



# HHS Public Access

Author manuscript

*Annu Rev Biophys.* Author manuscript; available in PMC 2020 March 14.

Published in final edited form as:

*Annu Rev Biophys.* 2018 May 20; 47: 41–61. doi:10.1146/annurev-biophys-060414-034042.

## Measuring Entropy in Molecular Recognition by Proteins

A. Joshua Wand, Kim A. Sharp

Johnson Research Foundation and Department of Biochemistry and Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104-6059, USA

### Abstract

Molecular recognition by proteins is fundamental to the molecular basis of biology. Dissection of the thermodynamic landscape governing protein–ligand interactions has proven difficult because determination of various entropic contributions is quite challenging. Nuclear magnetic resonance relaxation measurements, theory, and simulations suggest that conformational entropy can be accessed through a dynamical proxy. Here, we review the relationship between measures of fast side-chain motion and the underlying conformational entropy. The dynamical proxy reveals that the contribution of conformational entropy can range from highly favorable to highly unfavorable and demonstrates the potential of this key thermodynamic variable to modulate protein–ligand interactions. The dynamical so-called entropy meter also refines the role of solvent entropy and directly determines the loss in rotational–translational entropy that occurs upon formation of high-affinity complexes. The ability to quantify the roles of entropy through an entropy meter based on measurable dynamical properties promises to highlight its role in protein function.

### Keywords

protein dynamics; conformational entropy; solvation entropy; rotational–translational entropy; molecular recognition; NMR relaxation

## INTRODUCTION

How the fundamental forces of chemistry and physics combine to define the energy landscape of proteins remains a difficult and complex question to unravel and dissect through experiment, simulation, or theory. This is particularly true in the context of the residual conformational (configurational) entropy of the protein and the entropy of solvent. For nearly half a century, the magnificent but largely monolithic atomic-scale views of protein molecules primarily obtained by crystallography have provided deep insight into the energetic (enthalpic) origins of their function. For decades, theoretical explorations, molecular simulations, and a variety of spectroscopic approaches have emphasized that protein molecules are nevertheless dynamic and explore a range of microstates. It is now clear that the internal conformational entropy of proteins is significantly larger than that

---

wand@pennmedicine.upenn.edu.

### DISCLOSURE STATEMENT

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

indicated by common representations of the native-state well of their energy landscapes. The availability of large reservoirs of entropy in the native states of proteins makes it critical to determine whether this entropy is coupled to the energetics of functions that exist beyond those involved in the folding and stability of protein molecules. Here, we focus on the development of a new approach to measuring changes in conformational entropy using what is referred to as a dynamical proxy provided by advanced nuclear magnetic resonance (NMR) relaxation methods. Over the past decade or so, this approach has emerged to allow quantitative access to the contribution of conformational entropy to protein molecular recognition, which is one of the most fundamental roles carried out by proteins in biology. In addition, experimental access to conformational entropy has also allowed refinement of the calculation of the contribution of solvent entropy based on changes in solvent-accessible surface area and the direct determination of the reduction in rotational–translational entropy upon formation of high-affinity complexes.

## THERMODYNAMICS AND DYNAMICS OF MOLECULAR RECOGNITION

Experimental investigation of the thermodynamics of molecular recognition by proteins often begins from a calorimetric perspective where the heat or enthalpy ( $H_{\text{total}}$ ) and free energy ( $G_{\text{total}}$ ) are measured and the total binding entropy ( $S_{\text{total}}$ ) is determined by

$$\Delta G_{\text{total}} = \Delta H_{\text{total}} - T\Delta S_{\text{total}} = \Delta H_{\text{total}} - T \left( \Delta S_{\text{conf}}^{\text{protein}} + \Delta S_{\text{conf}}^{\text{ligand}} + \Delta S_{\text{solvent}} + \Delta S_{\text{r-t}} + \Delta S_{\text{other}} \right). \quad 1.$$

Detailed atomic resolution structural models provide great insight into the origins of the enthalpy of binding. Much less certain are the various contributions to the total binding entropy. In principle, several types of entropy are potentially important (see right side of Equation 1). Historically, entropy has most often entered the discussion in terms of the changes in the entropy of solvent water ( $S_{\text{solvent}}$ ) and framed in terms of the so-called hydrophobic effect (3, 14, 32, 98).  $S_{\text{solvent}}$  has, with some success, been related empirically to changes in accessible surface area (ASA) of the protein and ligand upon complexation (32). Changes in the conformational entropy ( $S_{\text{conf}}$ ) and the rotational–translational entropy ( $S_{\text{r-t}}$ ) of the interacting species have received far less attention, presumably because they have resisted experimental measurement.  $S_{\text{other}}$  refers to other processes, such as proton binding or release, that may contribute to the total binding entropy (50). The goal is to measure these contributions to the binding of ligands by protein molecules. It turns out that an indirect approach utilizing a dynamical proxy obtained by NMR relaxation methods has proven quite successful.

## CHARACTERIZATION OF MOTION BY NMR RELAXATION METHODS

Solution NMR spectroscopy has emerged as a powerful means for the site-resolved measurement of internal motion in protein molecules over an impressive range of timescales in proteins of significant size and complexity (88). NMR spectroscopy has a bounty of capabilities to characterize time-dependent phenomena and can access time regimes from picoseconds to days (8, 90). As is apparent below, motion on the fast timescale is anticipated

to be rich in the underlying conformational entropy, and we focus on the features of solution NMR relaxation that sample this time regime. Here, we are ultimately most focused on those motions that *express* large contributions to protein conformational entropy (i.e., motions that extensively sample the many states available to the folded native state of the protein molecule). It has long been recognized from early simulations that the extremely fast soft-mode torsional oscillations of amino acid side chains contain significant entropy and that vibrational entropy will vary little as proteins move from state to state (41). Classical NMR relaxation phenomena will probe the interconversion of these states on the picosecond–nanosecond timescale.

### Mechanisms of NMR Relaxation

Very briefly, an ensemble of nuclear spins can relax from a nonequilibrium state via a range of potential interactions. For example, the internuclear dipole–dipole interaction between spatially proximal hydrogens gives rise to the nuclear Overhauser effect (NOE), which is likely familiar to the nonexpert reader as a means to measure interatomic distances for the determination of molecular structure in solution (70, 114). The strength of this interaction depends on the time average of the distance between the nuclei and of the angle of the internuclear vector with the applied magnetic field. In favorable situations, such as when the two nuclei are bonded, the distance dependence is effectively constant (though see 30, 82) and the interaction is temporally modified only by the change in the orientation of the bond vector with respect to the magnetic field. Examples include the  $^{15}\text{N}$ – $^1\text{H}$  amide bond (38) and the  $^{13}\text{C}$ – $^1\text{H}$  bond in a variety of contexts, such as in a methyl group (34). Other nuclear spin interactions are perhaps more obscure to the reader. These include chemical shift anisotropy, which reflects the interaction of the nuclear spin with the asymmetric distribution of electrons around it (91), and the interaction between the nuclear quadrupolar moment of spin quantum number of one or greater and the surrounding electric field gradient (37). In some cases, the interference between relaxation mechanisms can offer insight into motion (e.g., 17, 68, 84, 107, 111, 116).

Relaxation to an equilibrium distribution of nuclear spins in a liquid sample is mediated by the fluctuation of local fields arising from chemical shielding anisotropy, dipole–dipole, and quadrupolar interactions brought about by local (internal) motion and by macromolecular tumbling. Internal protein motion results in time modulation of the local fields that, in turn, contribute to relaxation. The theoretical connection between the observed relaxation phenomena and motion is complicated. The interested reader is referred to a recent monograph on this and related subjects (8). Very briefly, the density matrix formalism is employed to represent the properties of the statistical ensemble of spins in a liquid sample, while the surroundings (or lattice) are generally treated classically (1, 33). To obtain the correct behavior of the density matrix as it evolves in time, it is replaced by its nonequilibrium value through the action of a definable propagator. Once the time behavior of the density matrix is known, the behavior of an observable is obtained by taking a trace of the product between its matrix representation and the corresponding density matrix operators. Examples can be found elsewhere (34).

## Measures of NMR Relaxation

NMR observables that are measured can be expressed as linear combinations of the so-called spectral density functions. The spectral densities are related to the time dependence of the orientation of the vector defining the particular type of interaction. Consider a  $^{13}\text{C}$  nucleus with a single attached  $^1\text{H}$  (e.g., a methine alpha carbon of an amino acid). The motion of the  $^{13}\text{C}$ - $^1\text{H}$  interaction vector (along the rigid bond between them) can be described by the autocorrelation function of the dot product of the interaction vector's orientation at some time  $t$  and its orientation at some time  $t'$ . The time dependence will have two components: a contribution from the slower global tumbling of the protein and an assumed faster component due to motion within the molecular frame. The global tumbling of the protein in solution defines the upper limit of the timescale that can be investigated by these types of measurements. The spectral densities are linearly combined as required by the physics of the specific NMR relaxation mechanism to define an observable relaxation (see 34 for illustrative derivations). For example, the longitudinal relaxation rate ( $1/T_1$  or  $R_1$ ) of the  $^{13}\text{C}$  nucleus by a single bonded  $^1\text{H}$  nucleus is given by

$$R_1 = \frac{d^2}{4} [J(\omega_{\text{H}} - \omega_{\text{C}}) + 3J(\omega_{\text{C}}) + 6J(\omega_{\text{H}} + \omega_{\text{C}})], \quad 2.$$

where  $d^2$  is composed of fundamental constants and the effective C-H bond length and  $\omega_{\text{H}}$  and  $\omega_{\text{C}}$  are the resonance frequencies of  $^1\text{H}$  and  $^{13}\text{C}$  nuclear spins, respectively.

## Primary Interpretation of NMR Relaxation

The dominant procedure for describing the essential character of the motion that gives rise to NMR relaxation phenomena of the type considered here is the so-called model-free approach of Lipari & Szabo (66, 67). The Lipari-Szabo (L-S) form of the spectral density is remarkably robust (20). In the simplest case, the L-S treatment leads to three parameters: a correlation time for isotropic macromolecular reorientation ( $\tau_{\text{m}}$ ), an effective correlation time ( $\tau_{\text{e}}$ ), and a measure of the angular disorder of the interaction vector termed the squared generalized order parameter ( $O^2$ )<sup>1</sup> expressed as

$$J(\omega) = \frac{O^2 \tau_{\text{m}}}{1 + \omega^2 \tau_{\text{m}}^2} + \frac{(1 - O^2) \tau}{1 + \omega^2 \tau^2}, \quad 3.$$

where  $\tau^{-1} = \tau_{\text{m}}^{-1} + \tau_{\text{e}}^{-1}$ .

The order parameter by definition ranges from zero to one, corresponding to complete isotropic disorder and complete rigidity of the interaction vector within the molecular frame, respectively. It is this motional parameter that provides access to conformational entropy. The effective correlation time, however, is too model dependent to be used as a robust descriptor of the time constant(s) for the underlying motion (66).

<sup>1</sup>To avoid confusion with entropy ( $S$ ), we refer to the Lipari-Szabo squared generalized order parameter as  $O^2$  rather than its original designation as  $S^2$ .

For each site of interest, there are two unknowns ( $\mathcal{O}^2$  and  $\tau_c$ ) plus one global variable ( $\tau_m$ ) defining isotropic tumbling of the protein in solution. The situation can become more complicated, of course. For example, the character of the tumbling of the macromolecule may be anisotropic to some degree. This is directly handled using appropriate diffusion equations and data filtering (46, 55, 83, 103).

Some cases may require (justify) more complex forms of the model-free spectral density shown above (10, 72). Nevertheless, the experimental prescription is clear: Resolve relaxation data at  $n$  individual sites in a protein and measure as many relaxation parameters (e.g.,  $T_1$ ,  $T_2$ ) at as many magnetic fields (to vary  $\omega_H$ ,  $\omega_C$ ) as needed to provide a robust determination of the  $2n + 1$  (in the case of isotropic macromolecular tumbling) parameters.

The advent of multidimensional NMR spectroscopy provides a means to resolve literally hundreds of probe sites for motion in proteins. To realize a comprehensive use of NMR relaxation phenomena to characterize protein internal motion requires several key steps. Increasingly sophisticated isotopic labeling schemes have been introduced to simplify the complexity of the NMR relaxation as much as possible to make its measurement and subsequent analysis more robust (e.g. 21, 22, 27, 40, 45, 53, 56, 62–65, 75, 76, 81, 88, 92, 102, 109). The basic strategy is to reduce the number of relaxation mechanisms (interactions) as much as possible. Finally, the cornerstone was the development of NMR experiments that prepared so-called pure NMR observables of NMR relaxation that could be directly interpreted. Lewis Kay and colleagues are largely identified as the developers of the NMR machinery necessary to measure  $^{15}\text{N}$  (16, 44),  $^2\text{H}$  (77, 80), and  $^{13}\text{C}$  (81) autorelaxation in proteins. Tugarinov and coworkers (61) have extended the number of deuterium relaxation experiments that provide a context for extracting even more fundamental relaxation rate constants at a single magnetic field. Notable contributions for implementing  $^{13}\text{C}$  relaxation in the context of proteins came from Torchia and colleagues (36) in their unraveling of the complexity of this particular mechanism of relaxation in the context of proteins. More recently, cross-correlated relaxation in  $^{13}\text{CH}_3$  groups has emerged as a highly sensitive though perhaps not as accurate route to measures of protein motion in even very large protein systems (99, 100, 105). Finally, computational strategies were needed to confidently extract the desired model-free parameters (12, 72, 103).

### Preparation of Samples

Unfortunately, it turns out the relaxation properties of most  $^{13}\text{C}$  nuclei with bonded hydrogen(s) in proteins of significant size are unfavorable for detailed quantitative analysis. This has important ramifications as is discussed below. The important general exception is the methyl carbon, which, owing to the fast rotation of the methyl group, can be used as a relaxation probe in even very large proteins (88, 104). Isotopic labeling is designed to render relaxation in the methyl group as simple and tractable as possible. The  $\text{CHD}_2$  isotopomer is targeted for  $^{13}\text{C}$  relaxation and the  $\text{CH}_2\text{D}$  isotopomer for deuterium relaxation. In the case of deuterium methyl relaxation, which is most suited for smaller proteins, the  $\text{CH}_2\text{D}$  isotopomer is created by expression of the protein on minimal media in ~55%  $\text{D}_2\text{O}$ . For  $^{13}\text{C}$ -methyl relaxation studies, the labeling strategy has evolved significantly over the past two decades. A simple approach is to employ randomly partially deuterated  $^{13}\text{C}_3$ -pyruvate as a

general carbon source and the appropriate spin systems isolated by low-pass filtering (53). A better approach is to express the target protein during growth on unlabeled glucose and appropriately labeled metabolic precursors for valine, leucine, and isoleucine (27). Protein expression is carried out in “100%” D<sub>2</sub>O to ensure elimination of <sup>1</sup>H spins at nonmethyl sites in the protein (35).

Proteins produced in this manner are largely deuterated with selective <sup>1</sup>H and <sup>13</sup>C labeling in methyl groups. The methyl groups are mixtures of isotopomers (i.e., CD<sub>3</sub>, CHD<sub>2</sub>, CH<sub>2</sub>D, CH<sub>3</sub>). Only the three carrying at least one hydrogen are observed in the two-dimensional <sup>1</sup>H–<sup>13</sup>C chemical shift correlation spectrum and give cross-peaks that are at slightly different positions. The appropriate isotopomer, in this case the CHD<sub>2</sub>, is selected by spectroscopic manipulation (104). Recently, new precursors (21) and expression strategies (78) have become available for increasingly precise and efficient introduction of appropriate combinations of nuclear spins into proteins.

The NMR relaxation experiment is designed to follow the return of a particular type of magnetization from a nonequilibrium state back to equilibrium. For carbon relaxation in proteins, longitudinal and transverse relaxation processes are most useful and are likely familiar to the nonspecialist reader as  $T_1$  and  $T_2$  relaxation, respectively. The NOE is not so useful in the context of carbon relaxation in proteins of significant size since this observable reaches a limit as molecular tumbling slows. Relaxation at a specific methyl is quantified by the variation of the intensity of the corresponding <sup>1</sup>H–<sup>13</sup>C cross-peak with the relaxation time period. These types of data are then used to determine the L-S order parameter, which includes the effects of the fast rotation of the methyl group about its symmetry axis and the motion of the symmetry axis itself. Since the geometry of the methyl group can be assumed, the L-S order parameter for the symmetry axis,  $O_{\text{axis}}^2$ , is obtained directly.

## PRELUDE TO A DYNAMICAL PROXY FOR CONFORMATIONAL ENTROPY

As described above, various technical barriers to obtaining accurate and interpretable measures of site-resolved relaxation phenomena have been surmounted. Analytical strategies for analyzing these data are now largely established with the L-S model-free approach, providing the dominant framework. The next task was to try to make sense of the physical significance of measured squared generalized order parameters and, to a lesser extent, obtained effective correlation times.

### A Model-Independent Statistical View

Analysis of the first dozen or so comprehensive examples of methyl dynamics in proteins suggested that there are three basic classes or types of motion of methyl groups in proteins. Individual proteins often, though not always, displayed a distinct trimodal distribution centered roughly on  $O_{\text{axis}}^2$  values of ~0.35, ~0.6, and ~0.8 (34). The trimodal distribution is exemplified by the calmodulin complexes (19, 51, 54) (Figure 1a). It now seems clear that the three classes correspond to distinct types of methyl-bearing side-chain motion, which have only a modest relationship with the length of the side chain and are determined mostly by local context (93). The low order-parameter distribution centered at ~0.35 is identified as

having extensive rotamer interconversion that leads to averaging of  $J$ -coupling constants (52) and is termed the  $J$ -class (34). The distribution of order parameters centered at  $\sim 0.8$  is identified with highly restricted motion within a single rotameric well reminiscent of the rigidity of the protein backbone and is termed the  $\omega$ -class (34). Molecular dynamics (MD) simulations also suggest a slight narrowing of the rotamer potential at high  $O_{\text{axis}}^2$  values (93). The intermediate band centered on an  $O_{\text{axis}}^2$  value of  $\sim 0.6$  is suggested to involve larger amplitude motion within a rotamer well with occasional rotamer jumps (52, 93).

This interpretation of the trimodal distribution of methyl-bearing side-chain motion also brings with it the notion that the different classes of motion correspond to different quantities of conformational entropy. A calmodulin complex provided the first clear insight into the response of a protein to binding a high-affinity ligand. Calmodulin interacts with literally hundreds of proteins in the course of eukaryotic signaling (39). Most calmodulin-binding domains of protein regulated by calmodulin are satisfactorily represented by relatively short peptides that generally have an amphiphilic character (e.g., 47). Using deuterium relaxation, it was found that there is a significant redistribution and loss of methyl-bearing side-chain motion in calmodulin upon binding a peptide corresponding to the calmodulin-binding domain of the smooth-muscle myosin light chain kinase (51). From this perspective, Wand and coworkers (19) measured the variation in methyl-bearing side-chain dynamics across six calmodulin-peptide complexes. The startling result was the population of  $J$ ,  $\alpha$ , and  $\omega$  classes of methyl side-chain motion varied linearly with the total binding entropy (Figure 1b) (19). This provided a relatively model-independent view of the potential role for conformational entropy in protein-ligand interactions and led subsequently to a more quantitative interpretation.

### The Model-Dependent Oscillator Inventory Approach

One can imagine an approach that uses a specific potential energy function to obtain a quantitative connection between the motion interconverting states and the corresponding entropy. There is a long history of employing dynamical models to interpret experimental NMR relaxation data (e.g., 113 and references therein). The formal definition of the L-S squared generalized order parameter indicates how this might be done (66):

$$O^2 = \iint d\Omega_i d\Omega_j p_{\text{eq}}(\Omega_i) P_2(\cos\theta_{ij}) p_{\text{eq}}(\Omega_j), \quad 4.$$

where  $\Omega_i$  is the polar angles for the  $i$ th orientation of the interaction vector and  $\theta_{ij}$  is the angle between them. The key is  $p_{\text{eq}}(\Omega)$ , which describes the probability distribution of the orientations. Knowledge of  $p_{\text{eq}}(\Omega)$  provides the partition function of the system and hence gives access to fundamental thermodynamic quantities such as entropy. A particularly influential contribution came from Akke and coworkers (2), who illustrated how, using a specific potential, one could make a connection to the thermodynamics of the system. Taking this view, Yang & Kay (115) and Li and coworkers (60) employed free diffusion in an infinite square well and a harmonic oscillator potential, respectively, to make a parametric connection between entropy and the L-S order parameter. Importantly, it also turns out that

vastly different potentials give roughly the same change in entropy with a change in motion (order parameter) (52).

It is obvious that there are many difficulties and qualifications with using what we term the oscillator inventory approach. To name but a few, (a) the true nature of the potential governing motion is uncertain and may vary within and among proteins; (b) current NMR methods provide routine access to the dynamics of only a limited number of sites (e.g., the amide N–H bond vector and the methyl symmetry axis); (c) NMR relaxation has a distinct geometric dependence (see Equation 4) and so may not sense all internal motions to the same degree; and (d) motions of different NMR probes may be correlated. Notwithstanding these strong criticisms, application of the oscillator inventory analysis to the calmodulin complexes recapitulated the linear correlation between the apparent change in conformational entropy of calmodulin and the total binding entropy (Figure 1c). This once again strongly suggested that the entropy represented by the change in internal motion of the protein constituted a significant fraction of the total entropy of binding.

## CONSTRUCTION OF A DYNAMICAL PROXY FOR CONFORMATIONAL ENTROPY

To avoid the various limitations of the oscillator inventory approach, we embarked on a strategy that would empirically calibrate the relationship between changes in internal motion of methyl-bearing side chains and the corresponding change in total side-chain conformational entropy. Here, the view changes. We now seek to use the methyl group as a reporter not only of its own disorder but also as a response to the disorder of its surrounding.

### The Model-Independent Entropy Meter

The construction of the entropy meter rests on two fundamental assumptions: (a) that methyl-bearing side chains are numerous enough to provide good coverage throughout the protein and (b) that they are sufficiently coupled to neighboring side chains to report on the local disorder (entropy). This results in a simple relationship between what can be measured (protein motion) or calculated (solvent entropy) and what is desired (conformational entropy) (42, 74), expressed as

$$(\Delta S_{\text{tot}} - \Delta S_{\text{sol}}) = s_d \left[ \left( a \cdot \langle \Delta O_{\text{axis}}^2 \rangle^{\text{protein}} + b \cdot \langle \Delta O_{\text{axis}}^2 \rangle^{\text{ligand}} \right) \right] + \Delta S_{\text{RT}} + \Delta S_{\text{other}}, \quad 5.$$

where  $S_{\text{tot}}$ ,  $S_{\text{sol}}$ ,  $S_{\text{conf}}$ ,  $S_{\text{r-t}}$ , and  $S_{\text{other}}$  are the changes in total system entropy, solvent entropy, conformational entropy, rotational–translational entropy, and undocumented entropy, respectively. By postulate,  $S_{\text{conf}}$  is linearly related to the weighted change in the dynamics of the ligand and the protein upon binding. The weighting factors ( $a$  and  $b$ ) serve to project the measured changes in motion of the methyl-group reporters across the entire protein and ligand. In its original formulation (74), the weighting factors were simply the number of amino acid residues [the ligand was a peptide; nonpeptide ligands may require more detailed analysis (see 13, for example)], but this has subsequently been shown to be insufficient (42, 108), as is described below. A linear relationship between changes in the L-S order parameter and entropy appears to persist for order parameters between 0.1 and 0.9



(34). The scaling factor  $s_d$  is what we desire. In effect, the objections that might be made against the oscillator inventory approach are subsumed into this empirical parameter.

Initial application of this approach to the aforementioned calmodulin complexes proved most convincing (74). A high linear correlation for Equation 5 was observed. Subsequent use of this strategy in the analysis of various mutants of the catabolite activator protein binding DNA resulted in an even more impressive linearity and precision of the scaling factor  $s_d$  (106). However, comparison of the calmodulin and catabolite activator protein studies revealed that the two independently determined  $s_d$  values differed greatly (108). On one hand, perhaps they truly are different, but, on the other hand, one suspects that the relationship between motion and the underlying conformational entropy is a universal property of protein molecules.

### Insights from Computation and Simulation

To gain insight into this apparent discrepancy, we turned to MD simulations. MD simulations have improved greatly over the past decade in their ability to recapitulate experimental order  $O_{\text{axis}}^2$  values, and multi-nanosecond-length simulations are now sufficiently accurate to provide useful guidance (e.g., 5, 15, 25, 42, 48, 49, 57–59, 87). Utilizing a set of proteins with a range of dynamical properties, we examined some of the basic assumptions of the entropy meter approach outlined above (42, 95).

Perhaps the strongest premise of the entropy meter is the need for a quantitative linkage between the motional averaging of the methyl-group symmetry axis and the underlying conformational entropy. To examine this question, the rotamer population distributions of all the methyl-containing amino acids were calculated from the MD trajectories of each protein (42). An excellent linear correlation was seen between the rotamer entropy ( $S_b$ ) of each methyl-containing amino acid normalized by the number of side-chain torsion angles ( $N_\chi$ ) in that residue and the corresponding methyl side-chain order parameters calculated from the MD simulations (Figure 2). Local normalization by the number of torsion angles (soft modes) has been recommended by earlier simulations (57, 58). Importantly, it turns out that this linearity extends globally when the coupling between side chains is weak (95). Moreover, as might be expected, the distributions from the different proteins overlay one another, strongly suggesting that the empirical scaling of motion to the corresponding entropy should be universal.

A second critical assumption in the construction of the entropy meter is that the motion of the methyl group is sufficiently coupled to its surroundings to faithfully report on local disorder (74). This central issue was explored by comparing the rotamer entropy from the probability distributions for all side-chain torsion angles to the dynamical proxy (i.e., methyl-group motion). An excellent linear correlation was found indicating that methyl motions (disorder) do indeed report quantitatively on the disorder of surrounding amino acid side chains (Figure 3). The final factor is the extent to which motions of individual side chains are coupled and to what extent this coupling reduces the total internal conformational entropy. Again, MD simulations provide the insight that this coupling primarily manifests as a linear scaling, reducing the side-chain entropy by a constant factor of about 17% (42).

## Experimental Calibration of the Entropy Meter

We have recently experimentally tested the hypothesis that the entropy meter is universally applicable—in other words, that the scaling ( $s_d$ ) between changes in fast motion and changes in conformational entropy is constant (7). The 28 protein–ligand complexes used for the calibration span a broad range of binding affinities ( $K_d$ :  $10^{-4}$  to  $10^{-10}$  M) and ligand types, including nucleic acids, enzyme substrates and cofactors, carbohydrates, peptides, and proteins.

Usually, the contribution of solvent entropy to the thermodynamics of an equilibrium is calculated using empirically determined coefficients relating changes in accessible polar and apolar surface area (32). These coefficients have been derived with various assumptions about the nature of conformational entropy in the folded state. It is important to note that significant heat capacity changes accompany hydration of apolar and polar groups. This temperature variation is well described experimentally by the relations  $a_1 = dC_{p1} \ln(T/385 \text{ K})$  (3) and  $a_2 = dC_{p2} \ln(T/176 \text{ K})$  (71), where  $dC_{p1}$  and  $dC_{p2}$  are the hydration heat capacities per unit area of apolar and polar surface, respectively. The unprecedented extent of the dynamical data set used (7) allowed us to fit directly for these solvent entropy coefficients. In addition, the intercept of Equation 5 contains the loss in rotational–translational entropy upon formation of the complex  $G_{r-t}$ . We therefore recast Equation 5 as

$$\begin{aligned} \Delta S_{\text{total}} = s_d & \left[ \left( N_{\chi}^{\text{protein}} \Delta \langle O_{\text{axis}}^2 \rangle^{\text{protein}} \right) + \left( N_{\chi}^{\text{ligand}} \Delta \langle O_{\text{axis}}^2 \rangle^{\text{ligand}} \right) \right] \\ & + \left[ dC_{p1} \ln(T/385 \text{ K}) \Delta \text{ASA}_{\text{apolar}} + dC_{p2} \ln(T/176 \text{ K}) \Delta \text{ASA}_{\text{polar}} \right] \\ & + \left[ \Delta S_{r-t} + \Delta S_{\text{other}} \right] \end{aligned} \quad 6.$$

Equation 6 is written for the situation of having a ligand with soft torsional modes and would be modified for a rigid molecule. Violations of the few assumptions used to construct the calibration line for the entropy meter will tend to result in deviation from linearity. These include the assumptions that methyl groups are sufficiently numerous, are well distributed, and are adequately coupled to non-methyl-bearing amino acids such that their motions provide comprehensive coverage of internal motion in the protein. As discussed above, these assumptions are strongly supported by simulation (42).

Fitting of Equation 6 to the database of 28 protein complexes gives an  $R$  of 0.85 and  $p$  of  $\sim 10^{-8}$ . The fitted values of  $s_d$ ,  $S_{r-t}$ ,  $dC_{p1}$ , and  $dC_{p2}$  are listed in Table 1. The parameter of central interest ( $s_d$ ) is well determined to be  $-4.8 \pm 0.5 \text{ J mol}^{-1} \text{ K}^{-1}$  and provides for the first time a robust and apparently general means to experimentally obtain the change in conformational entropy upon protein–ligand association. It is important to appreciate that the effects of correlated motion noted above in the context of MD simulations are inherently taken care of by the experimental calibration of  $s_d$ . It is also notable that, in this treatment, we have ignored the contribution to the binding entropy from the backbone of the protein. Recent simulations suggest that the binding of ligands by *structured* proteins will involve little contribution from the polypeptide chain (95). Unfortunately, only six of the complexes used have sufficiently characterized dynamics to allow backbone motion be included in our

analysis. However, when analyzed in a similar fashion, this subset indicates that the contribution by the backbone to the binding entropy is less than 5% (7).

## ENTROPY IN MOLECULAR RECOGNITION BY PROTEINS

Utilizing the determined value of  $s_d$ , we can establish that the contribution of conformational entropy to molecular recognition by proteins is quite variable among complexes.

Conformational entropy can highly disfavor, have no effect on, or strongly favor association and is often a large determinant of the thermodynamics of binding (Figure 5). The structural origins of this behavior remain a mystery, but the sheer variation will undoubtedly motivate many future investigations.

Other entropic contributions to protein–ligand associations are also made accessible. The entropy meter represented by Equation 6 allows that there may be other unknown sources of entropy. These might include, for example, (de)protonation events associated with binding (50). Clearly, if  $S_{\text{other}}$  is both significant and varies between complexes or if  $S_{\text{r-t}}$  varies greatly between complexes, then the linearity of Equation 6 will be degraded. The observed linear correlation strongly suggests that neither of these occur. In this case, the ordinate intercept represents the loss in rotational–translational entropy ( $S_{\text{r-t}}$ ) upon formation of high-affinity complexes. The change in rotational–translational entropy accompanying formation of complexes involving proteins has been the subject of extensive theoretical debate (e.g., 6, 18, 23, 24, 29, 31, 69, 89, 97, 101, 117) but, like conformational entropy, has largely resisted experimental definition. From the fitted ordinate intercept in Figure 4, the apparent  $S_{\text{r-t}}$  is found to be  $\sim 0.10 \pm 0.01 \text{ kJ mol}^{-1} \text{ K}^{-1}$ , which is somewhat higher than that recently obtained through MD simulations (29). Because it is a freely fitted parameter, the rotational–translational entropy term serves as an internal control for the veracity of the entropy meter, and it is satisfying that the obtained value is largely in agreement with modern treatments. It is interesting to note that the dynamic response of a weak complex ( $K_d \approx 4 \text{ mmol}$ ) between serotonin and a mutant histamine-binding protein binary complex with histamine falls near the calibration line (Figure 4). This suggests that even for weak-associating ligands, the loss of translational–rotational entropy remains close to that obtained for the higher-affinity complexes used in the calibration (7).

Finally, the contribution of solvent entropy to processes involving proteins has usually been obtained from empirical relationships related to changes in solvent-accessible surface area. Underlying such approaches have been strong assumptions regarding conformational entropy in the folded state (32), which are not required by the entropy meter. Simultaneous fitting of the heat capacity coefficients in Equation 6 indicate that the burial of apolar and polar surfaces upon binding both produce a positive (favorable) change in solvent entropy. This is because hydration of polar groups, like that of apolar groups, has a negative entropy of hydration, in agreement with a wide range of thermodynamic data on solute, ion, and protein hydration (4, 71, 73). The surface area coefficients for apolar ( $a_1$ ) and polar ( $a_2$ ) desolvation entropy at 298 K are found to be  $-0.096 \pm 0.029 \text{ J mol}^{-1} \text{ K}^{-1} \text{ \AA}^{-2}$  and  $-0.027 \pm 0.005 \text{ J K}^{-1} \text{ \AA}^{-2}$ , respectively (Table 1). The corresponding hydration heat capacity surface area coefficients are also listed in Table 1. Burial of hydrophobic area stabilizes the complex through the usual hydrophobic effect. Concomitantly, burial of polar area also provides a

stabilizing entropic contribution to the complex via release of its hydrating water into the bulk, less ordered state. The coefficients are smaller than those obtained previously (26, 32) and, in part, reflect the impact of inadequately assessing the contribution of residual side-chain entropy in the folded state of proteins in prior work. Also interesting to note is that the smaller magnitude also likely arises from a difference in the nature of solvation. Previously, solvation entropies have been estimated from group transfers between water and organic solvents or global unfolding of globular proteins (e.g., 3, 26, 32, 71, 85, 86, 96), which involves complete solvation of side chains. In contrast, the studies here focus on the (de)solvation of more extended protein surfaces and may reflect differences in length scale, detailed geometry, and inherent dynamics of protein surfaces (e.g., 9, 94, 110). This remains to be further explored.

## LOOKING FORWARD

To see what the contribution of conformational entropy to ligand binding by proteins might be, we can turn to results from the proteins used to calibrate the entropy meter (Figure 5). This group of protein complexes shows that conformational entropy contributed by the response of amino acid side chains to the binding of a ligand can vary from highly unfavorable to negligible to highly favorable. In some cases, conformational entropy is essential for high-affinity binding. Indeed, it is important to note that, without the favorable contribution by conformational entropy, one quarter of these complexes would be reduced to biologically meaningless affinities. The structural origin of the variable utilization of conformational entropy in molecular recognition remains unknown and will undoubtedly be the focus of future investigation. The entropy meter will allow investigation of the role of entropy in context, including transition state and Michaelis complexes in enzymology, extremely high-affinity binding interactions such as those seen in immunology, and the underlying mechanisms of allostery guided by entropic forces (11, 28, 79, 112).

## ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health and the Mathers Foundation.

## TERMS AND DEFINITIONS

### **Binding entropy**

the change in total system entropy upon the formation of a protein–ligand complex

### **Configurational entropy**

the distribution of states accessed by both soft modes (torsions) and hard modes (vibrations) of motion

### **Conformational entropy**

entropy largely associated with distribution of rotameric (soft modes) states of amino acid side chains

### **Dynamical proxy**

the use of motion between states to implicitly count the states visited

**Entropy meter**

the empirically calibrated conversion of the dynamical proxy into a quantitative statement about conformational entropy

**Hard modes**

the motions (vibrations) involving atoms in fixed covalent geometry (e.g., bonds, germinal angles)

**Nuclear magnetic resonance relaxation**

the process of returning nonequilibrium distribution of spin states to equilibrium; intimately related to motion

**Rotational–translational entropy**

the entropy associated with the rotational and translational freedom that is lost when a protein and ligand combine to form a complex

**Soft modes**

the motions associated with torsion angles connecting different rotameric states

**Solvent entropy**

change in the entropy of water due to release from the protein and ligand surfaces to form a dry intermolecular interface

**LITERATURE CITED**

1. Abragam A 1961 Principles of Nuclear Magnetism. Oxford, UK: Clarendon Press
2. Akke M, Brüschweiler R, Palmer AG. 1993 NMR order parameters and free energy: an analytical approach and its application to cooperative Ca<sup>2+</sup> binding by calbindin-D9k. *J. Am. Chem. Soc.* 115:9832–33 Used specific potential energy functions to relate motion measurable by NMR relaxation and thermodynamic quantities.
3. Baldwin RL. 1986 Temperature dependence of the hydrophobic interaction in protein folding. *PNAS* 83:8069–72 [PubMed: 3464944]
4. Ben-Naim A, Marcus Y. 1984 Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* 81:2016–27
5. Bowman GR. 2015 Accurately modeling nanosecond protein dynamics requires at least microseconds of simulation. *J. Comp. Chem.* 37:558–66 [PubMed: 26077712]
6. Carlsson J, Aqvist J. 2005 Absolute and relative entropies from computer simulation with applications to ligand binding. *J. Phys. Chem. B* 109:6448–56 [PubMed: 16851719]
7. Caro JA, Kasinath V, Harpole KW, Lim J, Granja J, et al. 2016 Entropy in molecular recognition by proteins. *PNAS* 114:6563–68 Showed the general applicability of the entropy meter to protein–ligand associations.
8. Cavanagh J, Fairbrother WJ, Palmer AGI, Rance M, Skelton NJ. 2006 Protein NMR Spectroscopy: Principles and Practice. Burlington, MA: Elsevier
9. Chandler D 2005 Interfaces and the driving force of hydrophobic assembly. *Nature* 437:640–47 [PubMed: 16193038]
10. Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. 1990 Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112:4989–91
11. Cooper A, Dryden DTF. 1984 Allostery without conformational change: a plausible model. *Eur. Biophys. J.* 11:103–9 [PubMed: 6544679]

12. Dellwo MJ, Wand AJ. 1989 Model-independent and model-dependent analysis of the global and internal dynamics of cyclosporine A. *J. Am. Chem. Soc.* 111:4571–78
13. DeLorbe JE, Clements JH, Teresk MG, Benfield AP, Plake HR, et al. 2009 Thermodynamic and structural effects of conformational constraints in protein–ligand interactions. Entropic paradox associated with ligand preorganization. *J. Am. Chem. Soc.* 131:16758–70 [PubMed: 19886660]
14. Dill KA. 1990 Dominant forces in protein folding. *Biochemistry* 29:7133–55 [PubMed: 2207096]
15. DuBay KH, Bowman GR, Geissler PL. 2015 Fluctuations within folded proteins: implications for thermodynamic and allosteric regulation. *Acc. Chem. Res.* 48:1098–105 [PubMed: 25688669]
16. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, et al. 1994 Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation. *Biochemistry* 33:5984–6003 [PubMed: 7514039]
17. Ferrage F, Pelupessy P, Cowburn D, Bodenhausen G. 2006 Protein backbone dynamics through <sup>13</sup>C′–<sup>13</sup>Ca cross-relaxation in NMR spectroscopy. *J. Am. Chem. Soc.* 128:11072–78 [PubMed: 16925424]
18. Finkelstein AV, Janin J. 1989 The price of lost freedom: entropy of bimolecular complex formation. *Protein Eng.* 3:1–3 [PubMed: 2813338]
19. Frederick KK, Marlow MS, Valentine KG, Wand AJ. 2007 Conformational entropy in molecular recognition by proteins. *Nature* 448:325–29 [PubMed: 17637663] The first clear model-dependent indication that conformational entropy could contribute significantly to protein–ligand associations.
20. Frederick KK, Sharp KA, Warischalk N, Wand AJ. 2008 Re-evaluation of the model-free analysis of fast internal motion in proteins using NMR relaxation. *J. Phys. Chem. B* 112:2095–103 [PubMed: 18229915]
21. Gans P, Hamelin O, Sounier R, Ayala I, Durá MA, et al. 2010 Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew. Chem. Int. Ed.* 49:1958–62
22. Gardner KH, Kay LE. 1998 The use of <sup>2</sup>H, <sup>13</sup>C, <sup>15</sup>N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 27:357–406 [PubMed: 9646872]
23. Gilson MK, Given JA, Bush BL, McCammon JA. 1997 The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* 72:1047–69 [PubMed: 9138555]
24. Gilson MK, Zhou H-X. 2007 Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* 36:21–42 [PubMed: 17201676]
25. Glass DC, Krishnan M, Smith JC, Baudry J. 2013 Three entropic classes of side chain in a globular protein. *J. Phys. Chem. B* 117:3127–34 [PubMed: 23421556]
26. Gomez J, Hilser VJ, Xie D, Freire E. 1995 The heat capacity of proteins. *Proteins* 22:404–12 [PubMed: 7479713]
27. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. 1999 A robust and cost-effective method for the production of Val, Leu, Ile ( $\delta$ 1) methyl-protonated <sup>15</sup>N-, <sup>13</sup>C-, <sup>2</sup>H-labeled proteins. *J. Biomol. NMR* 13:369–74 [PubMed: 10383198]
28. Guo J, Zhou H-X. 2015 Dynamically driven protein allostery exhibits disparate responses for fast and slow motions. *Biophys. J.* 108:2771–74 [PubMed: 26083915]
29. Harpole KW, Sharp KA. 2011 Calculation of configurational entropy with a Boltzmann–quasi-harmonic model: the origin of high-affinity protein–ligand binding. *J. Phys. Chem. B* 115:9461–72 [PubMed: 21678965]
30. Henry ER, Szabo A. 1985 Influence of vibrational motion on solid-state line-shapes and NMR relaxation. *J. Chem. Phys.* 82:4753–61
31. Hermans J, Wang L. 1997 Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.* 119:2707–14
32. Hilser VJ, Garcia-Moreno EB, Oas TG, Kapp G, Whitten ST. 2006 A statistical thermodynamic model of the protein ensemble. *Chem. Rev.* 106:1545–58 [PubMed: 16683744]
33. Hoffman RA. 1970 Line shapes in high-resolution NMR. *Adv. Magn. Reson.* 4:88–198

34. Igumenova TI, Frederick KK, Wand AJ. 2006 Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem. Rev.* 106:1672–99 [PubMed: 16683749]
35. Ishima R, Louis JM, Torchia DA. 2001 Optimized labeling of <sup>13</sup>CHD<sub>2</sub> methyl isotopomers in perdeuterated proteins: potential advantages for <sup>13</sup>C relaxation studies of methyl dynamics of larger proteins. *J. Biomol. NMR* 21:167–71 [PubMed: 11727980]
36. Ishima R, Petkova AP, Louis JM, Torchia DA. 2001 Comparison of methyl rotation axis order parameters derived from model-free analyses of <sup>2</sup>H and <sup>13</sup>C longitudinal and transverse relaxation rates measured in the same protein sample. *J. Am. Chem. Soc.* 123:6164–71 [PubMed: 11414851]
37. Jacobsen JP, Bildsøe HK, Schaumburg K. 1976 Application of density matrix formalism in NMR spectroscopy. II. One-spin-1 case in anisotropic phase. *J. Magn. Reson.* 23:153–64
38. Jarymowycz VA, Stone MJ. 2006 Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* 106:1624–71 [PubMed: 16683748]
39. Kahl CR, Means AR. 2003 Regulation of cell cycle progression by calcium/calmodulin-dependent pathways. *Endocr. Rev.* 24:719–36 [PubMed: 14671000]
40. Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Guntert P. 2006 Optimal isotope labelling for NMR protein structure determinations. *Nature* 440:52–57 [PubMed: 16511487]
41. Karplus M, Ichiye T, Pettitt BM. 1987 Configurational entropy of native proteins. *Biophys. J.* 52:1083–85 [PubMed: 3427197]
42. Kasinath V, Sharp KA, Wand AJ. 2013 Microscopic Insights into the NMR relaxation-based protein conformational entropy meter. *J. Am. Chem. Soc.* 135:15092–100 [PubMed: 24007504] Simulation and theory was used to refine the original construction of the entropy meter.
43. Kasinath V, Valentine KG, Wand AJ. 2013 A <sup>13</sup>C labeling strategy reveals a range of aromatic side chain motion in calmodulin. *J. Am. Chem. Soc.* 135:9560–63 [PubMed: 23767407]
44. Kay LE, Torchia DA, Bax A. 1989 Backbone dynamics of proteins as studied by <sup>15</sup>N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28:8972–79 [PubMed: 2690953]
45. Kerfah R, Plevin MJ, Sounier R, Gans P, Boisbouvier J. 2015 Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* 32:113–22 [PubMed: 25881211] Summary of the isotopic labeling important for the NMR-based studies described here.
46. Kneller JM, Lu M, Bracken C. 2002 An effective method for the discrimination of motional anisotropy and chemical exchange. *J. Am. Chem. Soc.* 124:1852–53 [PubMed: 11866588]
47. Kranz JK, Lee EK, Nairn AC, Wand AJ. 2002 A direct test of the reductionist approach to structural studies of calmodulin activity: relevance of peptide models of target proteins. *J. Biol. Chem.* 277:16351–54 [PubMed: 11904288]
48. Krishnan M, Kurkal-Siebert V, Smith JC. 2008 Methyl group dynamics and the onset of anharmonicity in myoglobin. *J. Phys. Chem. B* 112:5522–33 [PubMed: 18399677]
49. Krishnan M, Smith JC. 2012 Reconstruction of protein side-chain conformational free energy surfaces from NMR-derived methyl axis order parameters. *J. Phys. Chem. B* 116:4124–33 [PubMed: 22401582]
50. Leavitt S, Freire E. 2001 Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* 11:560–66 [PubMed: 11785756]
51. Lee AL, Kinnear SA, Wand AJ. 2000 Redistribution and loss of side chain entropy upon formation of a calmodulin–peptide complex. *Nat. Struct. Biol.* 7:72–77 [PubMed: 10625431]
52. Lee AL, Sharp KA, Kranz JK, Song X-J, Wand AJ. 2002 Temperature dependence of the internal dynamics of a calmodulin–peptide complex. *Biochemistry* 41:13814–25 [PubMed: 12427045]
53. Lee AL, Urbauer JL, Wand AJ. 1997 Improved labeling strategy for <sup>13</sup>C relaxation measurements of methyl groups in proteins. *J. Biomol. NMR* 9:437–40 [PubMed: 9255947]
54. Lee AL, Wand AJ. 2001 Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature* 411:501–4 [PubMed: 11373686]

55. Lee LK, Rance M, Chazin WJ, Palmer AG. 1997 Rotational diffusion anisotropy of proteins from simultaneous analysis of  $^{15}\text{N}$  and  $^{13}\text{C}$  nuclear spin relaxation. *J. Biomol. NMR* 9:287–98 [PubMed: 9204557]
56. LeMaster DM, Kushlan DM. 1996 Dynamical mapping of *E. coli* thioredoxin via  $^{13}\text{C}$  NMR relaxation analysis. *J. Am. Chem. Soc.* 118:9255–64
57. Li DW, Brüschweiler R. 2009 A dictionary for protein side-chain entropies from NMR order parameters. *J. Am. Chem. Soc.* 131:7226–27 [PubMed: 19422234]
58. Li DW, Brüschweiler R. 2009 In silico relationship between configurational entropy and soft degrees of freedom in proteins and peptides. *Phys. Rev. Lett.* 102:118108
59. Li DW, Showalter SA, Brüschweiler R. 2010 Entropy localization in proteins. *J. Phys. Chem. B* 114:16036–44 [PubMed: 21077640]
60. Li Z, Raychaudhuri S, Wand AJ. 1996 Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein Sci.* 5:2647–50 [PubMed: 8976574]
61. Liao X, Long D, Li D-W, Brüschweiler R, Tugarinov V. 2012 Probing side-chain dynamics in proteins by the measurement of nine deuterium relaxation rates per methyl group. *J. Phys. Chem. B* 116:606–20 [PubMed: 22098066]
62. Lichtenecker RJ. 2014 Synthesis of aromatic  $^{13}\text{C}/^2\text{H}$ - $\alpha$ -ketoacid precursors to be used in selective phenylalanine and tyrosine protein labelling. *Organ. Biomol. Chem.* 12:7551–60
63. Lichtenecker RJ, Coudeville N, Konrat R, Schmid W. 2013 Selective isotope labelling of leucine residues by using  $\alpha$ -ketoacid precursor compounds. *ChemBioChem* 14:818–21 [PubMed: 23564734]
64. Lichtenecker RJ, Weinhaupl K, Reuther L, Schorghuber J, Schmid W, Konrat R. 2013 Independent valine and leucine isotope labeling in *Escherichia coli* protein overexpression systems. *J. Biomol. NMR* 57:205–9 [PubMed: 24078042]
65. Lichtenecker RJ, Weinhaupl K, Schmid W, Konrat R. 2013  $\alpha$ -Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. *J. Biomol. NMR* 57:327–31 [PubMed: 24264768]
66. 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.* 104:4546–59 The model-free formalism provides the fundamental basis for interpretation of NMR relaxation phenomena in terms of molecular motion.
67. Lipari G, Szabo A. 1982 Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* 104:4559–70
68. Liu W, Zheng Y, Cistola DP, Yang DW. 2003 Measurement of methyl  $^{13}\text{C}$ - $^1\text{H}$  cross-correlation in uniformly  $^{13}\text{C}$ -,  $^{15}\text{N}$ -, labeled proteins. *J. Biomol. NMR* 27:351–64 [PubMed: 14512732]
69. Luo H, Sharp K. 2002 On the calculation of absolute macromolecular binding free energies. *PNAS* 99:10399–404 [PubMed: 12149474]
70. Macura S, Ernst RR. 1980 Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Mol. Phys.* 41:95–117
71. Makhataдзе GI, Privalov PL. 1993 Contributions of hydration to protein-folding thermodynamics. 1. The enthalpy of hydration. *J. Mol. Biol.* 232:639–59 [PubMed: 8393940]
72. Mandel AM, Akke M, Palmer AG. 1995 Backbone dynamics of *Escherichia coli* ribonuclease H1: correlations with structure and function in an active enzyme. *J. Mol. Biol.* 246:144–63 [PubMed: 7531772]
73. Marcus Y 1994 A simple empirical model describing the thermodynamics of hydration of ions of widely varying charges, sizes, and shapes. *Biophys. Chem.* 51:111–27
74. Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. 2010 The role of conformational entropy in molecular recognition by calmodulin. *Nat. Chem. Biol.* 6:352–58 [PubMed: 20383153]
75. Mas G, Crublet E, Hamelin O, Gans P, Boisbouvier J. 2013 Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. *J. Biomol. NMR* 57:251–62 [PubMed: 24078041]



76. McIntosh LP, Dahlquist FW. 1990 Biosynthetic incorporation of  $^{15}\text{N}$  and  $^{13}\text{C}$  for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q. Rev. Biophys.* 23:1–38 [PubMed: 2188278]
77. Millet O, Muhandiram DR, Skrynnikov NR, Kay LE. 2002 Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in  $^{13}\text{C}$ -labeled and fractionally  $^2\text{H}$ -enriched proteins in solution. *J. Am. Chem. Soc.* 124:6439–48 [PubMed: 12033875]
78. Monneau YR, Ishida Y, Rossi P, Saio T, Tzeng S-R, et al. 2016 Exploiting *E. coli* auxotrophs for leucine, valine, and threonine specific methyl labeling of large proteins for NMR applications. *J. Biomol. NMR* 65:99–108 [PubMed: 27255761]
79. Motlagh HN, Wrabl JO, Li J, Hilser VJ. 2014 The ensemble nature of allostery. *Nature* 508:331–39 [PubMed: 24740064]
80. Muhandiram DR, Yamazaki T, Sykes BD, Kay LE. 1995 Measurement of  $^2\text{H}$   $T_1$  and  $T_{1\rho}$  relaxation-times in uniformly  $^{13}\text{C}$ -labeled and fractionally  $^2\text{H}$ -labeled proteins in solution. *J. Am. Chem. Soc.* 117:11536–44 Pioneering development of methods for deuterium relaxation in proteins.
81. Ollerenshaw JE, Tugarinov V, Skrynnikov NR, Kay LE. 2005 Comparison of  $^{13}\text{CH}_3$ ,  $^{13}\text{CH}_2\text{D}$ , and  $^{13}\text{CHD}_2$  methyl labeling strategies in proteins. *J. Biomol. NMR* 33:25–41 [PubMed: 16222555]
82. Ottiger M, Bax A. 1998 Determination of relative  $\text{N-HN}$ ,  $\text{N-C}'$ ,  $\text{C}\alpha\text{-C}'$ , and  $\text{C}\alpha\text{-H}\alpha$  effective bond lengths in a protein by NMR in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* 120:12334–41
83. Pawley NH, Wang C, Koide S, Nicholson LK. 2001 An improved method for distinguishing between anisotropic tumbling and chemical exchange in analysis of  $^{15}\text{N}$  relaxation parameters. *J. Biomol. NMR* 20:149–65 [PubMed: 11495246]
84. Pelupesy P, Ravindranathan S, Bodenhausen G. 2003 Correlated motions of successive amide N-H bonds in proteins. *J. Biomol. NMR* 25:265–80 [PubMed: 12766390]
85. Privalov PL, Makhatadze GI. 1992 Contribution of hydration and noncovalent interactions to the heat-capacity effect on protein unfolding. *J. Mol. Biol.* 224:715–23 [PubMed: 1314903]
86. Privalov PL, Makhatadze GI. 1993 Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J. Mol. Biol.* 232:660–79 [PubMed: 8393941]
87. Rajeshwar R, Smith JC, Krishnan M. 2014 Hidden regularity and universal classification of fast side chain motions in proteins. *J. Am. Chem. Soc.* 136:8590–605 [PubMed: 24844417]
88. Rosenzweig R, Kay LE. 2014 Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annu. Rev. Biochem.* 83:291–315 [PubMed: 24905784]
89. Ross PD, Subramanian S. 1981 Thermodynamics of protein association reactions: forces contributing to stability. *Biochemistry* 20:3096–102 [PubMed: 7248271]
90. Rule GS, Hitchens TK. 2006 *Fundamentals of Protein NMR Spectroscopy*. Dordrecht, Neth: Springer
91. Saito H, Ando I, Ramamoorthy A. 2010 Chemical shift tensor—the heart of NMR: insights into biological aspects of proteins. *Prog. NMR Spectrosc.* 57:181–228
92. Schrohober J, Sara T, Bisaccia M, Schmid W, Konrat R, Lichtenecker RJ. 2015 Novel approaches in selective tryptophan isotope labeling by using *Escherichia coli* overexpression media. *ChemBioChem* 16:746–51 [PubMed: 25703586]
93. Sharp KA, Kasinath V, Wand AJ. 2014 Banding of NMR-derived methyl order parameters: implications for protein dynamics. *Proteins* 82:2106–17 [PubMed: 24677353]
94. Sharp KA, Nicholls A, Fine RF, Honig B. 1991 Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* 252:106–9 [PubMed: 2011744]
95. Sharp KA, O'Brien E, Kasinath V, Wand AJ. 2015 On the relationship between NMR-derived amide order parameters and protein backbone entropy changes. *Proteins* 83:922–30 [PubMed: 25739366]
96. Spolar RS, Livingstone JR, Record MT Jr. 1992 Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry* 31:3947–55 [PubMed: 1567847]

97. Steinberg IZ, Scheraga HA. 1963 Entropy changes accompanying association reactions of proteins. *J. Biol. Chem.* 238:172–81 [PubMed: 13983721]
98. Sturtevant JM. 1977 Heat-capacity and entropy changes in processes involving proteins. *PNAS* 74:2236–40 [PubMed: 196283]
99. Sun H, Godoy-Ruiz R, Tugarinov V. 2012 Estimating side-chain order in methyl-protonated, perdeuterated proteins via multiple-quantum relaxation violated coherence transfer NMR spectroscopy. *J. Biomol. NMR* 52:233–43 [PubMed: 22314703]
100. Sun H, Kay LE, Tugarinov V. 2011 An optimized relaxation-based coherence transfer NMR experiment for the measurement of side-chain order in methyl-protonated, highly deuterated proteins. *J. Phys. Chem. B* 115:14878–84 [PubMed: 22040035]
101. Tamura A, Privalov PL. 1997 The entropy cost of protein association. *J. Mol. Biol.* 273:1048–60 [PubMed: 9367790]
102. Teilum K, Brath U, Lundstrom P, Akke M. 2006 Biosynthetic <sup>13</sup>C labeling of aromatic side chains in proteins for NMR relaxation measurements. *J. Am. Chem. Soc.* 128:2506–7 [PubMed: 16492013]
103. Tjandra N, Feller SE, Pastor RW, Bax A. 1995 Rotational diffusion anisotropy of human ubiquitin from <sup>15</sup>N NMR relaxation. *J. Am. Chem. Soc.* 117:12562–66
104. Tugarinov V, Kay LE. 2005 Quantitative <sup>13</sup>C and <sup>2</sup>H NMR relaxation studies of the 723-residue enzyme malate synthase G reveal a dynamic binding interface. *Biochemistry* 44:15970–77 [PubMed: 16331956]
105. Tugarinov V, Sprangers R, Kay LE. 2007 Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. *J. Am. Chem. Soc.* 129:1743–50 [PubMed: 17249677]
106. Tzeng S-R, Kalodimos CG. 2012 Protein activity regulation by conformational entropy. *Nature* 488:236–40 [PubMed: 22801505]
107. Vugmeyster L, Pelupessy P, Vugmeister BE, Abergel D, Bodenhausen G. 2004 Cross-correlated relaxation in NMR of macromolecules in the presence of fast and slow internal dynamics. *C. R. Phys.* 5:377–86
108. Wand AJ. 2013 The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr. Opin. Struct. Biol.* 23:75–81 [PubMed: 23246280]
109. Wand AJ, Bieber RJ, Urbauer JL, McEvoy RP, Gan ZH. 1995 Carbon relaxation in randomly fractionally <sup>13</sup>C-enriched proteins. *J. Magn. Reson. Ser. B* 108:173–75 [PubMed: 7648015]
110. Wang L, Berne BJ, Friesner RA. 2011 Ligand binding to protein-binding pockets with wet and dry regions. *PNAS* 108:1326–30 [PubMed: 21205906]
111. Wang T, Frederick KK, Igumenova TI, Wand AJ, Zuiderweg ERP. 2005 Changes in calmodulin main-chain dynamics upon ligand binding revealed by cross-correlated NMR relaxation measurements. *J. Am. Chem. Soc.* 127:828–29 [PubMed: 15656608]
112. Welch GR, Somogyi B, Damjanovich S. 1982 The role of protein fluctuations in enzyme action: a review. *Prog. Biophys. Mol. Biol.* 39:109–46 [PubMed: 7048419] Discussion of the potential role of entropy in protein function, which is only now directly testable by experiment.
113. Wittebort RJ, Szabo A. 1978 Theory of NMR relaxation in macromolecules: restricted diffusion and jump models for multiple internal rotations in amino acid side chains. *J. Chem. Phys.* 69:1722–36
114. Wüthrich K, Wider G, Wagner G, Braun W. 1982 Sequential resonance assignments as a basis for determination of spatial protein structures by high-resolution proton nuclear magnetic resonance. *J. Mol. Biol.* 155:311–19 [PubMed: 7077675]
115. Yang D, Kay LE. 1996 Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.* 263:369–82 [PubMed: 8913313]
116. Zhang X, Sui X, Yang D. 2006 Probing methyl dynamics from <sup>13</sup>C autocorrelated and cross-correlated relaxation. *J. Am. Chem. Soc.* 128:5073–81 [PubMed: 16608341]

117. Zhou H-X, Gilson MK. 2009 Theory of free energy and entropy in noncovalent binding. *Chem. Rev.* 109:4092–107 [PubMed: 19588959] Superb review of the theoretical issues of protein–ligand associations.

Author Manuscript

Author Manuscript

Author Manuscript

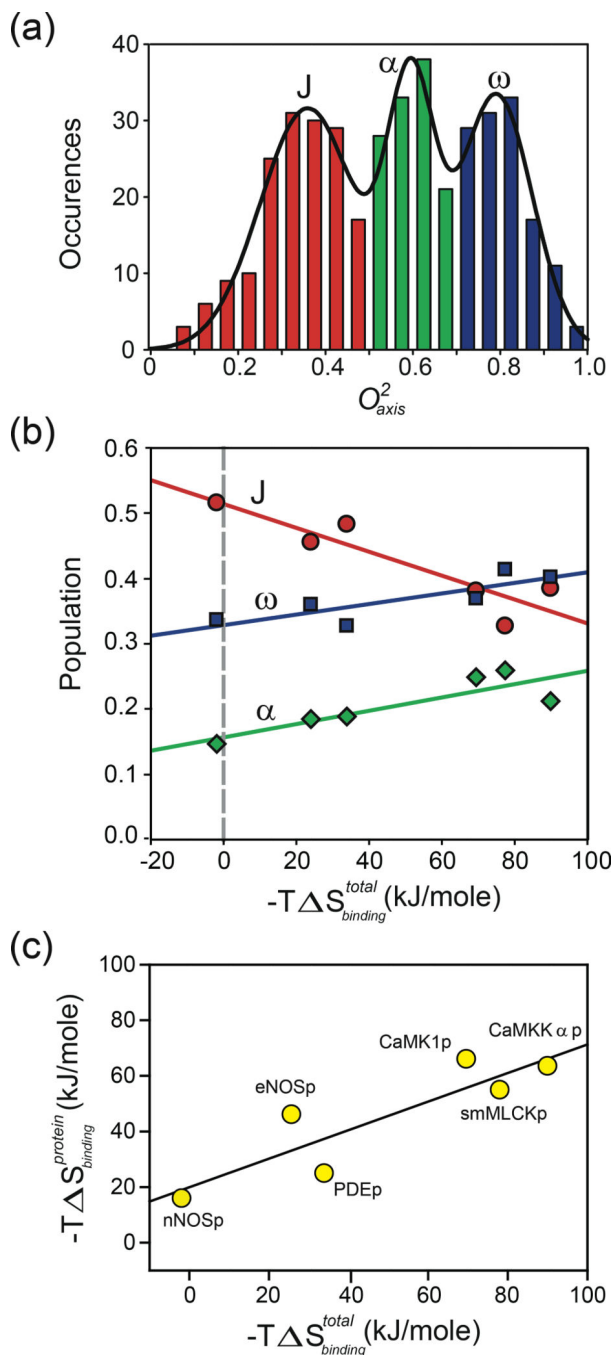
Author Manuscript

**SUMMARY POINTS**

1. Solution nuclear magnetic resonance (NMR) relaxation methods can provide extensive site-resolved measurement of internal dynamics of proteins, which can be interpreted in a robust, relatively model-free way using the Lipari-Szabo formulation.
2. The dynamical proxy for protein conformational entropy provided by NMR relaxation has been empirically calibrated. The resulting entropy meter allows the estimation of changes in conformational entropy upon a change in state (e.g., the binding of a ligand).
3. The entropy meter has revealed an unexpectedly broad range of responses by proteins to the binding of high-affinity ligands.

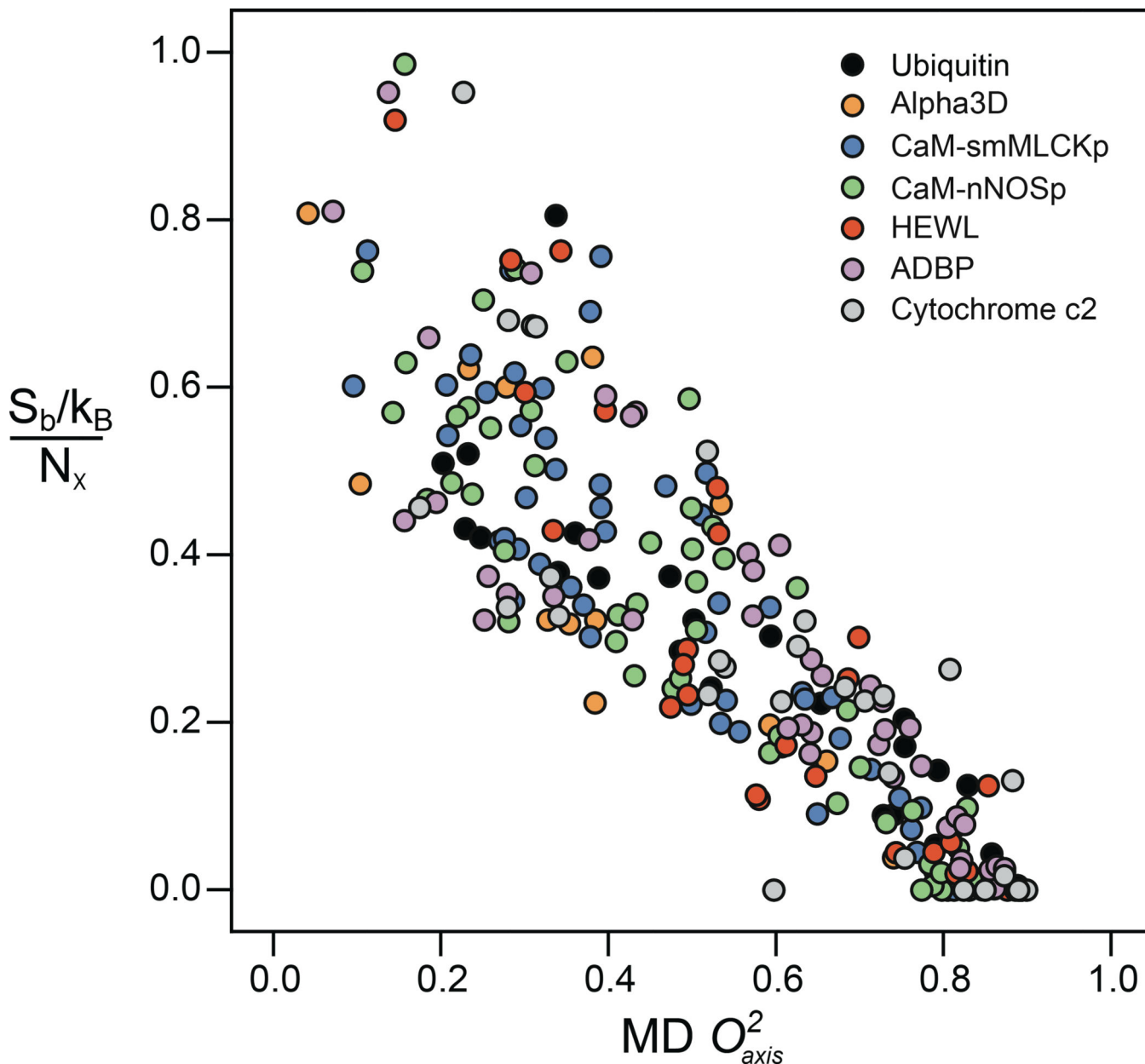
### FUTURE ISSUES

1. How can the entropy meter be expanded to incorporate slow timescales and additional probes, such as aromatic and polar side chains?
2. What strategies need to be developed for molecular dynamics simulations to reach quantitative agreement with nuclear magnetic resonance relaxation?
3. How does conformational entropy participate in very high- and low-affinity interactions?
4. What is the nature of protein surface hydration and its contribution to the thermodynamics of protein–ligand interactions?
5. Does conformational entropy participate in the activation entropy of enzyme catalysis or modulate the affinities of various Michaelis complexes?
6. How has conformational entropy contributed to the evolution of protein structure–function relationships?



**Figure 1.** Correlation of the apparent contribution of conformational entropy to the binding of target domains by calmodulin with total binding entropy. The apparent contribution reported by a model-dependent local interpretation of changes in methyl side-chain dynamics in calmodulin in complex with six different calmodulin-binding domains (19). (a) Summed histogram of the distribution of methyl group  $O_{axis}^2$  values of calmodulin in various complexes. (b) Correlation of the fractional population of the three classes of methyl-group motion with the total binding entropy measured by isothermal titration calorimetry for the

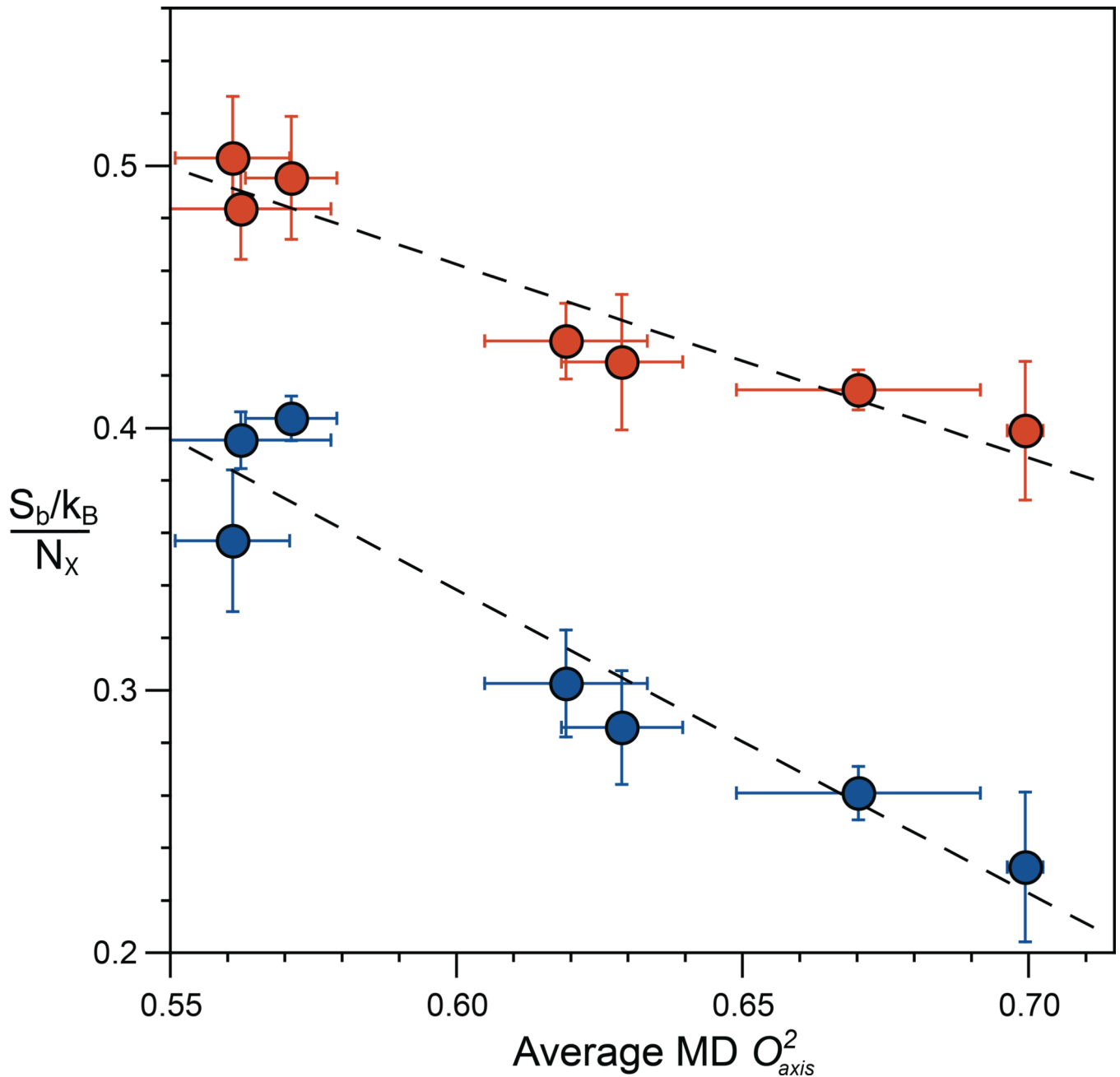
six calmodulin complexes: *J*-class (*red circles*), *α*-class (*green diamonds*), and *ω*-class (*blue squares*). (c) Application of the oscillator inventory approach to assess the role of conformational entropy of calmodulin in the binding of calmodulin-binding domains. The harmonic oscillator potential was used to convert changes in methyl-group motion to estimates of changes in local side-chain entropy. The simple sum of apparent entropies is plotted against the total binding entropy. A The p at the end of the terms in panel *c* signifies that the various calmodulin-binding domains are represented by peptides. daped from Reference 19. Abbreviations: CaMK1p, calmodulin-binding domain of calcium/calmodulin-dependent protein kinase1; CaMKKαp, calmodulin-binding domain of calcium/calmodulin-dependent protein kinase kinase alpha; eNOSp, calmodulin-binding domain of endothelial nitric oxide synthase; nNOSp, calmodulin-binding domain of neuronal nitric oxide synthase; PDEp, calmodulin-binding domain of phosphodiesterase; smMLCKp; calmodulin-binding domain of smooth muscle myosin light chain kinase.



**Figure 2.**

Correlation of the methyl rotamer entropy versus  $O_{axis}^2$  from molecular dynamics (MD) simulations. Normalized entropy  $S_b/N_\chi$  given for every side-chain methyl probe. The correlation was highly linear ( $R^2$  of 0.77) with a slope of  $-0.88 \pm 0.03$  and an intercept of  $0.78 \pm 0.02$ . The different proteins are color coded as indicated in the inset key. Modified from Reference 42, copyright American Chemical Society. Abbreviations: ADBP, porcine procarboxypeptidase B; CaM, calmodulin;  $k_B$ , Boltzmann constant; HEWL, hen egg white lysozyme; nNOSp, calmodulin-binding domain of neuronal nitric oxide synthase;  $N_\chi$ , total number of side-chain torsion angles;  $S_b$ , rotamer entropy; smMLCKp, calmodulin binding domain of the smooth muscle myosin light chain kinase.





**Figure 3.**

The dynamic proxy of methyl groups is an excellent reporter of both methyl and total side-chain rotameric entropy. (*Blue circles*) The normalized methyl rotameric entropy for each protein is calculated as the summation of  $S_b$  for individual methyl-bearing amino acids divided by the number of associated rotamer angles ( $N_\chi$ ). (*Red circles*) The total rotameric entropy for each protein is calculated as the summation of  $S_b$  for all residues and is normalized by the respective total number of rotamer angles ( $N_\chi$ ). The average methyl  $O^2_{axis}$  parameter for all methyl-bearing residues, including Ala, is that obtained from MD simulations. A very high linear correlation is observed for both methyl side-chain rotamer

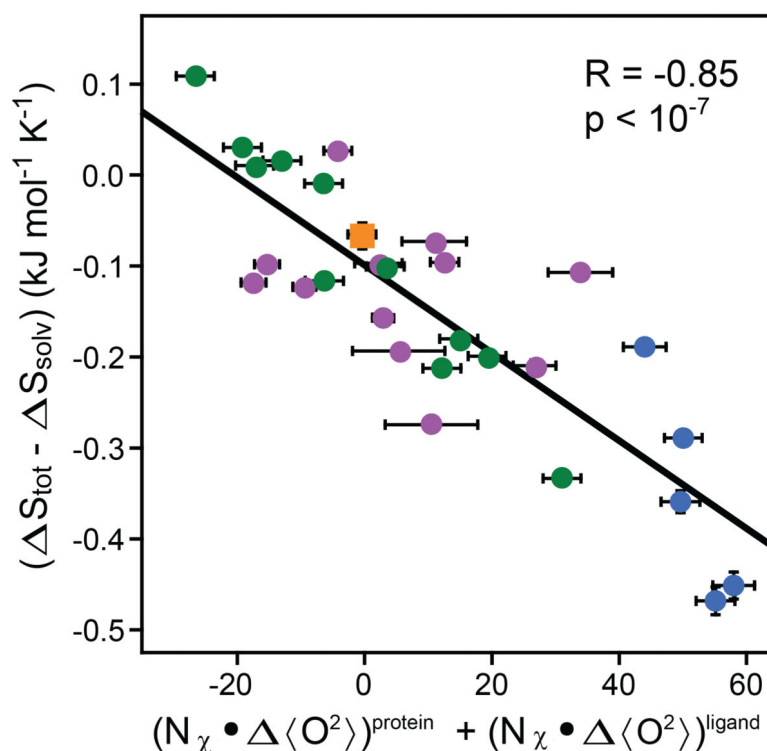
entropy (slope =  $-1.16 \pm 0.17$ ,  $R^2 = 0.90$ ) and total rotamer entropy (slope =  $-0.74 \pm 0.10$ ,  $R^2 = 0.91$ ), indicating that methyl-group motion is a good predictor of total side-chain conformational entropy. Adapted from Reference 43.

Author Manuscript

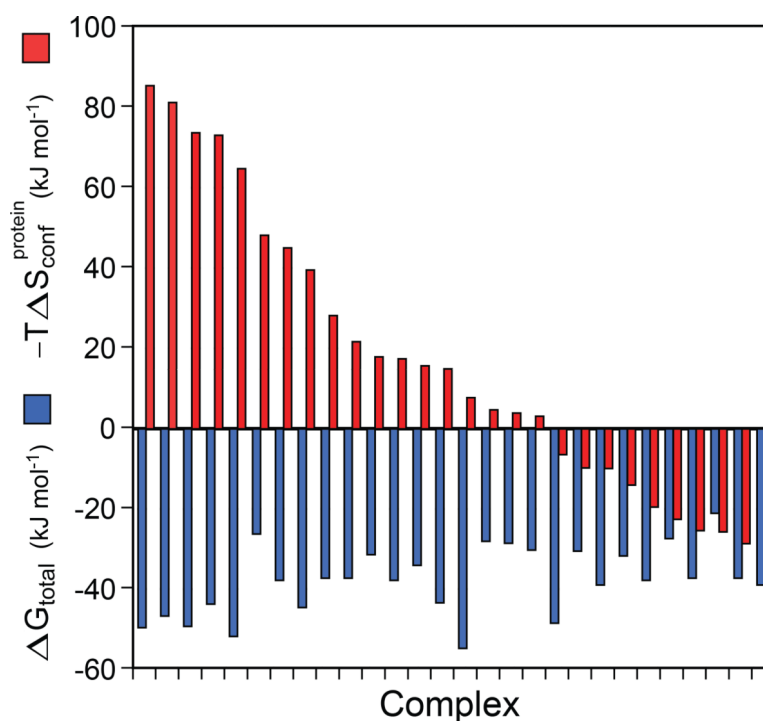
Author Manuscript

Author Manuscript

Author Manuscript



**Figure 4.** Calibration of the dynamical proxy for protein conformational entropy. Fitting of Equation 6 to data provided by 28 protein–ligand associations. The difference in the measured total binding entropy and calculated solvent entropy is plotted against the change in the dynamical proxy upon binding of ligands. The dynamical proxy is the average Lipari-Szabo squared generalized order parameter of methyl-group symmetry axes ( $\langle O_{\text{axis}}^2 \rangle$ ). The fitted slope ( $s_d$ ) of  $-4.8 \pm 0.5 \text{ J mol}^{-1} \text{ K}^{-1}$  allows for the conversion between measured changes in methyl-bearing side-chain motion and the associated conformational entropy. Other parameters of the entropy meter are summarized in Table 1. The CaM (calmodulin) and CAP (calmodulin-associated peptide) data subsets are shown in blue and green, respectively. Purple circles represent other binary complexes, as summarized elsewhere (7). The orange square represents the HBP(D24R)–histamine binary complex binding to serotonin, which has a dissociation constant in the millimolar range and was not used in the calibration. See text for details. Adapted from Reference 7.



**Figure 5.**

Contribution of protein conformational entropy to the free energy of ligand binding to proteins. The broad range of contributions available to proteins for high-affinity binding of ligands is illustrated by the protein–ligand complexes used to calibrate the entropy meter (Figure 4). The 28 protein–ligand complexes are arranged in descending order of the contribution of conformational entropy (*red bars*) to the total free energy of binding (*blue bars*). Conformational entropy contributed by the response of amino acid side chains to the binding of a ligand can vary from highly unfavorable to negligible to highly favorable. In some cases, conformational entropy is essential for high-affinity binding. The structural origins of the variable utilization of conformational entropy in molecular recognition are unknown. In most cases, the change in solvent entropy remains a dominant contribution. Note that  $-T\Delta S_{\text{R-t}}$ ,  $\Delta S_{\text{ligand}}$ , and  $\Delta S_{\text{solvent}}$  are not shown here. Adapted from Reference 7

**Table 1**Calibration of the dynamical proxy for conformational entropy<sup>a</sup>

Parameter	Value
Ordinate intercept ( $S_{r-t} + S_{other}$ )	$-(0.10 \pm 0.01) \times 10^{-1} \text{ kJ mol}^{-1} \text{ K}^{-1}$
Slope ( $s_d$ )	$-(4.8 \pm 0.5) \times 10^{-3} \text{ kJ mol}^{-1} \text{ K}^{-1}$
Hydration heat capacity—apolar ( $dC_{p1}$ )	$+(37 \pm 11) \times 10^{-5} \text{ kJ mol}^{-1} \text{ K}^{-1} \text{ \AA}^{-2}$
Hydration heat capacity—polar ( $dC_{p2}$ )	$-(5.2 \pm 1.0) \times 10^{-5} \text{ kJ mol}^{-1} \text{ K}^{-1} \text{ \AA}^{-2}$
Solvent entropy coefficient—apolar ( $a_1$ ) at 298 K	$-(9.6 \pm 2.9) \times 10^{-5} \text{ kJ mol}^{-1} \text{ K}^{-1} \text{ \AA}^{-2}$
Solvent entropy coefficient—polar ( $a_2$ ) at 298 K	$-(2.7 \pm 0.5) \times 10^{-5} \text{ kJ mol}^{-1} \text{ K}^{-1} \text{ \AA}^{-2}$

<sup>a</sup>Derived from a global fit of nuclear magnetic resonance–derived dynamical data. Precision determined by Monte Carlo sampling. See Reference 7 for details.  $a_1 = dC_{p1} \ln(T/385 \text{ K})$  (3) and  $a_2 = dC_{p2} \ln(T/176 \text{ K})$  (71), where  $dC_{p1}$  and  $dC_{p2}$  are the hydration heat capacities per unit area of apolar and polar surface, respectively. Substituting these relations for  $a_1$  and  $a_2$  into Equation 6, the parameters  $dC_{p1}$ ,  $dC_{p2}$ ,  $S_{r-t}$ , and  $s_d$  are obtained from a global fit using the experimental binding entropy changes, order parameter changes, and temperatures.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript