left[1, \exp\left(-\beta\lambda\left[\Delta E_{\mathrm{non-bonded}}^{(\mathrm{tunnel})} + \Delta E_{\mathrm{implicit\ solvent}}\right]\right)\right] \quad (23)$$

Here, $\Delta E_{\mathrm{non\text{-}bonded}}^{(\mathrm{tunnel})}$ and $\Delta E_{\mathrm{implicit\ solvent}}$ are changes in interaction energies resulting from the trial move. This acceptance probability does not involve changes in $E_{\mathrm{dihedrals}}$, whose statistics are fully addressed by the generation probability of Eq. (22).

Calculations were performed for many different values of λ (including $\lambda = 0$ and $\lambda = 1$, corresponding to the noninteracting reference system with dihedral bias and the regularized full potential, respectively). All simulations were repeated multiple times starting from randomly chosen initial side-chain configurations. Errors were estimated from variances among these trials or sets of trials.

Straightforward Metropolis MC sampling was sufficient to generate much of the data presented here. In the case of the CaM–ligand complexes, however, precise estimates could only be obtained with umbrella sampling techniques. For this purpose, we employed the adaptive method of Wang and Landau (WL).[39] Their original procedure was used to first construct a rough bias function, which was subsequently refined in several additional steps. During each refinement step, multiple independent simulations were performed using the same bias, and their resulting energy distributions were pooled to obtain a new estimate for the density of states.[76] This procedure was repeated until the density of states could be confidently constructed over a range of energies spanning those characteristic of physiological temperatures. Physical averages were finally computed in a nonadaptive run according to

$$\langle\cdot\rangle = \frac{\langle\cdot\exp[-\beta E(\Theta,\Phi)]\exp[W(E)]\rangle_{W(E)}}{\langle\exp[-\beta E(\Theta,\Phi)]\exp[W(E)]\rangle_{W(E)}}, \quad (24)$$

where $W(E)$ denotes the WL bias potential in units of k_BT and $\langle\cdot\rangle_{W(E)}$ indicates an average over the WL-biased ensemble.

Acknowledgements

This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. K.H.D. was supported by a National Science Foundation Graduate Research Fellowship and the Berkeley Fellowship.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.05.068

References

1. Chothia, C. (1975). Structural invariants in protein folding. *Nature*, **254**, 304–308.
2. Misura, K. M. S., Morozov, A. V. & Baker, D. (2004). Analysis of anisotropic side-chain packing in proteins and application to high-resolution structure prediction. *J. Mol. Biol.* **342**, 651–664.
3. Igumenova, T. I., Frederick, K. K. & Wand, A. J. (2006). Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem. Rev.* **106**, 1672–1699.
4. Wong, K. B. & Daggett, V. (1998). Barstar has a highly dynamic hydrophobic core: evidence from molecular dynamics simulations and nuclear magnetic resonance relaxation data. *Biochemistry*, **37**, 11182–11192.
5. Li, Z., Raychaudhuri, S. & Wand, A. J. (1996). Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein Sci.* **5**, 2647–2650.
6. Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4559–4570.
7. Best, R. B., Clarke, J. & Karplus, M. (2005). What contributions to protein side-chain dynamics are probed by NMR experiments? A molecular dynamics simulation analysis. *J. Mol. Biol.* **349**, 185–203.
8. Hu, H., Hermans, J. & Lee, A. L. (2005). Relating side-chain mobility in proteins to rotameric transitions: insights from molecular dynamics simulations and NMR. *J. Biomol. NMR*, **32**, 151–162.
9. Mittermaier, A. & Kay, L. E. (2001). Chi1 torsion angle dynamics in proteins from dipolar couplings. *J. Am. Chem. Soc.* **123**, 6892–6903.
10. Clarkson, M. W., Gilmore, S. A., Edgell, M. H. & Lee, A. L. (2006). Dynamic coupling and allosteric behavior in a nonallosteric protein. *Biochemistry*, **45**, 7693–7699.
11. Shapovalov, M. V. & Dunbrack, R. L. (2007). Statistical and conformational analysis of the electron density of protein side chains. *Proteins*, **66**, 279–303.
12. Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. (2007). Conformational entropy in molecular recognition by proteins. *Nature*, **448**, 325–329.
13. Kussell, E., Shimada, J. & Shakhnovich, E. I. (2001). Excluded volume in protein side-chain packing. *J. Mol. Biol.* **311**, 183–193.
14. Koehl, P. & Delarue, M. (1994). Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* **239**, 249–275.
15. Karplus, M. & Kushick, J. N. (1981). Method for estimating the configurational entropy of macromolecules. *Macromolecules*, **14**, 325–332.
16. Gohlke, H. & Case, D. A. (2004). Converging free energy estimates: MM-PB(GB)SA studies on the protein–protein complex Ras–Raf. *J. Comput. Chem.* **25**, 238–250.

17. Killian, B. J., Kravitz, J. Y. & Gilson, M. K. (2007). Extraction of configurational entropy from molecular simulations via an expansion approximation. *J. Chem. Phys.* **127**, 024107.
18. Chang, C.-E. A., McLaughlin, W. A., Baron, R., Wang, W. & McCammon, J. A. (2008). Entropic contributions and the influence of the hydrophobic environment in promiscuous protein–protein association. *Proc. Natl Acad. Sci. USA*, **105**, 7456–7461.
19. Gautier, R. & Tuffery, P. (2003). Critical assessment of side-chain conformational space sampling procedures designed for quantifying the effect of side-chain environment. *J. Comput. Chem.* **24**, 1950–1961.
20. Hu, X. & Kuhlman, B. (2006). Protein design simulations suggest that side-chain conformational entropy is not a strong determinant of amino acid environmental preferences. *Proteins*, **62**, 739–748.
21. Zhang, J. & Liu, J. S. (2006). On side-chain conformational entropy of proteins. *PLoS Comput. Biol.* **2**, 1586–1591.
22. Zídek, L., Novotny, M. V. & Stone, M. J. (1999). Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat. Struct. Biol.* **6**, 1118–1121.
23. Bernini, A., Ciutti, A., Spiga, O., Scarselli, M., Klein, S., Vannetti, S. *et al.* (2004). NMR and MD studies on the interaction between ligand peptides and alpha-bungarotoxin. *J. Mol. Biol.* **339**, 1169–1177.
24. Arumugam, S., Gao, G., Patton, B. L., Semenchenko, V., Brew, K. & Doren, S. R. V. (2003). Increased backbone mobility in beta-barrel enhances entropy gain driving binding of N-TIMP-1 to MMP-3. *J. Mol. Biol.* **327**, 719–734.
25. Mittermaier, A., Kay, L. E. & Forman-Kay, J. D. (1999). Analysis of deuterium relaxation-derived methyl axis order parameters and correlation with local structure. *J. Biomol. NMR*, **13**, 181–185.
26. Shetty, R. P., Bakker, P. I. W. D., DePristo, M. A. & Blundell, T. L. (2003). Advantages of fine-grained side chain conformer libraries. *Protein Eng.* **16**, 669–963.
27. Chandler, D. & Andersen, H. (1972). Optimized cluster expansions for classical fluids. II. Theory of molecular liquids. *J. Chem. Phys.* **57**, 1930–1931.
28. Schweizer, K. S. & Curro, J. G. (1987). Integral-equation theory of the structure of polymer melts. *Phys. Rev. Lett.* **58**, 246–249.
29. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93.
30. Weeks, J., Chandler, D. & Andersen, H. (1971). Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* **54**, 5237–5247.
31. Mayo, S. L., Olafson, B. D. & Goddard, W. A., III (1990). DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **94**, 8897–8909.
32. Tsai, J., Taylor, R., Chothia, C. & Gerstein, M. (1999). The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* **290**, 253–266.
33. Stickle, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. (1992). Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226**, 1143–1159.
34. Gordon, D. B., Marshall, S. A. & Mayo, S. L. (1999). Energy functions for protein design. *Curr. Opin. Struct. Biol.* **9**, 509–513.
35. Chandler, D. (2005). Interfaces and the driving force of hydrophobic assembly. *Nature*, **437**, 640–647.
36. Giovambattista, N., Lopez, C. F., Rossky, P. J. & Debenedetti, P. G. (2008). Hydrophobicity of protein surfaces: separating geometry from chemistry. *Proc. Natl Acad. Sci. USA*, **105**, 2274–2279.
37. Sharp, K. A., Nicholls, A., Fine, R. F. & Honig, B. (1991). Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science*, **252**, 106–109.
38. Shrake, A. & Rupley, J. A. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351–371.
39. Wang, F. & Landau, D. P. (2001). Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86**, 2050–2053.
40. Fersht, A. (1999). *Structure and Mechanism in Protein Science.* W. H. Freeman and Company, New York.
41. Hattori, M., Li, H., Yamada, H., Akasaka, K., Hengstenberg, W., Gronwald, W. & Kalbitzer, H. R. (2004). Infrequent cavity-forming fluctuations in HPr from *Staphylococcus carnosus* revealed by pressure- and temperature-dependent tyrosine ring flips. *Protein Sci.* **13**, 3104–3114.
42. Showalter, S. A., Johnson, E., Rance, M. & Brüschweiler, R. (2007). Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. *J. Am. Chem. Soc.* **129**, 14146–14147.
43. Doig, A. J. & Sternberg, M. J. (1995). Side-chain conformational entropy in protein folding. *Protein Sci.* **4**, 2247–2251.
44. Chang, C. A., Chen, W. & Gilson, M. K. (2007). Ligand configurational entropy and protein binding. *Proc. Natl Acad. Sci. USA*, **104**, 1534–1539.
45. Ratnaparkhi, G. S., Ramachandran, S., Udgaonkar, J. B. & Varadarajan, R. (1998). Discrepancies between the NMR and X-ray structures of uncomplexed barstar: analysis suggests that packing densities of protein structures determined by NMR are unreliable. *Biochemistry*, **37**, 6958–6966.
46. Babu, Y. S., Bugg, C. E. & Cook, W. J. (1988). Structure of calmodulin refined at 2.2 Å resolution. *J. Mol. Biol.* **204**, 191–204.
47. Teeter, M. M., Roe, S. M. & Heo, N. H. (1993). Atomic resolution (0.83 Å) crystal structure of the hydrophobic protein crambin at 130 K. *J. Mol. Biol.* **230**, 292–311.
48. Bode, W., Papamokos, E. & Musil, D. (1987). The high-resolution x-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech *Hirudo medicinalis*. Structural analysis, subtilisin structure and interface geometry. *Eur. J. Biochem.* **166**, 673–692.
49. Derrick, J. P. & Wigley, D. B. (1994). The third igg-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab. *J. Mol. Biol.* **243**, 906–918.
50. O'Neill, J. W., Kim, D. E., Baker, D. & Zhang, K. Y. (2001). Structures of the B1 domain of protein L from *Peptostreptococcus magnus* with a tyrosine to tryptophan substitution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **57**, 480–487.
51. Brudler, R., Meyer, T. E., Genick, U. K., Devanathan, S., Woo, T. T., Millar, D. P. *et al.* (2000). Coupling of hydrogen bonding to chromophore conformation and function in photoactive yellow protein. *Biochemistry*, **39**, 13478–13486.
52. Kang, B. S., Devedjiev, Y., Derewenda, U. & Derewenda, Z. S. (2004). The pdz2 domain of syntenin at ultra-high resolution: bridging the gap between macromolecular and small molecule crystallography. *J. Mol. Biol.* **338**, 483–493.

53. Poy, F., Yaffe, M. B., Sayos, J., Saxena, K., Morra, M., Sumegi, J. *et al.* (1999). Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. *Mol. Cell*, **4**, 555–561.
54. Schindelin, H., Jiang, W., Inouye, M. & Heinemann, U. (1994). Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **91**, 5119–5123.
55. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194**, 531–544.
56. Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. P. (1992). Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science*, **258**, 987–991.
57. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087.
58. Marlow, M. S. & Wand, A. J. (2006). Conformational dynamics of calmodulin in complex with the calmodulin-dependent kinase kinase alpha calmodulin-binding domain. *Biochemistry*, **45**, 8732–8741.
59. Frederick, K. K., Kranz, J. K. & Wand, A. J. (2006). Characterization of the backbone and side chain dynamics of the CaM–CaMKip complex reveals microscopic contributions to protein conformational entropy. *Biochemistry*, **45**, 9841–9848.
60. Osawa, M., Tokumitsu, H., Swindells, M. B., Kurihara, H., Orita, M., Shibanuma, T. *et al.* (1999). A novel target recognition revealed by calmodulin in complex with Ca^{2+}-calmodulin-dependent kinase kinase. *Nat. Struct. Biol.* **6**, 819–824.
61. Meador, W. E., Means, A. R. & Quiocho, F. A. (1992). Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin–peptide complex. *Science*, **257**, 1251–1255.
62. Clapperton, J. A., Martin, S. R., Smerdon, S. J., Gamblin, S. J. & Bayley, P. M. (2002). Structure of the complex of calmodulin with the target sequence of calmodulin-dependent protein kinase I: studies of the kinase activation mechanism. *Biochemistry*, **41**, 14669–14679.
63. Aoyagi, M., Arvai, A. S., Tainer, J. A. & Getzoff, E. D. (2003). Structural basis for endothelial nitric oxide synthase binding to calmodulin. *EMBO J.* **22**, 766–775.
64. Fallon, J. L. & Quiocho, F. A. (2003). A closed compact structure of native Ca(2+)-calmodulin. *Structure*, **11**, 1303–1307.
65. Zhou, Y., Vitkup, D. & Karplus, M. (1999). Native proteins are surface-molten solids: application of the Lindemann criterion for the solid versus liquid state. *J. Mol. Biol.* **285**, 1371–1375.
66. DeLano, W. (2007). MacPyMOL: a PyMOL-based molecular graphics application for MacOS X, DeLano Scientific LLC, Palo Alto, CA.
67. Mittermaier, A. & Kay, L. E. (2006). New tools provide new insights in NMR studies of protein dynamics. *Science*, **312**, 224–228.
68. Prabhu, N. V., Lee, A. L., Wand, A. J. & Sharp, K. A. (2003). Dynamics and entropy of a calmodulin–peptide complex studied by NMR and molecular dynamics. *Biochemistry*, **42**, 562–570.
69. Shehu, A., Kavraki, L. E. & Clementi, C. (2007). On the characterization of protein native state ensembles. *Biophys. J.* **92**, 1503–1511.
70. Grunberg, R., Nilges, M. & Leckner, J. (2006). Flexibility and conformational entropy in protein–protein binding. *Structure*, **14**, 683–693.
71. Deng, Y. & Roux, B. (2008). Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. *J. Chem. Phys.* **128**, 115103.
72. Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. (2002). Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **53**, 291–318.
73. Lee, A. L., Kinnear, S. A. & Wand, A. J. (2000). Redistribution and loss of side chain entropy upon formation of a calmodulin–peptide complex. *Nat. Struct. Biol.* **7**, 72–77.
74. Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. (2000). The penultimate rotamer library. *Proteins*, **40**, 389–408.
75. Fraczkiewicz, R. & Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and the gradients for macromolecules. *J. Comp. Chem.* **19**, 319–333.
76. Jayasri, D., Sastry, V. S. S. & Murthy, K. P. N. (2005). Wang–Landau Monte Carlo simulation of isotropic–nematic transition in liquid crystals. *Phys. Rev., E Stat. Nonlinear Soft Matter Phys.* **72**, 036702.