Published in final edited form as:

Proteins. 2014 September ; 82(9): 2106–2117. doi:10.1002/prot.24566.

Banding of NMR-derived Methyl Order Parameters: Implications for Protein Dynamics

Kim A. Sharp*, Vignesh Kasinath, and A. Joshua Wand*

Graduate Group in Biochemistry and Molecular Biophysics, Johnson Research Foundation and Department of Biochemistry & Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia 19104, United States of America

Abstract

Our understanding of protein folding, stability and function has begun to more explicitly incorporate dynamical aspects. Nuclear magnetic resonance has emerged as a powerful experimental method for obtaining comprehensive site-resolved insight into protein motion. It has been observed that methyl-group motion tends to cluster into three “classes” when expressed in terms of the popular Lipari-Szabo model-free squared generalized order parameter. Here the origins of the three classes or bands in the distribution of order parameters are examined. As a first step, a Bayesian based approach, which makes no *a priori* assumption about the existence or number of bands, is developed to detect the banding of O^2_{axis} values derived either from NMR experiments or molecular dynamics simulations. The analysis is applied to seven proteins with extensive molecular dynamics simulations of these proteins in explicit water to examine the relationship between O^2 and fine details of the motion of methyl bearing side chains. All of the proteins studied display banding, with some subtle differences. We propose a very simple yet plausible physical mechanism for banding. Finally, our Bayesian method is used to analyze the measured distributions of methyl group motions in the catabolite activating protein and several of its mutants in various liganded states and discuss the functional implications of the observed banding to protein dynamics and function.

Keywords

protein motion; NMR relaxation; amino acid side chain motion; conformational entropy; Bayesian analysis; ubiquitin; calmodulin; catabolite activation protein; molecular recognition

Introduction

The shift in perspective on proteins from static structure to a dynamic molecule has proven to be crucial in our understanding of protein folding, stability and function.^{1,2}

Multidimensional nuclear magnetic resonance (NMR) has emerged as the pre-eminent

*Correspondence to: A. J. Wand (wand@mail.med.upenn.edu) or K. A. Sharp (sharpk@mail.med.upenn.edu), Department of Biochemistry & Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104-6059). Contact information: Professor A. J. Wand, Department of Biochemistry & Biophysics, University of Pennsylvania, 905 Stellar-Chance Laboratories, 422 Curie Blvd, Philadelphia, Pennsylvania 19104-6059, telephone: 215-573-7288, facsimile: 215-573-7288, wand@mail.med.upenn.edu

experimental means of obtaining comprehensive site-resolved insight into motion on almost any timescale in solution.^{3–7} Over the past two decades considerable progress has been made in the use of deuterium and carbon relaxation in methyl groups as probes of protein side chain motion. Of particular interest is the potential for using internal protein motion as an indirect measure of or proxy for conformational entropy.^{8,9} In parallel, an interpretative framework has been constructed to allow the empirical determination of changes in conformational entropy from changes in methyl group motion induced, for example, by a change in functional state (e.g. binding of a ligand).^{10–12} This approach requires few assumptions about the nature of the underlying motion and appears to allow one to obtain robust estimates of the conformational entropy involved in protein function.¹² The assumptions involved in the construction of the empirical calibration, notably that the motion of methyl-bearing amino acid side chains are sufficiently coupled to their surroundings, were recently validated using molecular dynamics (MD) simulations of several proteins.¹² The excellent agreement between the average MD calculated order parameters of methyl side-chains versus those from NMR measurements expands the scope of this empirical calibration. NMR measured order parameters for any protein can be converted into a protein conformational entropy measure with remarkable accuracy. While this empirical conversion is a large step forward in accurately quantifying protein conformational entropy, a mechanistic understanding of the observed distribution of order parameter values in proteins is important for a full understanding of the conformational entropy measure of proteins.

Early deuterium methyl relaxation studies have indicated that the distribution of protein side chain flexibility is heterogeneous.⁴ While the chemical nature of the side chain (i.e. number of torsion angles) contributes somewhat,¹³ there is considerable variation in the dynamics of any individual residue type depending on its location and in which protein it resides. The overall dynamical character of a given side chain often does not follow intuitively anticipated dependencies (e.g. number of torsion angles, depth of burial, packing efficiency, etc.) or simple structural correlates such as secondary structure type.⁴ It has also been suggested that methyl-bearing side chain motion was not randomly distributed but tended to cluster into three “classes” of motion when expressed in terms of the popular Lipari-Szabo model-free squared generalized order parameter for the methyl group symmetry axis (O^2_{axis}).⁴ A characteristic tri-modal distribution was broadly observed for several different proteins.⁴ This distribution is particularly prevalent in the calmodulin-peptide complexes.^{10,14} The temperature dependence of the three classes of provided a simple though somewhat controversial explanation for the so-called glass transition in proteins.¹⁴ More interestingly, the observation that the three classes of motion responded differently to perturbations brought about by the binding of ligands by calmodulin suggested that these motional heterogeneities influence the thermodynamics of molecular recognition.¹⁰

The generality of the three-class interpretation of methyl group motion was investigated using molecular dynamics simulations where it was postulated that the tri-modal distribution seen for proteins could be explained by the fact that individual methyl containing amino acids from several different proteins themselves showed this distribution.^{15,16} This unfortunately still does not explain the origin of the trimodal distribution. Also, these early

molecular dynamics studies showed relatively poor correspondence between the generalized order parameters calculated from the trajectories and those derived from experiment, limiting the inferences that one can draw from the simulations. Since that time, MD simulation methodology, force fields and computational power available have all improved with the result that the accuracy of order parameters calculated from molecular dynamics simulations has increased markedly.^{12,17–19}

The origin of the tendency of O^2_{axis} values to cluster into a small number of groups or classes (hereafter termed ‘banding’) remains a mystery. This phenomenon is particularly obvious in ubiquitin (Figure 1). Shifts in the populations of these classes of methyl group motion have been used to gain insight into the effects of high pressure on this protein’s structure and dynamics.²⁰ Here we attempt to understand the origins of the banding in the distribution of order parameters in proteins in terms of the underlying features of the protein dynamics. As a first step, we apply a Bayesian based approach, which makes no *a priori* assumptions about the presence and number of bands to detect the banding of O^2_{axis} values derived either from NMR experiments or MD simulations. Since all number of bands, from one (no banding) upwards are given equal a priori probability, the Bayesian method is unbiased with regard to banding. The characteristic distribution of methyl order parameters agrees very well with our earlier non-Bayesian analysis of banding.^{Frederick, 2007 #12} In order to understand the fundamental origin of banding we combine statistical analysis of NMR derived O^2_{axis} values from several proteins with extensive molecular dynamics simulations of these proteins in explicit water to examine the relationship between O^2 and fine details of the motion of methyl bearing side chains. This analysis reveals that all the proteins studied display banding, with some subtle differences. We propose a very simple yet plausible physical mechanism for banding. Finally, our Bayesian method is used to analyze the measured distribution of O^2_{axis} in catabolite activating protein (CAP) and mutants of CAP bound to the same DNA ligand²¹ and discuss the functional implications of the observed banding to protein dynamics and function.

Materials and Methods

Bayesian analysis of O^2_{axis} distributions

Given a set of Lipari-Szabo methyl axis order parameters for a protein,¹⁰ it is assumed that these result from side chain motions belonging to an unknown number of motional classes, M , where $M=1$ corresponds to a homogeneous population i.e. no banding. The prior probability of M , $p(M|I)$, is uniform, i.e. there is no preference for any number of bands. It indicates recognition of the usual Bayesian background assumptions [Jeffreys, 1957 #44]. The fraction of methyl-bearing side chains belonging to each motional class is A_j , where

$\sum_j^M A_j = 1$. It is also assumed that within each motional class the actual values of the methyl side chain order parameter y_k are distributed around some characteristic mean value x_j , with some variance, ϵ^2 . The spread reflects both experimental uncertainties and heterogeneity in motional details from side chain to side chain within a class. With this model, the probability of obtaining a particular datum y_k is

$$p(y_k|M, \{A_j, x_j\}, \varepsilon, I) = \sum_{j=1}^M \frac{A_j}{\varepsilon \sqrt{2\pi}} \exp\left(-\frac{(y_k - x_j)^2}{2\varepsilon^2}\right) \quad (1)$$

where each term on the right hand side is the probability that side chain k is in motional class j (A_j) times the conditional probability that, if it is in class j , its order parameter deviates by $y_k - x_j$ from the class average. The Gaussian form is the least restrictive (maximum entropy) model for the distribution of values within a motional class given only that there must be some variance $\varepsilon^{2,22}$. The relative probability of obtaining the entire order parameter data set $\{y_k\}$, $k=1 \dots N$ for a given protein and set of conditions is then the product of the individual terms:

$$p(\{y_k\}|M, \{A_j, x_j\}, \varepsilon, I) = \prod_{k=1}^N \sum_{j=1}^M \frac{A_j}{\varepsilon \sqrt{2\pi}} \exp\left(-\frac{(y_k - x_j)^2}{2\varepsilon^2}\right) \quad (2)$$

With this likelihood function, the posterior probability that there are M motional classes (bands) is given by Bayes rule as

$$p(M|\{y_k\}, \{A_j, x_j\}, \varepsilon, I) \propto p(\{y_k\}|M, \{A_j, x_j\}, \varepsilon, I)p(M|I) \quad (3)$$

where the normalization factor $p(\{y_k\}|I)$ is omitted since it is constant across model parameters $M, \{A_j, x_j\}, \varepsilon$. For the purpose of determining M , the unknown band parameters $\{A_j, x_j\}$ and ε can be eliminated by marginalization. The evidence for a model with M motional classes is then

$$\frac{p(M|\{y_k\}, I)}{p(A_1|I) \dots p(A_M|I)p(x_1|I) \dots p(x_M|I)} \propto \int \dots \int p(M|\{y_k\}, \{A_j, x_j\}, \varepsilon, I)p(\varepsilon|I) d\varepsilon dA_1 \dots dA_M dx_1 \dots dx_M \quad (4)$$

where the prior $p(M|I)$ is absorbed into the constant of proportionality as it is uniform. Uniform priors, being the least restrictive, are also used for the unknown band parameters

$p(A_1|I) \dots p(x_M|I)$ over their physically allowable range $\{0 \dots 1\}$, subject to $\sum_{j=1}^M A_j = 1$. The prior for the width parameter was also assumed to be uniform in the range 0.1–0.3, which is effectively a measure of the band width. These priors can then be factored out of the integrand and absorbed in a constant of proportionality for each M . The multi-dimensional marginalization integration was performed by dividing the $2M$ dimensional parameter space into a uniform grid, evaluating the integrand using Eqs. 2 and 3 and summing. Grid dimensions of 8–10 were found to be both computationally tractable and numerically satisfactory, giving very similar results. The resulting model likelihood $p(M|\{y_k\}, I)$ was evaluated for $M=1 \dots 5$. Band numbers greater than five were not considered as the range of order parameter (0 to 1) is too small, and the data uncertainty too large, for this many bands to be meaningful. The maximum value of $p(M|\{y_k\}, I)$ gave the most likely number of bands in the data, M_m . For a given M_m , the band position and population parameters giving the maximum likelihood by Eq. 2 were determined by maximizing $p(\{y_k\} | M, \{A_j, x_j\}, \varepsilon, I)$

with respect to $\{A_j, x_j\}$ using the Newton-Raphson method. The starting values for the optimization were those giving the largest value of $p(\{y_k\} | M_m, \{A_j, x_j\}, \varepsilon, I)$ in the marginalization integration of Eq. 4. The uncertainty in optimal $\{A_j, x_j\}$ values was estimated by the ‘width’ of the 2M-dimensional likelihood function, which was obtained from the Hessian matrix computed at the likelihood function maximum.²² Uncertainties were displayed by contouring in parameter space at the 1-sigma level (i.e. where the likelihood has fallen to 1/e of its maximum value).

Using the maximum likelihood band parameters ($M, A_j, x_j, j=1, M$), each O^2 value y_k is then assigned to the band for which it has the maximum posterior probability calculated from Eq. 1. This is equivalent to defining the boundary between neighboring bands j and i at the O^2 value y where the posterior probabilities $p(y | M, A_j, x_j)$ and $p(y | M, A_i, x_i)$ given by Eq. 1 are equal (See, e.g. the dotted lines of Figure 1)

Molecular dynamics simulations & analysis

Molecular dynamics simulations of the seven proteins listed in Table I were carried out with NAMD2,²³ using the CHARMM27²⁴ all-atom parameter set and the TIP3P²⁵ water potential as described in detail elsewhere.¹² Simulation temperatures corresponded to those at which the NMR relaxation experiments were performed (Table I). Following equilibration runs of at least 1 ns, several 60 ns data production runs were performed with every subsequent 60 ns simulation starting from the final coordinates of the earlier run but with different initial velocities. For three protein systems, ubiquitin, the calmodulin-smMLCKp and calmodulin-nNOSp complexes, longer simulations were also run on the Anton supercomputer at the Pittsburgh Supercomputer Center run using the same force field and simulation conditions except for a non-bond cutoff of 14 Å. Other results of these simulations have been presented elsewhere.¹²

The Lipari-Szabo²⁶ squared generalized order parameters (O^2) were calculated from MD simulations by overlaying snapshots of the protein from the trajectories using a standard rigid-body alignment (C_α) procedure. For each snapshot, the unit vector along the methyl symmetry axis was obtained in terms of its vector components in Cartesian axes, x , y , and z . The O^2 parameter for a given methyl is then calculated using:²⁷

$$O^2 = \frac{3}{2} \left[\langle x^2 \rangle^2 + \langle y^2 \rangle^2 + \langle z^2 \rangle^2 + 2\langle xy \rangle^2 + 2\langle yz \rangle^2 + 2\langle xz \rangle^2 \right] - \frac{1}{2} \quad (5)$$

where $\langle \rangle$ indicates the average over the trajectory. In addition, side chain χ angles for each amino acid were calculated from the MD trajectories and classified as gauche⁺ [0°, 120°], trans [120°, 240°]; or gauche⁻ [240°, 360°]. The rotamer state of each side chain was determined from its chi angle(s) and the rotamer frequency histograms accumulated.

Modeling side chain packing interactions

The minimal model for the effect of side chain interactions on side chain dynamics and methyl order parameters was previously developed to explain anomalous temperature dependences of O^2 values in calmodulin.²⁸ Four side-chains are arranged in a cluster (Figure

7.) Each side chain i ($i=1\ldots 4$) can undergo angular motion over some range ($\pm \theta_{\max} = 180^\circ$) in a potential that consists of two components.

1. A step potential with $U = 0$ for $|\theta_s| \leq 30^\circ$, and $U = U_i$ for $|\theta_s| > 30^\circ$. This component depends only on single residue displacements, and U_i has the effect of favoring or disfavoring motion within different angular ranges, corresponding to different rotamer states.
2. A nearest neighbor overlap potential. The interaction of each pair of neighbors i and j is given an overlap penalty of $U_{ij} = 1$ kcal/mole if their angular displacement from each other is $(\theta_j - \theta_i) < \theta_{ij}$. The overlap thresholds were set to $\theta_{12} = \theta_{34} = 0^\circ$ and $\theta_{23} = \theta_{41} = 30^\circ$, which corresponds to tighter packing between residues 1 and 2, 3 and 4, and looser packing between residues 2 and 3, 4 and 1.

Precise numerical integration of the angular partition function $Z = \int d\theta_1 d\theta_2 d\theta_3 d\theta_4 e^{-U/kT}$ was performed by systematic evaluation of the total energy

$$U(\theta_1, \theta_2, \theta_3, \theta_4) = \sum_{i=1}^4 U_i + \sum_{i,j < i} U_{ij} \quad (6)$$

for all combinations of the four angles in increments of 1.5° . From the partition function the joint angular probability function $p(\theta_1, \theta_2, \theta_3, \theta_4)$ is obtained and then order parameters and any required average properties are calculated.

Results and Discussion

The tendency of O^2_{axis} values to cluster implies the existence of discrete sub-populations of motional behavior in protein methyl-bearing amino acid side chains. As such it is an example of the general problem of mixture model analysis, which is particularly amenable to Bayesian methods.²² Advantages of the Bayesian approach include the ability to calculate the relative likelihood of banding with no *a priori* assumption about the existence and number of bands, along with the simultaneous determination of their most likely position and uncertainty in position. Unlike our previous analysis of banding in a variety of contexts,^{4,10,14,20} no binning of the data -with its inevitably loss of resolution- is required. All the information in the data is retained. Nor is specification of a particular model for distribution of O^2 values within each band required. This is crucial for increased sensitivity of analysis of proteins with less obvious banding than ubiquitin. Analysis of the O^2_{axis} values was carried out for seven proteins for which comprehensive NMR relaxation data were available. These proteins varied in size, structure and overall dynamic behavior, as measured by the average value of O^2_{axis} for each protein (Table I). Extensive MD simulations of these proteins are also available.¹² A detailed comparison of the measured and calculated order parameter values has been published for this set of proteins.¹² In summary, the correlation of average O^2_{axis} values calculated from MD simulations with those measured from NMR experiments is very high while site-to-site agreement is somewhat lower and variable across this set of proteins. On average, however, the variation in protein side chain dynamics is well tracked by the MD simulations.

Banding of methyl order parameters

Analyzing the distribution of methyl order parameters for possible banding using the standard histogram-based method is often difficult since the observed banding is heavily dependent upon the bin size used for visualizing the distribution.²⁹ A slight change in the bin size can often alter the distribution enough to change the apparent number of bands. Therefore, we developed a Bayesian based approach for analyzing the distribution of order parameters with no *a priori* assumptions with respect to the number or position of bands, or even if there is banding at all. Importantly, this approach avoids binning. The analysis first determines the relative likelihood of having one or more bands in the data, irrespective of the positions and fractional occupancies of the bands. Figure 2 shows representative band likelihood plots for ubiquitin and α_3D . For the most likely number of bands (m), the most likely band centers and occupancies are then determined (Table II, Figure 3). We observe banding of the methyl order parameters ($m > 1$) for all seven proteins using either NMR measured order parameters or the MD calculated counterparts. A summary of the banding analysis of the experimental NMR data is presented in Table II. For each protein, the band centers and fractional populations for the model with the most likely number of bands are tabulated. In addition, the ratio of likelihoods for the most likely number of bands to the next most likely is given. For proteins with most probably three bands ($m = 3$), the next most likely model was always two bands, while the converse was true for the ‘two-band’ proteins. In all cases two or three bands was at least several orders of magnitude more likely than one band (i.e. than no banding at all), the exception being α_3D , for which no banding was only 50-fold less likely (Figure 2).

Although the O^2_{axis} values for each of the proteins are shown to fall into discrete clusters, there are subtle differences between the different proteins. For example, ubiquitin is observed to have three bands -J, α , and ω in increasing order of O^2_{axis} in the nomenclature of Igumenova et al.⁴ - whereas α_3D has only has two (Figures 1 and 2). The absence of the most rigid (ω band) in the latter is consistent with the view that α_3D is the most dynamic protein studied here and suggests that the differences in flexibility between different proteins result from differences in band population, as opposed to general changes in order parameter of all residues. Similarly, HEWL and Cyt c_2 , the two proteins with the most rigid methyl side-chain motions, present themselves with the majority of the methyl groups in the rigid ω -band with a very minor fraction in the most flexible J-band. Interestingly, like Cyt c_2 , CaM-nNOS and ALBP do not have the middle α band, whereas CaM-smMLCKp, which has very similar overall flexibility as judged by the *average* methyl order parameter has the α -band along with J and ω bands. This suggests an intriguing hypothesis that while the same protein (calmodulin in this case) can bind to multiple targets, namely the smMLCKp and nNOSp peptides here, it does so with different heterogeneous distribution of side chain dynamics which in turn directly influence the redistribution of conformational entropy. This redistribution of conformational entropy effectively aids in tuning the affinity of the protein to its different targets.^{10,11}

Overall, the centers of the J-, α -, and ω -bands show only minor variation whereas the corresponding populations differ more widely across the protein test set (Figure 3). The boundary between J- and α -bands is close to a $O^2_{axis} \sim 0.47$ while that between the α - and

ω -bands is an $O^2_{axis} \sim 0.68$, these boundaries are fairly consistent across all the proteins studied here. The over-arching conclusion is that the more objective Bayesian analysis confirms that banding is a real phenomenon produced by the underlying dynamics of the protein and that differences in banding are indicative of differences in structure, rigidity and binding behavior of the proteins.

Origin of banding

A physical basis for the banding observed above remains unclear. One possible explanation is the discrete physical differences in the methyl bearing side chains, namely that methyl groups are separated by 0, 1, 2 or 3 torsion angles from the backbone in Ala, Val, Ile, Leu and Met residues, respectively. We refer to this as the intrinsic residue type (IRT) model. Banding would arise not so much from the properties of the protein, but from shorter amino acids forming the ω band, longer residues forming the J band etc. Alternatively, it has been postulated that banding of O^2_{axis} values can occur because each amino acid type can show banding due to heterogeneous motional behavior.¹⁴ We refer to this model as the heterogeneous motion (HM) model. To distinguish between these two possibilities, we examined the distribution of each methyl type between the J, α and ω bands for both NMR derived and MD derived O^2_{axis} values for the seven proteins in Table I. Here, the gamma and delta methyl groups of Ile are distinguished as they are separated by one and two chi angles from the backbone, respectively. The frequency histograms of the experimental O^2_{axis} values indicate that each methyl type appears with appreciable frequency in all three bands (Figure 4). In particular the methyls of Leu and Ile, which typically account for 50% or more of all methyls in a protein, are well represented in all three bands. The most rigid class, Ala methyls, are under-represented in the most flexible J-band, as one might expect since it is rigidly attached to the backbone. Nevertheless, Ala methyls occur with almost equal frequency in the two more rigid α and ω bands. Conversely, Met methyls, are under-represented in the most rigid ω -band and occur with high frequency in both less rigid α - and J-bands. Since each methyl type occurs with high frequency in at least two of the bands, one concludes that for banding to occur, it must be manifest by the context of individual amino acids. In other words, we can reject the IRT model for banding, that different bands represent intrinsic differences in amino acids. We note that a previous study¹⁵ also found that O^2 values from each amino acid/methyl type were widely distributed over the range 0.2–0.9 spanning the three bands, although discrete banding was not evident in their pooled data from 18 proteins.

To address how otherwise chemically identical methyl groups of a given amino acid show a *discrete* distribution of O^2_{axis} in a given protein, we used MD simulations to examine the relationship between the methyl O^2_{axis} value, band identity, and conformational dynamics of the side chain. It has been previously shown that differences in O^2_{axis} are attributed largely to the differences in chi angle rotamer populations.^{15,28,30} We find that the rotamer distributions of methyl bearing side chains were characteristically different for the three bands. The J-band consists of residues which experienced frequent rotameric transitions, i.e. two or more rotamers experienced a probability of occupancy greater than ~30% whereas the ω -band containing the most rigid sites experienced few or no transitions which translated to a probability of occupancy of greater than ~90% in one rotamer. The α -band consisted of

residues that did not fit either of the above rotamer profiles, but otherwise was not starkly different compared to the J band: They underwent significant rotameric transitions, but not as frequently as for the J band. Indeed, if one has only the rotamer frequencies of a side chain one can predict what band its methyl O^2_{axis} will fall in using this 30%/90% rule with 94% accuracy for ubiquitin, and 75% accuracy for the calmodulin complexes (Figure 5). Although our aim here is not to predict band assignments but to understand their microscopic origins, this rotamer distribution analysis provides an alternative description of methyl residues banding. However, it still does not provide a concise explanation for the distinction between the J and the α -band or conversely, why some combinations of rotamer frequencies are disfavored. It does however suggest that correlations between different side chain motions, perhaps due to close packing, induce discrete and very discontinuous dynamic behavior. In other words, it appears that restrictions in the allowed sets of motions result in concomitant restrictions in the values that O^2_{axis} takes.

Following this line of analysis, we may rephrase the problem of banding origin by asking why some values of O^2_{axis} are markedly less frequent than others, i.e. focus on the gaps between bands rather than the bands themselves. A possible explanation for such gaps is a form of symmetry breaking arising from close packing interactions. Consider a single methyl type, say the δ -Leu. If the side chain made no interactions with other side chains, then the dynamics of every Leu residue would be almost identical on average, and one would expect the corresponding O^2_{axis} values to cluster narrowly round some intrinsic value. However, there are extensive steric interactions with other side chains, particularly in the well-packed core and perhaps even at the surface of a well-structured protein. There are three possible effects the interactions between two Leu side chains could have on corresponding O^2_{axis} values: i) no effect; ii) both O^2_{axis} values are affected the same way; iii) there are opposite effects on each O^2_{axis} value. The first possibility seems unlikely. Possibility (ii) would tend to make O^2 values of interacting residues the same, which is effectively the opposite of banding. Possibility (iii) means one O^2 value would increase and the other decrease, i.e. it would tend to create a *gap* between O^2 values. In effect one residue wins the competition between them for space in which to undergo rotamer transitions. Cases (ii) and (iii) also make opposite predictions about the relationship of the distance between residues and their O^2_{axis} values. The former predicts that pairs of residues in close contact will have more similar O^2 values than the average over all such pairs in a protein, while the latter predicts that the difference in O^2 values will be greater than average. Figure 6a shows the mean absolute difference in experimentally determined O^2_{axis} values ($|O^2_{axis}|$) between all pairs of identical methyl groups (β - β , γ - γ , δ - δ or ϵ - ϵ) binned by distance between methyls, aggregating data for all seven test proteins. The smallest distance bin with data corresponds to methyls in direct van der Waals contact. This analysis considers only pairs of methyls that are the same number of torsion angles from the backbone, in order to avoid biasing $|O^2_{axis}|$ by intrinsic side chain effects that might obscure the effects of side chain location and packing. Methyl pairs in direct contact (distance = 3.5 – 4.5 Å) have significantly more different O^2_{axis} values than the average over methyl pairs at all distances (significance at $p = 0.1$ by the T-test), but the effect drops off sharply. Pairs separated by an additional 1 Å or 2 Å (distances of 4.5–5.5 Å or 5.5–6.5 Å) apart are statistically no different from average ($p > 0.8$ by the T-test). This causes in turn a significantly greater tendency for contact methyl

pairs to fall into different bands (Figure 6b). The data clearly favor scenario (iii) above: a go-low/go-high form of symmetry breaking that favors gaps in O^2_{axis} values, and hence could explain banding. It also begins to explain why significant spatial clustering of similar “amplitudes” of motion are not generally seen in proteins.

A plausible mechanism for symmetry breaking

Analysis of the distance dependence of NMR-derived order parameters of pairs of identical methyl groups in the seven proteins in Table I clearly shows a tendency for residues in close contact to split their O^2_{axis} values into a high/low pair, but it provides no physical mechanism for this. Here we present a simple explanation based on steric interactions and competition for space to move in a crowded environment. To explain the anomalous temperature dependence of some methyl O^2_{axis} values in calmodulin, we developed a minimal model that captures the essential features of protein side chain motion in a tightly packed environment due to steric interactions and coupling with neighboring side chains.²⁸ Here we use this model to explore possible physical mechanisms giving rise to the symmetry breaking of methyl O^2_{axis} values.

In this model four side-chains are arranged in a planar cluster (Figure 7). Each can undergo angular motion over some range ($\pm \theta_{\max}$) subject to some intrinsic potential plus an overlap penalty of 1 kcal/mole if they get too close to either of their neighbors. The angular threshold at which the overlap penalty between two neighbors is incurred was set to correspond to tighter packing between residues **1** and **2**, **3** and **4**, and looser packing between residues **2** and **3**, **4** and **1**. A reference calculation was performed with no intrinsic potential term, so if there were no overlap term each residue would adopt ‘rotamer’ states uniformly within the range $-\theta_{\max} \rightarrow \theta_{\min} \rightarrow +\theta_{\max}$. The resulting O^2_{axis} values are given in the first line of Table III (Case A). Because of symmetry, each residue has one tight packing interaction, one loose packing interaction and no intrinsic potential and the order parameters are identical ($O^2_{axis} = 0.23$). Next, an intrinsic potential term of 0.8 kcal/mole was applied at residue **1**. This could represent some asymmetry due to interactions felt only by residue **1** with residues outside the cluster. It would favor residue **1** rotamer angles $\theta_1 < 30^\circ$. The resulting order parameters are given as Case B of Table III. Residue **1**’s order parameter increases to 0.54 and this causes a slight increase in the order parameter of its tight packed neighbor, residue **2**. However, the O^2_{axis} values of residues **3** and **4** decrease by 0.03 and 0.04 respectively, giving net splittings of 0.34 and 0.35. Introducing the same asymmetric interaction at residue **2** magnifies the splitting (Case C, Table III). Now O^2_{axis} for residues **1** and **2** increases to 0.57, while the O^2_{axis} for residues **3** and **4** decreases to 0.15, producing a splitting of 0.42. The explanation for this splitting is straightforward: If two residues **a** and **b** are in close contact in the interior of a protein, they compete for space in which to undergo torsion transitions, and both have higher order parameters than they would have outside the protein core. If, however, residue **a** is restricted by some interactions with other residues more than **b** is, **a**’s order parameter will rise, and then **b** can explore more effectively the space it shares with **a** so its order parameter decreases. Interestingly, while this effect arises from competing steric interactions between residues, it is not readily apparent from analysis of correlated motions. Table IV shows the fluctuation covariance matrices for Case A, full symmetry, and Case C, broken symmetry. The motions are only moderately correlated with

some positive and some negative values. However, in both cases the pattern of correlation coefficients is the same: residue 1's angular motion is positively correlated with residue 2, with which it shares a tighter packing interaction, and negatively correlated with residues 3 and 4. While there are some small changes, the pattern of positive and negative correlations is the same whether there is symmetry splitting of O^2 values or not. The evidence for symmetry splitting is apparent in the analysis of NMR data in Figure 6, but the model indicates that conventional correlative motion analysis of MD simulations is not the right strategy to detect relevant residue level dynamics.

Entropy redistribution upon DNA binding in CAP complexes

Recent work on the importance of protein internal motions has illuminated the role played by conformational entropy in tuning the affinity of proteins to ligands.⁹ Initial work on calmodulin bound to different target domains showed an interesting interplay between the three-band architecture in redistributing conformational entropy upon binding.¹⁰ Using the empirical entropy meter approach, Marlow and coworkers demonstrated that this statistical redistribution of internal protein motion reflected a significant change in conformational entropy.¹¹ Also using the empirical entropy meter approach and an array of mutational perturbations, Tzeng & Kalodimos found a significant contribution of conformational entropy to the free energy of DNA binding to the catabolite activation protein (CAP).²¹ Here we apply the Bayesian banding analysis here to the data of Tzeng & Kalodimos to gain some deeper insight in what heterogeneous dynamics reveals about the contribution of conformational entropy to the thermodynamics of binding.

Wild-type CAP and its mutants show banding of methyl O^2_{axis} parameters in both the unbound and DNA-bound states (Figure 7). However, there are subtle differences in the banding architecture of the different CAP mutants. For example, S62F, (T127L, S128I), and A144T display only two of the three bands compared to wild type CAP and other mutants of CAP. A summary of the banding analysis is presented in Table V. Intriguingly, we observe very strong correlations between the changes in band populations of the J-band (slope = -0.0078 ± 0.0014 kcal mol⁻¹; $R^2 = 0.76$) and ω -band (slope = $+0.0095 \pm 0.0014$ kcal mol⁻¹; $R^2 = 0.84$) with the total binding entropy (Figure 9). The populations of the J- and ω -bands are anti-correlated. This result is very similar to that observed in the previous calmodulin study¹⁰ and continues to suggest that the banding architecture contributes to the tuning of affinity. In contrast to the calmodulin complexes, the α -band is not correlated ($R^2 = 0.06$) with the total binding entropy. Interestingly, the entropy redistribution component due to band population changes occurs primarily through the α -band, unlike the calmodulin system where the populations of the α - and ω -bands are inversely correlated with that of the J-band. It is important to note that while the calmodulin peptide complexes represent naturally evolved diversity, the set of mutants of the CAP protein examined here largely represent laboratory constructed mutational perturbations. It remains to be seen whether the distinction in O^2 band shifts is a reflection of natural selection during evolution of proteins or simply an unrelated difference between calmodulin and the catabolite activation protein.

Acknowledgments

This work was supported by a grant from the Mathers Charitable Trust and NIH grant GM102447 and the Pittsburgh Supercomputing Center through NIH Award RC2GM093307 to CMU through the NRBSC.

References

1. Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. *Science*. 1991; 254:1598–1603. [PubMed: 1749933]
2. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*. 2002; 9:646–652. [PubMed: 12198485]
3. Blackledge M. Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Prog NMR Spectr*. 2005; 46:23–61.
4. Igumenova TI, Frederick KK, Wand AJ. Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem Rev*. 2006; 106:1672–1699. [PubMed: 16683749]
5. Jarymowycz VA, Stone MJ. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev*. 2006; 106:1624–1671. [PubMed: 16683748]
6. Palmer AG 3rd, Massi F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev*. 2006; 106:1700–1719. [PubMed: 16683750]
7. Baldwin AJ, Kay LE. NMR spectroscopy brings invisible protein states into focus. *Nat Chem Biol*. 2009; 5:808–814. [PubMed: 19841630]
8. Wand AJ. Dynamic activation of protein function: A view emerging from NMR spectroscopy. *Nat Struct Biol*. 2001; 8:926–931. [PubMed: 11685236]
9. Wand AJ. The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr Opin Struct Biol*. 2013; 23:75–81. [PubMed: 23246280]
10. Frederick KK, Marlow MS, Valentine KG, Wand AJ. Conformational entropy in molecular recognition by proteins. *Nature*. 2007; 448:325–U323. [PubMed: 17637663]
11. Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol*. 2010; 6:352–358. [PubMed: 20383153]
12. Kasinath V, Sharp KA, Wand AJ. Microscopic insights into the NMR relaxation based protein conformational entropy meter. *J Am Chem Soc*. 2013; 135:15092–15100. [PubMed: 24007504]
13. Mittermaier A, Kay LE, Forman-Kay JD. Analysis of deuterium relaxation-derived methyl axis order parameters and correlation with local structure. *J Biomol NMR*. 1999; 13:181–185. [PubMed: 20700817]
14. Lee AL, Wand AJ. Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature*. 2001; 411:501–504. [PubMed: 11373686]
15. Best RB, Clarke J, Karplus M. The origin of protein sidechain order parameter distributions. *J Am Chem Soc*. 2004; 126:7734–7735. [PubMed: 15212494]
16. Best RB, Clarke J, Karplus M. What contributions to protein side-chain dynamics are probed by NMR experiments? A molecular dynamics simulation analysis. *J Mol Biol*. 2005; 349:185–203. [PubMed: 15876377]
17. Showalter SA, Johnson E, Rance M, Bruschweiler R. Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. *J Am Chem Soc*. 2007; 129:14146–14147. [PubMed: 17973392]
18. Scouras AD, Daggett V. The dynamomics rotamer library: Amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. *Protein Sci*. 2011; 20:341–352. [PubMed: 21280126]
19. Glass DC, Krishnan M, Smith JC, Baudry J. Three entropic classes of side chain in a globular protein. *J Phys Chem B*. 2013; 117:3127–3134. [PubMed: 23421556]
20. Fu Y, Kasinath V, Moorman VR, Nucci NV, Hilser VJ, Wand AJ. Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc*. 2012; 134:8543–8550. [PubMed: 22452540]

21. Tzeng SR, Kalodimos CG. Protein activity regulation by conformational entropy. *Nature*. 2012; 488:236–240. [PubMed: 22801505]
22. Sivia, DS.; Skilling, J. *Data Analysis A Bayesian Tutorial*. Oxford University Press; Oxford: 2006.
23. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J Comp Chem*. 2005; 26:1781–1802. [PubMed: 16222654]
24. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoseck M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The biomolecular simulation program. *J Comp Chem*. 2009; 30:1545–1614. [PubMed: 19444816]
25. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983; 79:926.
26. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc*. 1982; 104:4546–4559.
27. Chatfield DC, Szabo A, Brooks BR. Molecular dynamics of staphylococcal nuclease: Comparison of simulation with N-15 and C-13 NMR relaxation data. *J Am Chem Soc*. 1998; 120:5301–5311.
28. Lee AL, Sharp KA, Kranz JK, Song XJ, Wand AJ. Temperature dependence of the internal dynamics of a calmodulin-peptide complex. *Biochemistry*. 2002; 41:13814–13825. [PubMed: 12427045]
29. Scott D. On optimal and data-based histograms. *Biometrika*. 1979; 10:605–610.
30. Chou JJ, Case DA, Bax A. Insights into the mobility of methyl-bearing side chains in proteins from (3)J(CC) and (3)J(CN) couplings. *J Am Chem Soc*. 2003; 125:8959–8966. [PubMed: 12862493]

Figure 1. Comparison of measured and simulated O^2_{axis} values for ubiquitin. Vertical dotted lines indicate boundaries between the three classes (bands) of motion identified by Bayesian analysis of the experimental NMR data (abscissa).

Figure 2.

Posterior log₁₀ likelihood ratios versus band number for NMR derived methyl O^2_{axis} order parameters for ubiquitin (□) and α₃D (■). Ratios are relative to the most likely band number for each protein.

Figure 3.

Summary of J (circles), α (squares) and ω (diamonds) band positions and populations of NMR-derived O^2_{axis} values determined by Bayesian analysis. Representative uncertainty in band position and population estimates is indicated by the 1-sigma likelihood contours for ubiquitin (dotted lines).

Figure 4. Distribution of amino acids among the J (bottom), α (middle) and ω (top) bands determined from Bayesian analysis of the NMR-derived O^2_{axis} values for the proteins in Table I.

Figure 5. Band assignment using Bayesian analysis of O^2_{axis} values ($\square$) of ubiquitin (left panel) and the CaM-smMLCKp complex (right panel). Band assignments using only χ torsion angle populations and the 30%/90% rule (see text) ($\bullet$).

Figure 6.

Mean absolute difference in methyl axis O^2 for methyl pairs on different residues with the same number of χ angles binned by distance between the methyl groups (left panel). The horizontal line indicates the mean difference (0.213) in O^2_{axis} over all methyl pairs within 11 Å. Standard error bars were obtained from the variance in $|\Delta O^2_{axis}|$ over all pairs of methyls. Percent of same- χ methyl pairs in a different band, binned by distance between the methyl groups (right panel). The horizontal line indicates the mean (72.9%) over all methyl pairs within 11 Å. Standard error bars were obtained from the variance over all pairs of methyl groups.

Figure 7.

Model for close packed side chain interactions between four residues. U_i is the intrinsic energy penalty for angular displacement of a side chain i from its equilibrium position by more than θ . U_{steric} is the energy penalty for overlapping any two side chains. θ_{ij} defines the density of packing between residues i and j . The steric overlap only occurs when the pair are displaced by more than θ_{ij} towards each other.

Figure 8.

Bayesian analysis of J (circles), α (squares) and ω (diamonds) banding in CAP and its mutants in both unbound (solid symbols) and DNA bound (open symbols) states. Most complexes displayed three bands except S62F, A144T, (T127L, S128I), which displayed two bands. See Table V. The boundaries of the bands are located at 0.47 for J and α , and at 0.68 for α and ω bands, respectively.

Figure 9.

Correlation of changes in band population with the total entropy of binding of a DNA oligomer to CAP. Significant linear correlations are observed for changes in the population of the J (circles), and ω (diamonds) bands but there is no clear correlation for the α -band (squares).

Characteristics of the protein set used for banding analysis

Table I

Protein ^a (PDB ID) ^b	Residues ^c	T (°C) ^d	Length ^e (ns)	$\langle O^2_{axis} \rangle^{expd}$	$\langle O^2_{axis} \rangle^{MD}$	R ^{2f}
α_3D (2A3D)	73 (27)	30	160	0.451	0.571	0.76
CaM-nNOSp (2060)	168 (62)	35	1120	0.534	0.560	0.50
CaM-smMLCKp (1CDL)	167 (67)	35	1280	0.583	0.562	0.62
ALBP (1LIB)	131 (40)	20	112	0.633	0.619	0.75
Ubiquitin (1UBQ)	76 (44)	25	260	0.664	0.629	0.85
HEWL (1ZLZA)	129 (50)	35	240	0.713	0.699	0.64
Cyt c ₂ (1C2R)	116 (47)	30	120	0.767	0.670	0.47

^a Abbreviations: ALBP, adipocyte lipid binding protein; Cyt c₂, cytochrome c₂; CaM-smMLCKp, calcium-saturated calmodulin (CaM) in complex with a peptide corresponding to the smooth muscle myosin light chain kinase calmodulin-binding domain (smMLCKp); CaM-nNOSp, calcium-saturated calmodulin in complex with a peptide corresponding to the neuronal nitric oxide synthase calmodulin-binding domain (nNOSp); HEWL, hen egg white lysozyme

^b PDB code of starting structure

^c Number of residues (number of methyl group axis order parameters measured)

^d Temperature of NMR experiments and simulations

^e Length of the simulation

^f Correlation between experimental and molecular dynamics O^2_{axis} values

Table II

Summary of Banding Analysis of NMR derived Order Parameters

Protein	m^a	ratio ^b	J-band		α -band		ω -band	
			center	fraction	center	fraction	center	fraction
α_3D	2	10	0.32	0.49	0.57	0.52		
CaM-nNOSp	2	7	0.36	0.51			0.72	0.50
CaM-smMLCKp	3	13	0.34	0.33	0.58	0.24	0.77	0.43
ALBP	2	3	0.34	0.36			0.79	0.65
Ubiquitin	3	23	0.26	0.21	0.60	0.33	0.89	0.47
HEWL	3	3	0.32	0.16	0.62	0.25	0.80	0.59
Cyt c2	2	4	0.38	0.23			0.82	0.77

^aMost likely number of bands.^bRelative likelihood compared to next most probably band number

Table III

Order parameters for a model of symmetry breaking

Conditions ^a	Order Parameter			
	Residue 1	Residue 2	Residue 3	Residue 4
A ($U_i = 0$, $i=1-4$)	0.23	0.23	0.23	0.23
B ($U_1 = 0.8$)	0.54	0.26	0.20	0.19
C ($U_1, U_2 = 0.8$)	0.57	0.57	0.15	0.15

^aEnergies in kcal/mole

Table IV

Residue-residue correlations for a model of symmetry breaking

Condition A ^a	1 ^b	2	3	4
1	52.8°	0.10	−0.10	−0.18
2	0.10	52.8°	−0.18	−0.10
3	−0.10	−0.18	52.8°	0.10
4	−0.18	−0.10	0.10	52.8°

Condition C	1	2	3	4
1	38.3°	0.08	−0.07	−0.14
2	0.08	38.3°	−0.14	−0.07
3	−0.07	−0.14	55.6°	0.07
4	−0.14	−0.07	0.07	55.6°

^aRefers to Table III^bResidue number. Diagonal term gives the mean square angular fluctuation of each residue. Off-diagonal terms give the fluctuation covariance in angular motion between the indicated residues.

Table V

Summary of banding analysis for CAP/DNA binding^a

Protein	Ligands ^b	J-band			α band			ω band		
		center	fraction	center	fraction	center	fraction	center	fraction	center
Wild type	cAMP	0.296	0.197	0.521	0.635	0.834	0.168			
Wild type	cAMP/DNA	0.343	0.196	0.511	0.378	0.787	0.427			
D53H	cAMP	0.303	0.283	0.583	0.485	0.817	0.233			
D53H	cAMP/DNA	0.360	0.369			0.733	0.640			
S62F	cAMP	0.373	0.326			0.722	0.674			
S62F	cAMP/DNA	0.396	0.44			0.705	0.561			
T127LS128I		0.286	0.100	0.571	0.329	0.857	0.571			
T127LS128I	DNA	0.266	0.115	0.558	0.322	0.842	0.555			
T127LS128I	cAMP	0.268	0.331	0.528	0.670					
T127LS128I	cAMP/DNA	0.311	0.445	0.601	0.556					
G141S	none	0.286	0.557	0.429	0.329	0.571	0.114			
G141S	DNA	0.286	0.671	0.429	0.214	0.571	0.114			
G141	cAMP	0.429	0.786	0.571	0.100	0.714	0.114			
G141S	cAMP/DNA	0.364	0.196	0.500	0.681	0.821	0.124			
G141S	cGMP	0.286	0.100	0.429	0.557	0.571	0.343			
G141S	cGMP/DNA	0.286	0.100	0.429	0.786	0.571	0.114			
A144T	none	0.330	0.277	0.579	0.724					
A144T	DNA	0.217	0.075	0.540	0.926					
A144T	cAMP			0.503	0.645	0.737	0.345			
A144T	cAMP/DNA	0.237	0.050	0.546	0.834	0.813	0.118			
A144T	cGMP	0.286	0.100	0.571	0.786	0.857	0.114			
A144T	cGMP/DNA	0.291	0.112	0.537	0.888					
Average Band Center		0.314		0.529		0.742				

^a taken from Tzeng & Kalodimos²¹^b all complexes with the same DNA oligomer