J Biomol NMR (2011) 50:59–70
DOI 10.1007/s10858-011-9504-6

ARTICLE

NMR order parameters calculated in an expanding reference frame: identifying sites of short- and long-range motion

Eric Johnson

Received: 21 December 2010 / Accepted: 9 February 2011 / Published online: 19 April 2011

Abstract NMR order parameters are calculated from molecular dynamics computer simulations of ubiquitin and the apo (Ca^{2+}-free) state of calbindin D_{9k}. Calculations are performed in an expanding reference frame so as to discriminate between the effects of short- and long-range motions. This approach reveals that the dominant contributions to the order parameters are short-range. Longer-range contributions are limited to specific sites, many of which have been recognized in previous studies of correlated motions. These sites are identified on the basis of an effective reorientational number, n_{eff}. Not only does this parameter identify sites of short- and long-range motion, it also provides a way of evaluating the separability condition that is key to the Lipari-Szabo model-free method. When analyzed in conjunction with the Prompers-Brüschweiler separability index, the n_{eff} values indicate that longer-range motions play a more prominent role in apo calbindin than they do in ubiquitin.

Keywords Protein dynamics · NMR order parameter · Molecular dynamics simulation · Ubiquitin · Calbindin D_{9k}

Introduction

NMR spin relaxation experiments have contributed significantly to our understanding of protein conformational dynamics on the pico- to nanosecond time scale (Cavanagh et al. 2007; Jarymowycz and Stone 2006; Palmer 2004; Korzhnev et al. 2001). A key development that facilitated the widespread application of these experiments was the Lipari-Szabo model-free method (Lipari and Szabo 1982). In this method of analysis, overall rotational diffusion of the molecule is characterized by an overall correlation time τ_m. Internal reorientation of an individual bond vector is similarly characterized by an effective internal correlation time τ_e as well as by a generalized order parameter S^2. The order parameter is a measure of the degree to which internal reorientation is restricted for a particular spin interaction. In the case of ^{15}N backbone relaxation, the dominant spin interaction vector corresponds to the ^{15}N–^{1}H bond vector. If a bond vector reorients isotropically in a molecular reference frame, then all orientations are equally probable in that reference frame and $S^2 = 0$. Alternatively, if the bond vector is fixed in the molecular reference frame, then $S^2 = 1$. A typical S^2 value for an N–H bond in a region of regular secondary structure is $\sim$0.85, although lower S^2 values are commonly observed in loops and near the chain termini (Goodman et al. 2000). These more flexible regions often play important functional roles in biologically significant processes such as binding, catalysis, and allostery (Jarymowycz and Stone 2006; Palmer 2004).

A key assumption in the Lipari-Szabo method is that overall rotation and internal reorientation are separable, representing two statistically independent processes (Wong et al. 2009; Wong and Case 2008; Meirovitch et al. 2006; Vugmeyster et al. 2003; Prompers and Brüschweiler 2002). Under this assumption, the effects of internal reorientation are clearly not global. Any global effect would likely alter the molecule's diffusion tensor and thereby violate the separability condition. Is it possible, however, that internal

Electronic supplementary material The online version of this article (doi:10.1007/s10858-011-9504-6) contains supplementary material, which is available to authorized users.

E. Johnson (✉)
Department of Chemistry and Physical Sciences,
College of Mount St. Joseph, Cincinnati, OH 45233, USA
e-mail: eric_johnson@mail.msj.edu

reorientation acts on a scale that is not entirely local? Might internal reorientation act on a semi-local or an intermediate length scale? If so, is the separability condition still fulfilled, and are the order parameters still meaningful? In order to address these questions, a useful distinction between short- and long-range motion was proposed by Brüschweiler (1995). In short-range motion, the ^{15}N–^{1}H bond vector reorients as a result of fluctuations in nearby internal coordinates. For example, the ^{15}N–^{1}H bond vector might reorient in the molecular reference frame as a result of fluctuations in the adjacent dihedral angles ϕ_i and ψ_{i-1} (Li et al. 2009). In contrast, long-range motion is attributed to fluctuations in more distant internal coordinates. Hinge-bending and other rigid body-type motions are potentially long-range, as may be the fraying effect that is commonly observed near the chain termini. These examples demonstrate that long-range contributions are sometimes projected onto a local environment (Johnson et al. 2008). Although the ^{15}N–^{1}H bond vector acts as a local probe, the origin of some effects that contribute to the corresponding S^2 value are nonlocal. The goal of this study is to provide a framework in which to detect these long-range effects.

The starting point for this study are molecular dynamics (MD) computer simulations of the proteins ubiquitin and apo (Ca^{2+}-free) calbindin D_{9k}. Order parameters are calculated from the simulations using a novel method that involves an expanding reference frame. This approach discriminates between short- and long-range motion by initially limiting the analysis to include only short-range effects and then gradually introducing long-range effects in a systematic manner. A key result is an effective reorientational number that indicates the length scale over which internal reorientation typically acts for a particular site. Ubiquitin and apo calbindin serve as useful model systems for this initial study due to the fact that the proteins have the same number of residues. Differences in the effective reorientational numbers can, therefore, be attributed to actual differences in long-range motions, rather than to scaling effects related to the chain length. A comparison of the effective reorientational numbers reveals that, not only is apo calbindin more flexible than ubiquitin, but the motions in calbindin are more long-range.

Methods

Molecular dynamics simulations

An X-ray structure (PDB code 1UBQ) was used as the starting structure for the ubiquitin simulations (Vijay-Kumar et al. 1987). Although no X-ray structure is available for apo calbindin, an NMR ensemble of structures (PDB code 1CLB) is available (Skelton et al. 1995). The THESEUS software (Theobald and Wuttke 2008; Theobald and Wuttke 2006) was used to select a representative member from the NMR ensemble. On the basis of this initial analysis, model 10 was selected as the starting structure for the calbindin simulations. In the case of calbindin, it was also necessary to add an N-terminal methionine. AmberTools (Case et al. 2005) was used to place each protein in a truncated octahedral box of TIP3P water molecules, with a minimum distance of 12 Å between the protein atoms and the edge of the box. No counterions were added to ubiquitin due to the fact that the net charge on the system was already zero. Seven Na^+ counterions were added, however, in order to neutralize the net charge on calbindin. The simulations were performed using the Amber10 software with the ff99SB force field (Hornak et al. 2006) and particle mesh Ewald periodic boundary conditions. The systems were energy minimized and heated to a target temperature of 300 K, which is consistent with the previous NMR relaxation studies (Lienin et al. 1998; Akke et al. 1993). Ten different seed values were used for the assignment of initial velocities, which then led to the production of ten trials for each protein. All of the quantities reported in this study have been averaged over ten trials. An equilibration period followed in the NPT ensemble, first for 200 ps and then for 100 ps with 1 and 0.1 kcal mol^{-1} Å^{-2} restraints, respectively, on all protein atoms. The restraints were then removed, and the systems were further equilibrated for 1 ns. Finally, a production phase was conducted for 4 ns in the NPT ensemble, which is comparable to the overall correlation times of the two proteins (4.03 and 4.10 ns for ubiquitin and calbindin, respectively) (Lienin et al. 1998; Akke et al. 1993). During the production phase, hydrogen atoms were constrained by the SHAKE algorithm, and a 2 fs time step was used along with an 8 Å direct-space cutoff for Lennard–Jones and electrostatic interactions. This procedure provided 10×4 ns $= 40$ ns of production phase data for each protein, which is comparable to other recent studies (Genheden et al. 2010).

Order parameter calculations

Order parameters were calculated using the isotropic reorientational eigenmode dynamics (iRED) method of Prompers and Brüschweiler (Prompers and Brüschweiler 2002). An isotropically averaged covariance matrix **M** was constructed whose elements are

$$M_{ij} = \frac{1}{2}\left\langle 3\left(\mathbf{e_i}\mathbf{e_j}\right)^2 - 1\right\rangle \quad (1)$$

where $\mathbf{e_i}$ and $\mathbf{e_j}$ are normalized bond vectors taken from the same snapshot. The angular brackets indicate averaging

over all snapshots of the MD trajectory. Note that it is not necessary to superimpose the snapshots prior to the analysis. The covariance matrix contains only inner products and therefore is rotationally invariant. In fact, separability is not assumed a priori. As a result, it is not necessary to decide which method and which set of atoms are most appropriate for the removal of overall rotation. Next, a principal component analysis was performed by solving the equation

$$\mathbf{M}|m\rangle = \lambda_m|m\rangle \tag{2}$$

where $m = 1,\ldots, N$. N is the number of bond vectors included in the calculation; λ_m are the eigenvalues (mode amplitudes); and $|m\rangle$ are the corresponding eigenvectors. The S_j^2 for vector j was then calculated as

$$S_j^2 = 1 - \sum_{m=1}^{N-5} \lambda_m \left| |m\rangle_j \right|^2 \tag{3}$$

The sum in Eq. 3 extends over the internal modes, which are the eigenvectors with the $N - 5$ smallest eigenvalues.

Expanding reference frame

Order parameters were calculated multiple times over an expanding reference frame. This approach involves a large number of calculations, which is facilitated by the computational efficiency of the iRED method. An initial iRED calculation was performed that included bond vectors from a single peptide plane, located at position p along the protein sequence. The following vectors were included in the calculation: N–H, C^{α}–C′, C′–O, C′–N, N–C^{α}, the cross-product of the N–H and N–C^{α} vectors, and the cross-product of the C^{α}–C′ and C′–O vectors. In the case of proline, the C^{δ} atom was substituted for the missing H atom. Next, the calculation was performed over a larger set of bond vectors. In addition to the original set of vectors, the second calculation included the bond vectors and cross-products from the two adjacent peptide planes, $p - 1$ on the N-terminal side and $p + 1$ on the C-terminal side. The expansion continued, with a peptide plane added on both the N- and C- terminal sides in each subsequent iRED calculation. Once the expansion reached the N- or the C-terminus, the remaining calculations proceeded in the direction in which expansion was still possible. Although multiple bond vector types were included in these calculations, the order parameters that are reported below correspond specifically to the N–H bond vectors (for which experimental data are available).

S^2 as a function of the number of peptide planes

The overall strategy described above was repeated so that each peptide plane served as the starting point for the expansion. These calculations provided a series of (n, S^2) data points for each plane. The independent variable n was the number of planes included in the calculation. The data points were then fit to the following empirical model.

$$S^2 = a_0 + \sum_{k=1}^{2} a_k \exp\left(-\frac{n-1}{n_k}\right) \tag{4}$$

This model includes five fitting parameters (a_0, a_1, a_2, n_1, and n_2), which are specific for a given plane. The fitting parameters were obtained for each plane by least-squares optimization subject to the following constraints: $0 \le a_0$, $0 \le a_k$, $a_0 + a_1 + a_2 \le 1$, and $0 < n_k \le n_T$. The total number of peptide planes n_T is 75 for both ubiquitin and calbindin, and the fitting parameters were sorted such that $n_1 < n_2$. The fitting procedure was performed in MATLAB using the *fmincon* function. An initial fit was performed using a monoexponential function whose parameters were subsequently used to specify a grid of starting values for the fit to the biexponential function represented in Eq. 4. The best-fit parameters provided the lowest sum-of-squares over this grid of starting values. Note also that the fits were performed separately for each of the trials (simulations with different starting velocities), of which there were ten for each protein. As a result, the fitting parameters that are reported for a given peptide plane have been averaged over ten trials.

Results and discussion

Comparison with experiment

The purpose of the current study is to calculate order parameters over an expanding reference frame. It is important to first demonstrate, however, that the N–H order parameters, when calculated in the context of the entire protein, are consistent with the experimental values reported in the previous NMR relaxation studies (Lienin et al. 1998; Akke et al. 1993). In what follows, S_T^2 will refer to the typical case in which all 75 of the peptide planes in either ubiquitin or calbindin are included in the calculations. In Fig. 1, the MD-derived S_T^2 values are plotted as a function of the residue number, along with the experimental trend lines.

The order parameters are relatively high throughout the secondary structural elements of the two proteins. Several of the loops, linkers, and termini exhibit much lower order parameters in comparison. In several cases, the MD-derived values are lower than what was reported experimentally. These discrepancies are relatively isolated across ubiquitin, primarily affecting residues 11, 33, 36, 41, 54, and 75. In calbindin, however, the discrepancies are

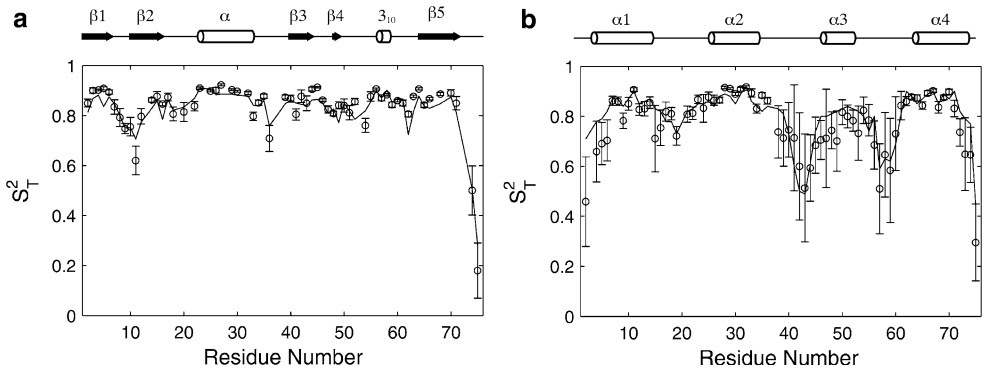


Fig. 1 Plot of S_T^2 as a function of residue number for ubiquitin (**a**) and calbindin (**b**). For a given residue, the *circle* and *error bar* correspond to the average and standard deviation, respectively, over ten MD simulations. The *curves* represent the experimental values for each of the two proteins (Lienin et al. 1998; Akke et al. 1993). The figure includes only those residues for which experimental data are available. In ubiquitin, the structural elements include: strand 1 (residues 1–7), strand 2 (10–17), α helix (23–34), strand 3 (40–45), strand 4 (48–50), 3_{10} helix (56–59), and strand 5 (64–72). In calbindin, the structural elements include: helix 1 (3–15), loop 1 (16–24), helix 2 (25–35), linker (36–45), helix 3 (46–53), loop 2 (54–62), and helix 4 (63–74)

more widespread, involving the termini, loops 1 and 2, and helix 3. These same regions also tend to exhibit large standard deviations over the ten trials, which will be reflected in the fitted parameters from Eq. 4. Despite these discrepancies, the correlations between the MD-derived and experimental order parameters are still relatively high ($r = 0.95$ for ubiquitin and $r = 0.85$ for calbindin), indicating that the simulations are a reasonable source of additional information.

Effect of the expanding reference frame

Before examining individual cases, it is useful to consider the general trend that is observed over the course of an expansion. At the start, the iRED calculation includes only the N–H bond vector of interest and the other vectors derived from the same peptide plane. This system of vectors reorients over the course of the simulation almost as if it were a rigid body. Therefore, the N–H order parameter is expected to be close to 1 in the initial calculation. The value that is observed in practice, however, is likely to be slightly less than 1 due to small fluctuations in the peptide plane geometry. As an increasing number of peptide planes is included in the iRED calculations, the expanding system of vectors behaves less and less like a rigid body. A shift thereby occurs in which motion that was previously assigned to overall rotation is reassigned to internal reorientation. The N–H order parameter responds accordingly by decreasing with the increasing size of the system. Eventually, the order parameter approaches the value that is traditionally calculated over the total length of the protein, which is designated here as S_T^2.

Although this general trend is perhaps expected, it is not known how quickly the order parameter approaches S_T^2. Nor is it known if the order parameter decays uniformly across a protein. This information is provided in Fig. 2, which plots the N–H order parameter as a function of the number of planes n that was included in the calculation. The best-fit curve is also represented in these plots (where applicable). The biexponential model represented in Eq. 4 provides reasonable fits for a majority of the residues. In what follows, the plane of interest will be identified by the residue number of the corresponding N atom. Residues E64 of ubiquitin and S2 of calbindin decay to very different S_T^2 values (0.91 and 0.32, respectively), but the biexponential model provides a reasonable fit in both cases (Fig. 2a, b). In some cases, irregularities are noted for one or more of the data points early in the expansion. Residue L43 of ubiquitin, for example, exhibits an initial decrease, followed by a slight increase in S^2 (Fig. 2c). Only a small number of cases (21 of 750 in ubiquitin and 23 of 750 in calbindin) were excluded from the analysis due to more significant irregularities. (750 corresponds to 10 trials with 75 residues per trial.) These cases were identified by the fact that the sum-of-squares did not decrease in going from the mono- to the bi-exponential model. In this group, neither model is able to account for the irregularities that are observed. Residue E48 of calbindin provides an example of this behavior in Fig. 2d. In ubiquitin, 10 of the 21 outliers originate from strand 1, while an additional 6 outliers originate from the C-terminus. In calbindin, the outliers are concentrated near the N-terminus and helix 1, which account for 4 and 6 outliers, respectively. The prolines are also over-represented, accounting for 5 of calbindin's 23 outliers. Additional information regarding the outliers is provided as Supplemental Material.

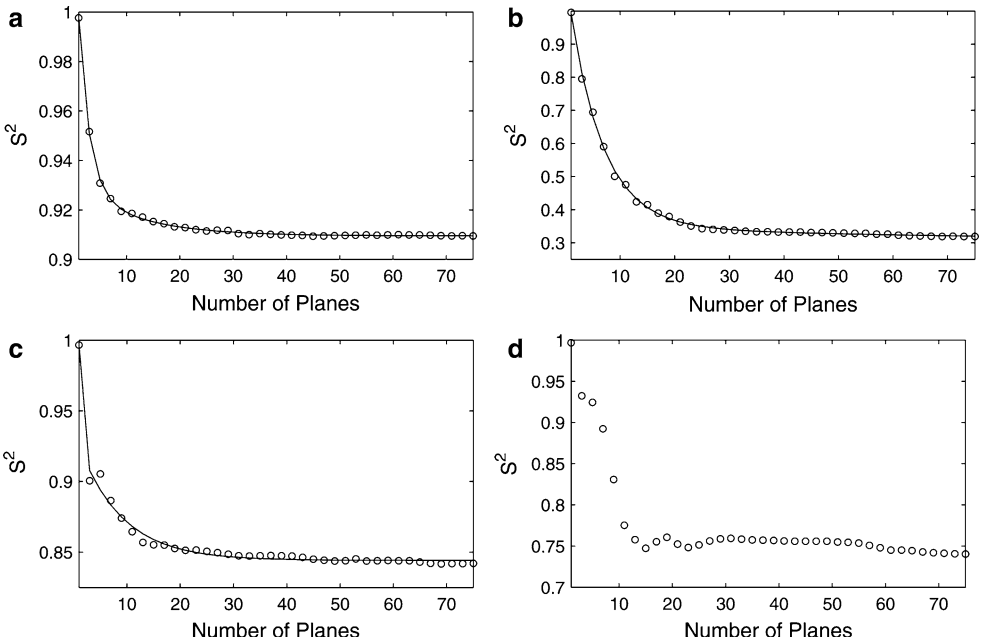


Fig. 2 Plot of S^2 as a function of the number of planes included in the iRED calculation. **a** and **c** represent data from Trial 1 for residues E64 and L43, respectively, of ubiquitin. **b** and **d** represent data from Trial 1 for residues S2 and E48, respectively, of calbindin. The best-fit curve from Eq. 4 is provided (where applicable)

Trends among the best-fit parameters

The fitting parameters provide a novel source of potentially useful information about the motions reflected in the NMR order parameters. Each of the fitting parameters is plotted in Fig. 3 as a function of residue number. Although the expansion occurs over a finite number of planes, the model represented in Eq. 4 provides a parameter a_0, which represents the value of S^2 at $n = \infty$. In practice, a_0 is very close to S_T^2. The rmsd between these two quantities is 0.01 for both ubiquitin and calbindin (data not shown), indicating that the order parameter nearly reaches its plateau value over the course of the expansion. Equation 4 also provides the value of S^2 at $n = 1$, which corresponds to the sum $(a_0 + a_1 + a_2)$. It was noted in the preceding section that S^2 is likely to be close to 1 when $n = 1$. This trend is confirmed among the data, which give a value of 0.997 ± 0.002 for the sum $(a_0 + a_1 + a_2)$, again for both ubiquitin and calbindin (data not shown). The values obtained for these two limiting cases suggest that the quantity $(1 - S_T^2) \approx (a_0 + a_1 + a_2) - a_0 = a_1 + a_2$. This result helps to explain an additional observation, which is the fact that the parameter a_1 exhibits a strong correlation with $(1 - S_T^2)$. Recall that the fitting parameters were sorted so that $n_1 < n_2$. The parameter a_1 may, therefore, be regarded as a short-range contribution to $(1 - S_T^2)$, while the parameter a_2 is a longer-range contribution. If the contribution from a_2 were constant, then the above equation would suggest an almost linear relationship between $(1 - S_T^2)$ and a_1, with a slope of 1 and a y-intercept corresponding to a constant value of a_2. Although this scenario is not realized in practice, the a_2 values are in fact much more narrowly distributed than are the values of a_1. Many of the a_1 values in Fig. 3c and d are less than 0.2, but some a_1 values are actually greater than 0.3. The a_2 values in contrast are all less than 0.3 (Fig. 3e, f). Overall, the data indicate that the variation in $(1 - S_T^2)$ that is observed across the protein is reflected primarily in the value of a_1.

Figure 4 provides a correlation plot between $(1 - S_T^2)$ and a_1 for each of the two proteins. A linear fit of the data gives a slope of 1.03 and a y-intercept of 0.063 for ubiquitin (Fig. 4a). The calbindin data yield a similar result, with a slope is 1.10 and a y-intercept is 0.084 (Fig. 4b). In each case, the slope is slightly higher than the value of 1 proposed above. The y-intercepts, however, are very close to the average a_2 values, which are 0.065 and 0.097 for ubiquitin and calbindin, respectively. These findings suggest that site-specific differences in $(1 - S_T^2)$ are primarily due to differences in the short-range parameter a_1, rather than in the longer-range parameter a_2.

An effective reorientational number

The remaining two fitting parameters, n_1 and n_2, are plotted as a function of residue number in Fig. 3g–j. These two parameters operate over different length scales, but they exhibit a similar trend, which is more clearly defined by the introduction of a new parameter, the effective reorientational number. The effective reorientational number n_{eff} provides a more general measure of the length scale over which the order parameter decays to its plateau value. This new quantity will be used to characterize the extent to

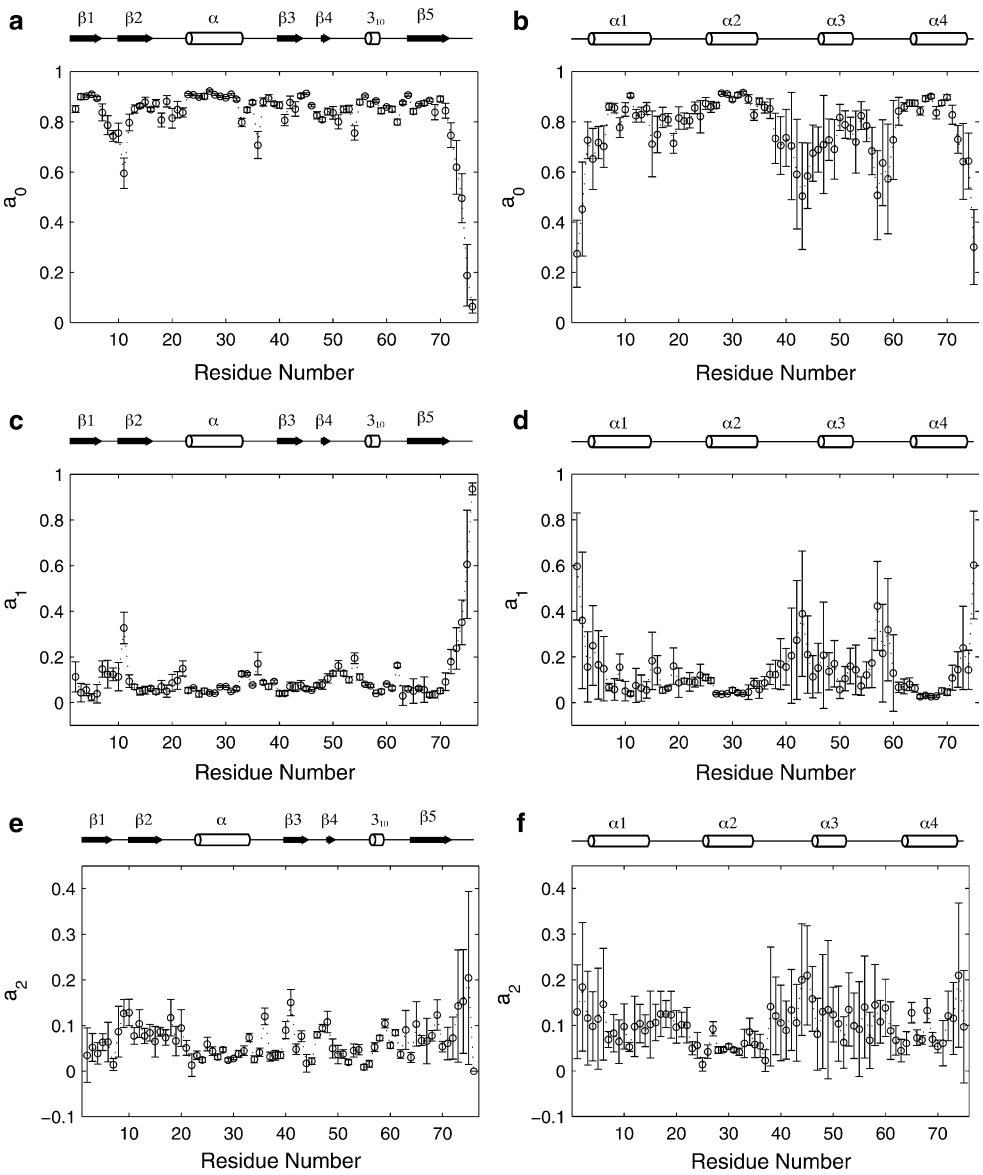


Fig. 3 Plot of the best-fit parameters as a function of residue number for ubiquitin (**a**, **c**, **e**, **g**, **i**) and calbindin (**b**, **d**, **f**, **h**, **j**). For a given residue, the *circle* and *error bar* correspond to the average and standard deviation, respectively, over the ten MD simulations

which nonlocal effects contribute to the order parameter. The quantity was calculated as follows. First, the following definite integral was evaluated

$$A = \int_1^{n_T} \sum_{k=1}^{2} a_k \exp\left(-\frac{n-1}{n_k}\right) dn = \sum_{k=1}^{2} a_k n_k \left[1 - \exp\left(-\frac{n_T - 1}{n_k}\right)\right] \quad (5)$$

which corresponds to the area A between the best-fit curve and the plateau at $S^2 = a_0$. It was previously mentioned that the fitted value of S_T^2 is just slightly larger than a_0. The gap, albeit small, between S_T^2 and a_0 makes an unwanted contribution to the area A. It is possible to remove this unwanted contribution by calculating the area A'

$$A' = A - \left(S_T^2 - a_0\right)(n_T - 1) \quad (6)$$

The effective reorientational number is then calculated by dividing the area A' by the difference between the fitted S^2 values at $n = 1$ and at $n = n_T$.

$$n_{eff} = \frac{A'}{a_0 + a_1 + a_2 - S_T^2} \quad (7)$$

One can simplify the above expression given certain approximations. If, for example, one were to set the fitted S_T^2 value equal to a_0 and ignore the fact that the expansion occurs over a finite number of planes (thereby changing the upper limit of integration in Eq. 5 from n_T to ∞), then Eq. 7 simplifies to

$$n_{eff} = \sum_{k=1}^{2} \left(\frac{a_k}{a_1 + a_2}\right) n_k \quad (8)$$

Fig. 3 continued

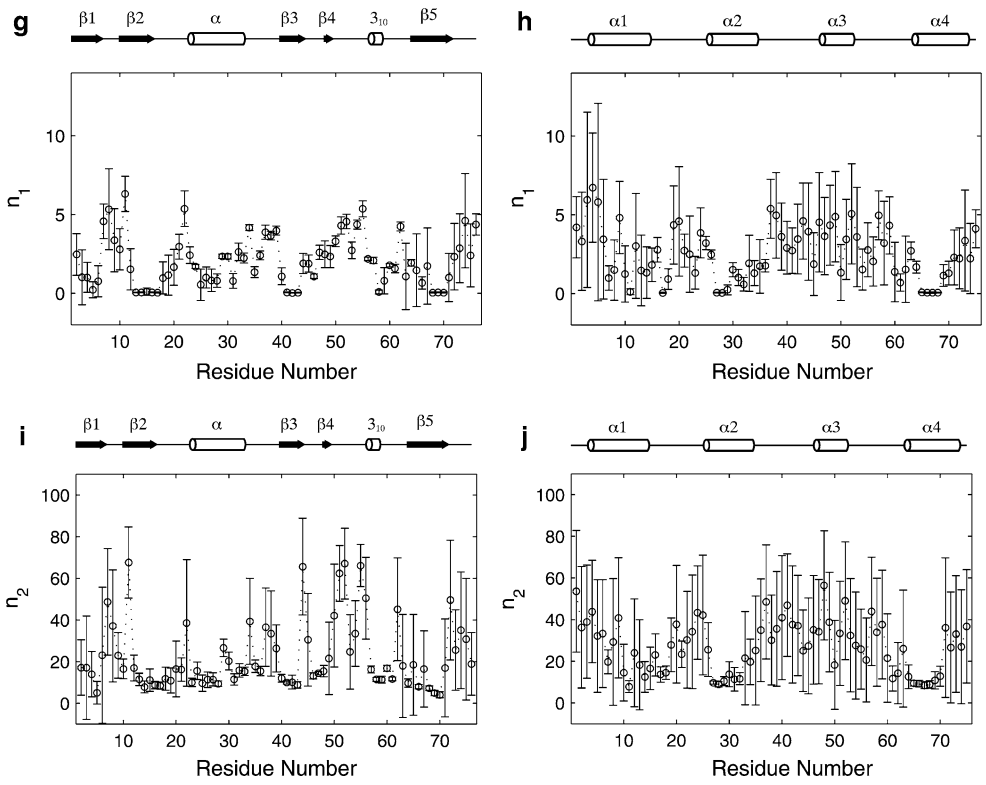
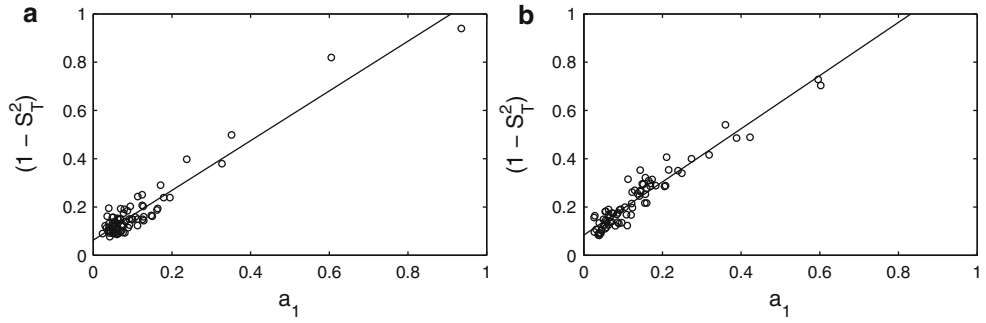


Fig. 4 Correlation plot between $(1 - S_T^2)$ and a_1 for ubiquitin (**a**) and calbindin (**b**). These quantities have been averaged over the ten MD simulations for each residue. In each case, the *line* represents a linear fit of the data



Equation 8 serves as a useful reference in that it expresses n_{eff} as a weighted-average of the n_1 and n_2 values. Note, however, that this simplified expression overestimates the values of n_{eff}. For that reason, the results that follow were obtained using Eq. 7.

The n_{eff} values provided by Eq. 7 are plotted as a function of residue number in Fig. 5. Figure 5 reveals that the n_{eff} values tend to be lower in ubiquitin than in calbindin. The median n_{eff} value for ubiquitin is 6.46, while that for calbindin is 7.97. Furthermore, only 2 residues in ubiquitin have n_{eff} values higher than 10, but there are 11 such residues in calbindin. This trend is significant given the fact that the two proteins have the same total number of residues. If this were not the case, then any differences in n_{eff} could perhaps be attributed to a scaling effect related to the chain length. By choosing to apply this method to two proteins with the same number of residues, it is much more likely that the differences in n_{eff} reflect actual differences in longer-range motions. This conclusion is also supported by the fact that the two proteins satisfy the separability condition to different degrees. A well-established measure of the separability between overall rotation and internal reorientation is the Prompers–Brüschweiler separability index. (See Eq. A26 from Prompers and Brüschweiler 2002.) It is high in cases where the separability condition is considered to be valid. The values are 6.00 ± 0.17 and 4.54 ± 0.41 for ubiquitin and calbindin, respectively. (In order to be consistent with the established literature, the separability indices were calculated here using only the N–H bond vectors.) Both of these observations, higher n_{eff} values and a lower separability index in calbindin, suggest that longer-range motions play a more prominent role in

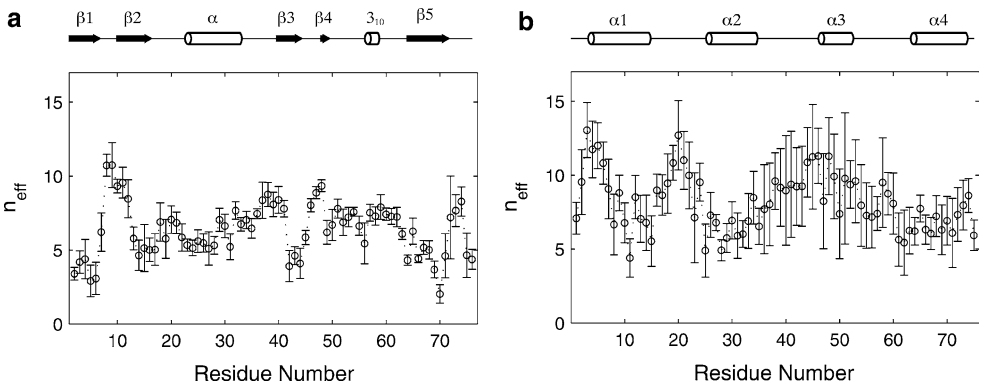


Fig. 5 Plot of n_{eff} (from Eq. 7) as a function of residue number for ubiquitin (**a**) and calbindin (**b**). For a given residue, the *circle* and *error bar* correspond to the average and standard deviation, respectively, over the ten MD simulations

calbindin than they do in ubiquitin. It should be noted, however, that the n_{eff} values in both proteins are much less than the total number of peptide planes, suggesting that, even in the case of calbindin, internal reorientation acts on a fairly limited length scale. At present, there are too few examples in the literature to know when the separability condition actually begins to break down. New insights may emerge by examining how the effective reorientational number varies in a system for which the separability is known to be particularly low. The disordered A-state of ubiquitin, for example, gives a separability index of 2.51 (Prompers and Brüschweiler 2002). Future studies are needed in order to determine if this low separability index is accompanied by high n_{eff} values.

It is also useful to analyze the current results in light of the ubiquitin and calbindin structures, which are provided in Fig. 6. Both of these proteins are considered to be single-domain proteins. The calbindin structure, however, can be divided into two EF-hand motifs connected by a ten-residue linker (Skelton et al. 1995). Reorientation of the EF-hands relative to one another can introduce a long-range effect that, among the proteins studied here, is specific to calbindin. This effect likely contributes to the systematically higher n_{eff} values that are observed in calbindin compared to ubiquitin.

New insights are also obtained by examining how the n_{eff} values vary within these proteins. Figure 6 represents the n_{eff} values using a color gradient mapped onto the ubiquitin and calbindin structures. The highest n_{eff} value for each protein is represented in red, the median value in white, and the lowest value in blue. In ubiquitin, the only residues that exhibit n_{eff} values greater than 10 are residues 8 and 9, which are located in the turn of the N-terminal β-hairpin (Fig. 6a). Interestingly, this segment has been recognized in previous studies as a site of correlated motion (Prompers and Brüschweiler 2001; Showalter and Brüschweiler 2007). In a recent validation study of the AMBER99SB force field, Showalter and Brüschweiler found a significant correlation in the internal reorientation of the N–H bond vectors associated with residues 8 and 9 of ubiquitin (Showalter and Brüschweiler 2007). This correlation was apparent in a modified iRED covariance matrix constructed from the $N - 5$ internal modes. This trend is also observed in the present study. A covariance matrix for internal reorientation was constructed for each of the 10 ubiquitin trials. In 1 of those 10 trials, the largest (most positive) off-diagonal element is in fact associated with residues 8 and 9. In an additional 4 of the 10 trials, the element associated with residues 8 and 9 is the second largest off-diagonal element (data not shown). Unfortunately, it is difficult to

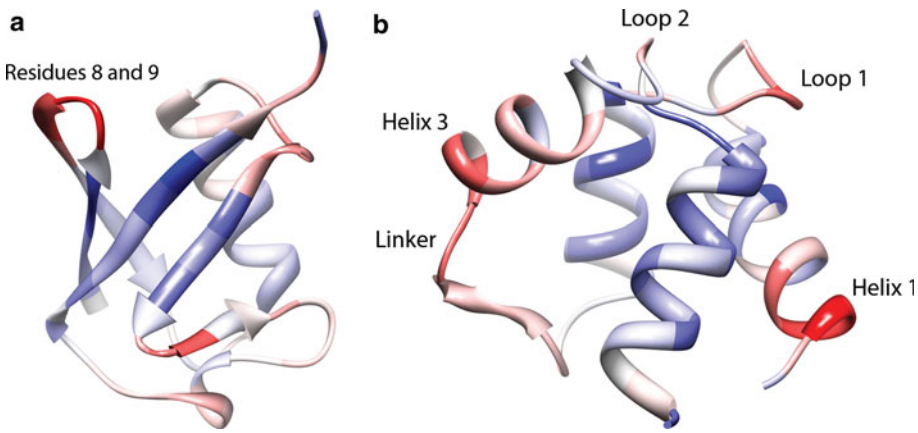


Fig. 6 The n_{eff} values plotted onto ribbon diagrams of ubiquitin (**a**) and calbindin (**b**). The highest n_{eff} value for each protein is represented in *red*, the median value in *white*, and the lowest value in *blue*. The images were produced using the UCSF Chimera package (Pettersen et al. 2004)

characterize these correlations experimentally. Future progress in this area will likely rely upon newly emerging techniques, such as the measurement of C′N multiple-quantum relaxation rates. These rates reflect correlated chemical shift modulations of neighboring $^{13}C'$ and ^{15}N nuclei. Abergel and coworkers recently applied this method to a lanthanide-substituted form of calbindin (Mori et al. 2010), with results that are in many ways consistent with the current study. In calbindin, the most prominent sites of longer-range motion are the start of helix 1, Ca^{2+}-binding loop 1, the end of the linker, and the start of helix 3 (Fig. 6b). These same sites are highly represented among the 16 outlying residues that were identified by Abergel and coworkers. The most significant deviation from the average rate was observed for residue E5 of helix 1 whose n_{eff} value in the present study is the third highest in calbindin. An additional 8 of the 16 outliers from the previous study map to either Ca^{2+}-binding loop 1, the end of the linker, or the start of helix 3. Collectively, these observations indicate that sites of longer-range motion are often outliers in computational and experimental studies of correlated motion.

Figures 5b and 6b also provide insight into the mechanism by which calbindin binds calcium. Each EF-hand exhibits a helix-loop-helix motif that binds Ca^{2+} via the loop segment (Kesvatera et al. 2001; Linse et al. 1991, 1987). As was noted above, two of the most readily identified sites of longer-range motion in calbindin are the start of helix 1 and Ca^{2+}-binding loop 1, both of which are located in the N-terminal EF-hand. Although a similar trend is observed in the C-terminal EF-hand at the start of helix 3, Ca^{2+}-binding loop 2 (residues 54–62) does not exhibit particularly high n_{eff} values. This observation is analyzed here in the context of the many well-documented differences between the two sub-domains (Mäler et al. 2000; Malmendal et al. 1998; Bertini et al. 2002). The N-terminal EF-hand, for example, is referred to as a pseudo-EF-hand. A pseudo-EF-hand binds Ca^{2+} mainly through backbone carbonyl oxygens and is highly selective for Ca^{2+}. The binding loop in a pseudo-EF-hand is also fairly rigid, exhibiting order parameters that are in some cases comparable to what is observed in the α-helices (Fig. 1b). This scenario differs from that of the C-terminal EF-hand, which is considered a consensus EF-hand. A consensus EF-hand binds Ca^{2+} primarily with side-chain carboxylate groups. It is less selective for Ca^{2+}, and its binding loop is relatively flexible. The picture that has emerged over the course of many studies is that the N-terminal pseudo-EF-hand is poised and ready to bind Ca^{2+} with only minor changes in its structure and dynamics. The C-terminal consensus EF-hand, on the other hand, requires more significant changes. The current findings suggest that longer-range motion is also more pronounced in the N-terminal EF-hand. The fact that the N-terminal EF-hand undergoes more limited changes upon Ca^{2+}-binding may be favored by the longer-range motion.

Alternative expansion strategies: spatial expansion

Distinctions between short- and long-range motion depend on one's definition of local. The results presented thus far were obtained by expanding the system of peptide planes in a linear fashion along the protein sequence. Effects are considered local in this scenario if they originate from nearby in the protein sequence. It is possible, however, that proximity in three-dimensional space is a more relevant indicator of what should be considered local. A more appropriate strategy in that case would be to expand the system across the protein's three-dimensional structure. In order to investigate this possibility, a second set of calculations was implemented for each protein as follows. Distances between the backbone N atoms were calculated from a snapshot taken midway through the production phase (at the 2 ns mark). These distances were then used to determine the order in which peptide planes would be introduced into the order parameter calculations. The initial calculation again included only the peptide plane containing the N–H bond vector of interest. Following that initial calculation, the peptide plane whose N atom gave the shortest distance to the N atom of the original plane was added to the system. The expansion continued in this manner by adding planes at an increasing distance from the original plane. The expansion was considered complete when the plane furthest from the original plane had been added to the system. In order to limit the time that would be needed to perform these calculations, the independent variable n was sampled over a restricted set of values. The first 10 planes were introduced into the system one at a time. Following these first 10 (n, S^2) data points, the value of n was sampled more sparsely. The values increased in increments of 2 from $n = 10$ to 20 and then in increments of 5 from $n = 20$ to 75. The resulting (n, S^2) data set was again fit to Eq. 4 and the effective reorientational number was calculated using the best-fit parameters. The results are shown in Fig. 7, which plots n_{eff} as a function of residue number.

In general, the results of the spatial expansion are consistent with those of the sequence expansion. Calbindin again exhibits slightly higher n_{eff} values than does ubiquitin, and the primary sites of longer-range motion coincide with those that were identified above. A closer examination of the data, however, reveals subtle differences between the sequence and spatial expansions. In ubiquitin, for example, the two types of secondary structure respond differently to the change in expansion strategy. In the α-helix, the n_{eff} values are on average 0.52 residues higher in the sequence

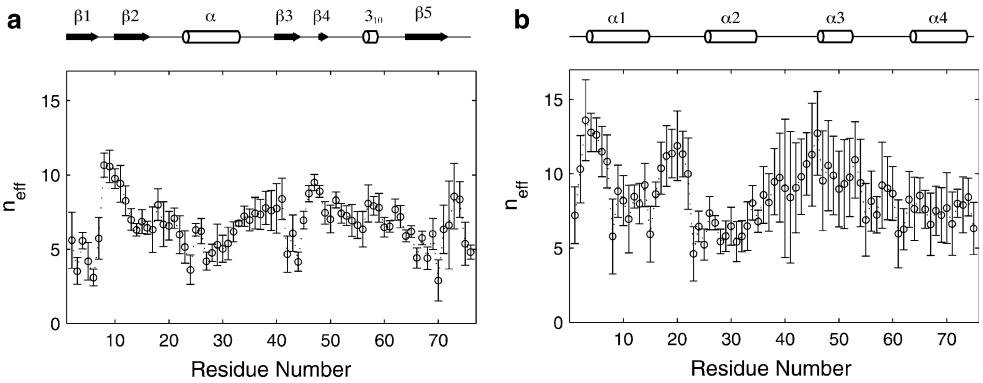


Fig. 7 Plot of n_{eff} as a function of residue number for ubiquitin (**a**) and calbindin (**b**), using an alternative expansion strategy that proceeds in space, rather than along the protein sequence. For a given residue, the *circle* and *error bar* correspond to the average and standard deviation, respectively, over the ten MD simulations

expansion than in the spatial expansion. The sequence expansion, therefore, favors the appearance of longer-range motion in the α-helix. In the β-strands, however, the n_{eff} values are on average 0.66 residues higher in the spatial expansion, suggesting that the spatial expansion favors the appearance of longer-range motion. The divergence in this case likely reflects fundamental differences in the composition of the two types of secondary structure. The α-helix is composed of residues that are in neighboring sequence positions. The sequence expansion, therefore, effectively captures many of the key interactions that give rise to its internal dynamics. Each β-strand, however, is part of a sheet that spans more distant sequence positions. Under these circumstances, the sequence expansion can neglect potentially significant interactions that are more effectively captured by the spatial expansion.

This secondary structural effect, albeit relatively clear in the case of ubiquitin, is not supported by the calbindin data. One would expect, in light of the ubiquitin data, that calbindin's α-helices would exhibit higher n_{eff} values in the sequence expansion. The results indicate, however, that the n_{eff} values for the α-helices are actually 0.55 residues higher (on average) in the *spatial* expansion. This result reflects a much more uniform response in calbindin to the change in expansion strategy. In calbindin, the n_{eff} values are higher in the spatial expansion, not only for the α-helices, but across the entire protein (on average by 0.36 residues). It appears, therefore, that the spatial expansion favors the appearance of longer-range motion more generally in calbindin. It may be the case that the secondary structural effect in ubiquitin is limited to relatively rigid proteins. Apo calbindin may simply be too flexible for such an effect to emerge. In some circumstances, the relative flexibility of the system may even limit one's choice of expansion strategy. In proteins that are significantly more flexible than calbindin, no one structure is likely to serve as a suitable reference for determining the order in which residues are incorporated into the spatial expansion. In that case, one might consider using multiple reference structures and averaging over the results or rely instead upon the sequence expansion.

Alternative expansion strategies: random expansion

One final possibility to consider is a random expansion strategy. This strategy serves as a sort of negative control. Following the initial calculation, which included only the peptide plane of interest, additional planes were randomly incorporated into the system until all 75 planes were represented. The results are presented in Fig. 8. Under the random expansion strategy, the n_{eff} values are consistently lower than what was observed above. In fact, the n_{eff} values converge to quite similar values for the two proteins. The median n_{eff} values are now 5.24 and 5.53 for ubiquitin and calbindin, respectively. Although some variation persists within each protein, the n_{eff} profiles lose many of their

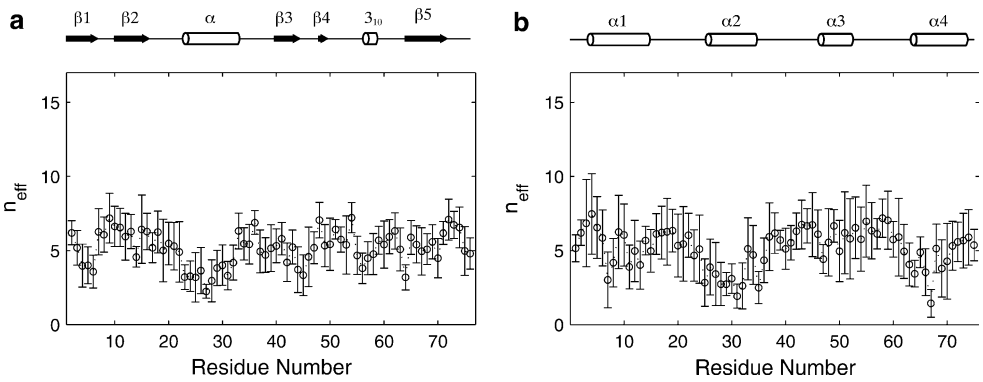


Fig. 8 Plot of n_{eff} as a function of residue number for ubiquitin (**a**) and calbindin (**b**), using a random expansion strategy. The *circle* and *error bar* for a residue correspond to the average and standard deviation, respectively, over the ten MD simulations

characteristic features, masking any evidence of longer-range motion. It appears, therefore, that new insights are likely to emerge only if the expansion starts locally and then incorporates more distant planes into the system according to some meaningful set of rules.

Conclusions

NMR relaxation experiments and MD computer simulations have long been regarded as complementary techniques (Brüschweiler 2003; Case 2002). Computational studies often rely on experimental order parameters as benchmarks in force field development, while experimental studies benefit from the mechanistic level of detail that is provided by simulations. In order for simulations to be useful, however, the information contained therein must be brought to light by the necessary tools of analysis. The present study provides one such tool. When applied to the proteins ubiquitin and calbindin, the method described here reveals that the order parameters are largely the result of short-range contributions. Longer-range motions appear to play a role, particularly in the case of calbindin, but are limited to specific sites, many of which have been identified in previous studies of correlated motion. It is likely that the method described here would also prove useful if applied to a protein with a less well-defined structure. Future applications of this sort would continue to define the limits of separability and would help to characterize protein dynamics across the order–disorder continuum.

Acknowledgments I thank Mark Fischer, Diana Davis, Jamie Titus, and Mark Rance for their many insights. I also thank Tricia Johnson for help in preparing the figures. The Rieveschl Endowment for the Sciences provided funding to purchase the computer used in this study.

References

Akke M, Skelton NJ, Kördel J, Palmer AG, Chazin WJ (1993) Effects of ion binding on the backbone dynamics of calbindin D_{9k} determined by ^{15}N NMR relaxation. Biochemistry 32: 9832–9844

Bertini I, Carrano CJ, Luchinat C, Piccioli M, Poggi L (2002) A ^{15}N NMR mobility study on the dicalcium P43 M calbindin D_{9k} and its mono-La^{3+}-substituted form. Biochemistry 41:5104–5111

Brüschweiler R (1995) Collective protein dynamics and nuclear spin relaxation. J Chem Phys 102:3396–3403

Brüschweiler R (2003) New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. Curr Opin Struct Biol 13:175–183

Case DA (2002) Molecular dynamics and NMR spin relaxation in proteins. Acc Chem Res 35:325–331

Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26: 1668–1688

Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ (2007) Protein NMR spectroscopy: principles and practice. Academic Press, San Diego

Genheden S, Diehl C, Akke M, Ryde U (2010) Starting-condition dependence of order parameters derived from molecular dynamics simulations. J Chem Theory Comput 6:2176–2190

Goodman JL, Pagel MD, Stone MJ (2000) Relationships between protein structure and dynamics from a database of NMR-derived backbone order parameters. J Mol Biol 295:963–978

Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725

Jarymowycz VA, Stone MJ (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. Chem Rev 106:1624–1671

Johnson E, Showalter SA, Brüschweiler R (2008) A multifaceted approach to the interpretation of NMR order parameters: a case study of a dynamic α-helix. J Phys Chem B 112:6203–6210

Kesvatera T, Jönsson B, Telling A, Tõugu V, Vija H, Thulin E, Linse S (2001) Calbindin D_{9k}: a protein optimized for calcium binding at neutral pH. Biochemistry 40:15334–15340

Korzhnev DM, Billeter M, Arseniev AS, Orekhov VY (2001) NMR studies of Brownian tumbling and internal motions in proteins. Prog Nucl Magn Reson Spectrosc 38:197–266

Li DW, Meng D, Brüschweiler R (2009) Short-range coherence of internal protein dynamics revealed by high-precision in silico study. J Am Chem Soc 131:14610–14611

Lienin SF, Bremi T, Brutscher B, Brüschweiler R, Ernst RR (1998) Anisotropic intramolecular backbone dynamics of ubiquitin characterized by NMR relaxation and MD computer simulation. J Am Chem Soc 120:9870–9879

Linse S, Brodin P, Drakenberg T, Thulin E, Sellers P, Elmdén K, Grundström T, Forsén S (1987) Structure-function relationships in EF-hand Ca^{2+}-binding proteins. Protein engineering and biophysical studies of calbindin D_{9k}. Biochemistry 26:6723–6735

Linse S, Johansson C, Brodin P, Grundström T, Drakenberg T, Forsén S (1991) Electrostatic contributions to the binding of Ca^{2+} in calbindin D_{9k}. Biochemistry 30:154–162

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 104:4546–4559

Mäler L, Blankenship J, Rance M, Chazin WJ (2000) Site-site communication in the EF-hand Ca^{2+}-binding protein calbindin D_{9k}. Nature Struct Biol 7:245–250

Malmendal A, Carlström G, Hambraeus C, Drakenberg T, Forsén S, Akke M (1998) Sequence and context dependence of EF-hand loop dynamics. An ^{15}N relaxation study of a calcium-binding site mutant of calbindin D_{9k}. Biochemistry 37:2586–2595

Meirovitch E, Shapiro YE, Polimero A, Freed JH (2006) Protein dynamics from NMR: the slowly relaxing local structure analysis compared with model-free analysis. J Phys Chem A 110:8366–8396

Mori M, Kateb F, Bodenhausen G, Piccioli M, Abergel D (2010) Toward structural dynamics: protein motions viewed by chemical shift modulations and direct detection of C′N multiple-quantum relaxation. J Am Chem Soc 132:3594–3600

Palmer AG (2004) NMR characterization of the dynamics of biomacromolecules. Chem Rev 104:3623–3640

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612

Prompers JJ, Brüschweiler R (2001) Reorientational eigenmode dynamics: a combined MD/NMR relaxation analysis method for

flexible parts in globular proteins. J Am Chem Soc 123: 7305–7313

Prompers JJ, Brüschweiler R (2002) General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. J Am Chem Soc 124:4522–4534

Showalter SA, Brüschweiler R (2007) Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field. J Chem Theory Comput 3:961–975

Skelton NJ, Kördel J, Chazin WJ (1995) Determination of the solution structure of apo calbindin D_{9k} by NMR spectroscopy. J Mol Biol 249:441–462

Theobald DL, Wuttke DS (2006) THESEUS: Maximum likelihood superpositioning and analysis of macromolecular structures. Bioinformatics 22:2171–2172

Theobald DL, Wuttke DS (2008) Accurate structural correlations from maximum likelihood superpositions PLoS Comput Biol 4:e43

Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 Å resolution. J Mol Biol 194:531–544

Vugmeyster L, Raleigh DP, Palmer AG, Vugmeister BE (2003) Beyond the decoupling approximation in the model free approach for the interpretation of NMR relaxation of macromolecules in solution. J Am Chem Soc 125:8400–8404

Wong V, Case DA (2008) Evaluating rotational diffusion from protein MD simulations. J Phys Chem B 112:6013–6024

Wong V, Case DA, Szabo A (2009) Influence of the coupling of interdomain and overall motions on NMR relaxation. Proc Natl Acad Sci USA 106:11016–11021