Dynamics and Entropy of a Calmodulin−Peptide Complex Studied by NMR and Molecular Dynamics†

Ninad V. Prabhu, Andrew L. Lee,‡ A. Joshua Wand, and Kim A. Sharp*

Johnson Research Foundation and Department of Biochemistry and Biophysics, University of Pennsylvania, 3700 Hamilton Walk, Philadelphia, Pennsylvania 19104-6059

Received July 31, 2002

ABSTRACT: All-atom, explicit water molecular dynamics simulations of calcium-loaded calmodulin complexed with a peptide corresponding to the smooth muscle myosin light chain kinase target were carried out at 295 and 346 K. Amide and side chain methyl angular generalized order parameters were calculated and analyzed in the context of the protein's structure and dynamics. The agreement between amide order parameters measured by NMR and those from the simulations was found to be good, especially at the higher temperature, indicating both better convergence for the latter and excellent transferrability of the CHARMM parameters to the higher temperature. Subtle dynamical features such as helix fraying were reproduced. A large range of order parameters for the nine calmodulin methionines was observed in the NMR, and reproduced quite well in the simulations. The major determinant of the methionine order parameter was found to be the proximity to side chains of aromatic residues. An upper bound estimate of the difference in backbone entropy between loop and helical regions was extracted from the order parameters using a model of motion in an effective potential. Although loop regions are more flexible than helical regions, it was found that the entropy loss per residue upon folding was only ∼20% less for loops than for helices. Pairwise correlated motions, which could significantly lower entropy estimates obtained from order parameter analysis alone, were found to be largely absent.

Calmodulin is a protein that regulates the activities of many proteins, including smooth muscle myosin light chain kinase, based on the intracellular calcium levels. Calmodulin also undergoes large conformational changes upon calcium binding and protein binding, and it forms intimate intermolecular contacts with domains of regulated target proteins. Upon formation of a high-affinity complex with a peptide corresponding to the smooth muscle myosin light chain kinase target (smMLCKp),[1] the backbone dynamics of calcium-saturated calmodulin (CaM) are largely unaffected, while the side chain dynamics are significantly perturbed (*1*). The role of protein dynamics in calmodulin function, and in the function of proteins in general, continues to be a subject of intense study and discussion (*2*). High-resolution NMR is a leading source of information about the dynamics of proteins. Using NMR relaxation methods, Lee et al. have measured nearly the entire set of backbone amide and side chain methyl generalized order parameters (O)[2] for CaM complexed with a peptide corresponding to the smooth muscle myosin light chain kinase target (smMLCKp) at seven temperatures ranging from 15 to 73 °C (*1*, *3*, *4*), providing an unprecedented amount of detail on the temperature-dependent dynamics of this protein. Lee and Wand have also published a global analysis of the temperature dependence of this protein's dynamic behavior, as revealed by the NMR order parameter methods (*4*). This analysis shows that the well-known glass-disorder transition in proteins can be explained in terms of the specific temperature dependencies of side chain dynamics as revealed by NMR relaxation measurements. The temperature dependence of the CaM order parameters has also been analyzed using several simple interpretive models, which indicate that changes in experimental generalized order parameters can be robustly related to changes in residual entropy (*3*). Thus, this protein provides a rich and important data set for analyzing and understanding protein dynamics.

NMR-derived order parameters can provide site-specific information about the dynamics of proteins, but a full understanding of dynamics requires some interpretation using a physical model. The generalized order parameter at a particular temperature is a function of the angular time correlation function of a group on the NMR time scale, and the "model free" approach provides a rigorous but general framework for interpreting this angular motion (*5*). A common approach for interpreting the NMR order parameter

† Financial support from NIH Grants GM54105 and PO1-GM48130 (K.A.S.), DK 39806 (A.J.W.), and GM18114 (A.L.L.) is gratefully acknowledged.

* To whom correspondence should be addressed. Telephone: (215) 573-3506. Fax: (215) 898-4217. E-mail: sharpk@mail.med.upenn.edu.

‡ Present address: Division of Medicinal Chemistry and Natural Products, School of Pharmacy, CB#7360, Beard Hall, University of North Carolina, Chapel Hill, NC 27599-7360.

[1] Abbreviations: CaM, calcium-saturated calmodulin; smMLCKp, peptide corresponding to the calmodulin domain of smooth muscle mysoin light chain kinase; O^2, squared generalized order parameter; MD, molecular dynamics; S, entropy.

[2] In this paper, we will use the symbols O and O^2 for the generalized order parameter and its square, respectively, instead of the usual symbols S and S^2, to avoid confusion with the use of S for entropy.

10.1021/bi026544q CCC: $25.00
Published on Web 12/10/2002

is to consider the group moving or "diffusing" in a time invariant "mean" or "effective" potential well of some general shape (*4*, *6*, *7*). The resulting value of O depends on the depth, width, and, in general, the shape of this well. For a given model, a higher value of O means less angular range of motion, but many specific angular distributions (and by implication many potential well shapes) are consistent with a given value of O at a given temperature. This makes it difficult to extract specific information about the shape of the potential surface or details of side chain dynamics from a set of O values measured at one temperature. Molecular dynamics (MD) simulations can be used to calculate order parameters which can be compared to experiment, and at the same time provide a more detailed description of the dynamics. Thus, MD simulations have become a useful adjunct to the interpretation of NMR order parameters (*7*−*16*). Since the extent of protein motion is intimately related to its configurational entropy, there has also been great interest in extracting entropies, or changes in entropy from order parameter data either directly from experiment or in combination with MD simulations (*6*, *7*, *16*−*19*). To relate order parameters and entropy quantitatively, it is necessary to know the angular probability distribution of the groups. This can be estimated by assuming particular forms for the effective potential (*6*, *7*, *17*). Alternatively, molecular dynamics simulations, since they contain the full information about the dynamics on an atomic level, can in principle be used to relate O and S quantitatively without assuming a simplified potential, although the accurate evaluation of the necessary probability distributions is a very difficult computation. An additional problem in relating order parameters to protein configurational entropy is the one of correlations. Even if the exact probability distribution of a particular group (or equivalently, the effective potential well in which it is moving) were known and this exactly reproduced the experimentally measured order parameter over the entire temperature range, this would provide only an upper bound on the entropy. Order parameters provide information only about the motion of individual groups, not about correlations between two or more motions, which would decrease the entropy. MD simulations have become increasingly sophisticated and accurate, and in principle, the full atomic level information about the dynamics provided by MD can also be used to examine the extent of correlations, and their effect on the calculated entropy, although how accurately MD can reproduce the configurational entropy of a protein is currently unknown.

The goal of this study was to use molecular dynamics simulations to provide a more detailed analysis of the temperature-dependent dynamics of calmodulin: (1) to compare calculated and experimental order parameters, particularly side chain order parameters, (2) to interpret side chain order parameter variations in terms of protein structure and dynamics, and (3) to relate changes in the order parameter to changes in the conformational entropy of the protein.

MATERIALS AND METHODS

Molecular dynamics simulations were carried out with the CHARMM25 package, using the CHARMM22 all-atom parameter set and the TIP3P water potential (*20*−*22*). Coordinates of the CaM−smLCKp complex were obtained from entry 1CDL of the Protein Data Bank (*23*). The asymmetric unit contains four replicas. Replica B was chosen since it is the most complete, missing only two lysine side chains, both of which are far from the peptide. The two missing side chains were built and hydrogens added using CHARMM, and the complex was minimized for 100 steps of steepest descent, followed by 3000 steps of adopted basis Newton−Raphason. The complex was then placed in a cubic box of pre-equilibrated TIP3P water, 63 Å on a side. Overlapping waters were removed, and the entire system was subjected to 500 steps of steepest descent minimization to remove residual strain from water−protein overlap. Molecular dynamics simulations were run at either 295 or 346 K using the following setup and parameters: nonbonded cutoff of 13 Å with vshift and fshift options, nonbonded lists updated every 10 steps, constant temperature, constant pressure of 1 atm, periodic orthorhombic boundary conditions, minimum image convention, Nose-Hoover thermostat and barostat, and a time step of 1 fs. All bonds to H atoms were constrained with the shake algorithm; in addition, the H−O−H angle of the water was constrained. A 50 ps equilibration period was followed by 12 × 100 ps data collection periods. Coordinates were saved every 0.1 ps for analysis. Equilibration of the system was monitored by following the protein radius of gyration mean energy, and by the order parameters themselves (see below).

NMR order parameters were calculated from the MD coordinates as follows. The snapshots were overlaid on each other using a standard mass-weighted rigid body alignment. For each snapshot, the unit vector (μ) of each amide group in the NH direction was calculated. For each methyl group, the unit vector along the carbon centroid of the three hydrogens was calculated; i.e., we assumed free methyl rotation, which is consistent with both experimental and simulation results (refs *9* and *24* and this work). The "long" time average O^2 of each site was then calculated from the x, y, and z Cartesian components of the unit vector (*9*):

$$O^2 = {}^3/_2(\langle x^2\rangle^2 + \langle y^2\rangle^2 + \langle z^2\rangle^2 + 2\langle xy\rangle^2 + 2\langle yz\rangle^2 + 2\langle zx\rangle^2) - {}^1/_2 \quad (1)$$

where the broken brackets ($\langle\rangle$) denote the average over the trajectory. In addition, for selected amide and methyl groups the angular time correlation function

$$C(t) = \langle P_2[\mu(\tau)\,\mu(\tau + t)]\rangle \quad (2)$$

was calculated from the time series of unit vectors $\mu(t)$ using the method of Fourier transforms, where $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial. Here the broken brackets ($\langle\rangle$) denote the average over all time origins in the trajectory. The time course of $C(t)$ was used to judge the length of simulation needed for convergence of the calculated values of O^2. Uncertainties in the precision of O^2 were estimated from the standard deviation among the 10 batches of snapshots.

Correlations between atomic motions were calculated from the overlaid snapshots by calculating the mean coordinates of each atom, and the mean squared coordinate fluctuation covariance matrix (see ref *25*). The diagonals of this matrix are the mean squared fluctuation in atomic coordinates, $\langle \Delta x_i^2\rangle$, where Δx_i is the displacement of the ith coordinate

from its mean value. The off-diagonal terms, $\langle \Delta x_i \Delta x_j \rangle$, were normalized to give the coordinate fluctation correlation coefficients:

$$\mathbf{c}_{ij} = \frac{\langle \Delta x_i \Delta x_j \rangle}{\sqrt{\langle \Delta x_i^2 \rangle \langle \Delta x_j^2 \rangle}} \quad (3)$$

Solvent accessible surface areas of atoms and residues were calculated using the program SURFCV (*26*). Side chain packing fractions and accessible volumes were calculated using the voronoi polyhedron method (*27*).

We examined the relationship between the order parameter and entropy for motion of a group in several simple azimuthally symmetric potential well models where the restoring potential $U(\theta)$ depends only on the power of the angle θ with respect to the symmetry axis [$U(\theta) = K\theta^n$ for $n = 2$, 4, and 6] and K is the force constant. The order parameter was calculated to high precision by numerical integration of the angular partition function

$$Z = \int_0^{\pi} \mathrm{d}\theta \sin\theta \mathrm{e}^{-U(\theta)/kT} \quad (4)$$

followed by evaluation of O^2 from the mean squared cosine angular correlation

$$O^2 = {}^3/_2\langle \cos^2\theta \rangle - {}^1/_2 = {}^3/_2[\int_0^{\pi} \cos^2\theta\, p(\theta)\, \mathrm{d}\theta]^2 + {}^3/_4[\int_0^{\pi} \sin^2\theta\, p(\theta)\, \mathrm{d}\theta]^2 - {}^1/_2 \quad (5)$$

where $p(\theta)$ (=$\sin\theta \mathrm{e}^{-U/kT}/Z$) is the angular probability function obtained from the partition function. The entropy was evaluated from $p(\theta)$ using

$$S/k = -\int_0^{\pi} \mathrm{d}\theta\, P(\theta) \ln[P(\theta)] - S_{\mathrm{ref}}/k \quad (6)$$

where S is excess relative to the entropy of unconstrained motion and $S_{\mathrm{ref}} = k \ln(4\pi)$. For each potential, the dependence of S on O was evaluated by varying K to generate order parameters spanning the experimentally observed range (0.2−0.9).

RESULTS

Effect of Sequence Position and Temperature. Amide NH order parameters were calculated from the MD simulations at 22 and 73 °C and compared to experiment. Panels A and B of Figure 1 show the calculated and measured amide NH generalized order parameter as a function of sequence position. The error bars indicate the rms variation in calculated values between 0.1 ns windows. The location of helices is indicated below the plots. The general pattern of amide order parameters at 22 °C is typical of that seen previously in proteins, with higher values within helices decreasing toward the ends reminiscent of helix "fraying" and lower values in loop regions. The absolute value and sequence profile of simulated order parameters closely follow the experiment. The agreement is better at 73 °C than at 22 °C, primarily because of the better agreement with the lower O^2 values in the loop region. Example angular time correlation functions for high O^2 amide, low O^2 amide, and methyl groups are shown in Figure 2. The correlation functions of sites having high O^2 values converge rapidly within the simulation time. The correlation functions of sites having low O^2 values converge more slowly, but are still fully converged during the simulation at the high temperature. At the lower temperature, there are some lower-frequency oscillations ~0.1 in magnitude superimposed on a rapidly converging ca. 0.7 contribution, indicating the contribution from longer time scale motions of the loops. This increases the uncertainty in the calculated order parameters, which is echoed by the larger uncertainty limits in Figure 1. Thus, one reason for the better agreement with experiment at 73 °C is superior convergence of the loop amide O^2 at the higher temperature. The methyl convergence profile is generally good, but shows some longer time scale drift and oscillation, also indicating contributions from longer time scale motions, which introduces a corresponding uncertainty in the values.

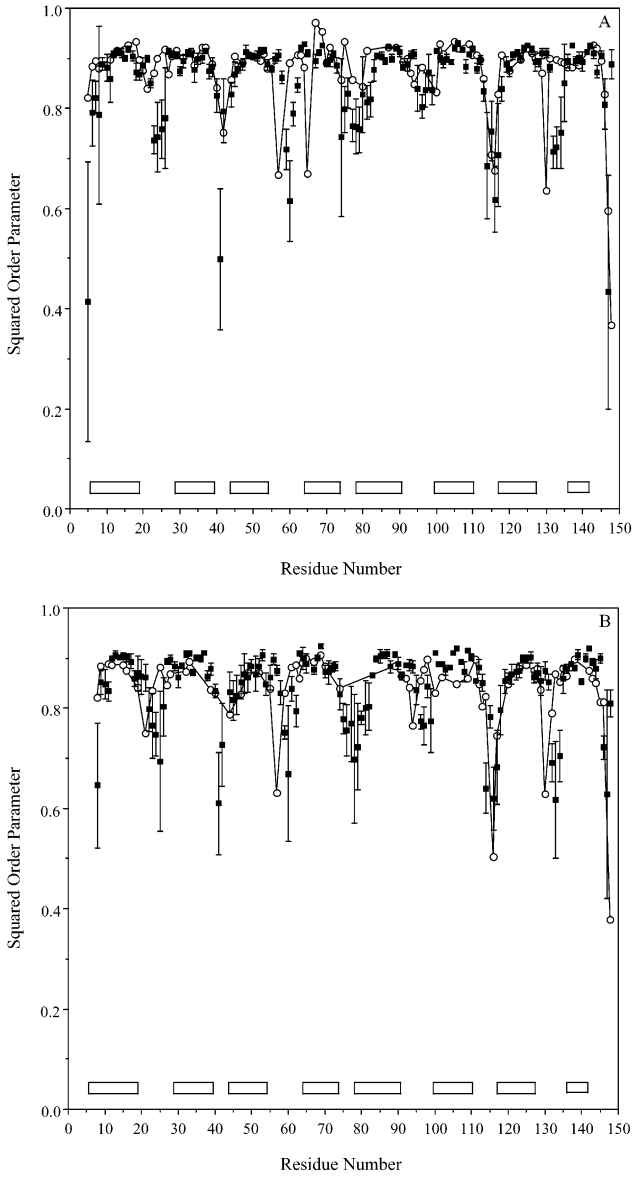

FIGURE 1: Calculated (■) and measured (○) O^2 as a function of sequence. Helices indicated as bars below the plot: (A) 22 and (B) 73 °C.

The effect of temperature and secondary structure on the calculated average order parameter values is summarized in Table 1. The amide order parameters decrease significantly between 22 and 73 °C but largely follow the same trends seen in the lower-temperature simulations. The general

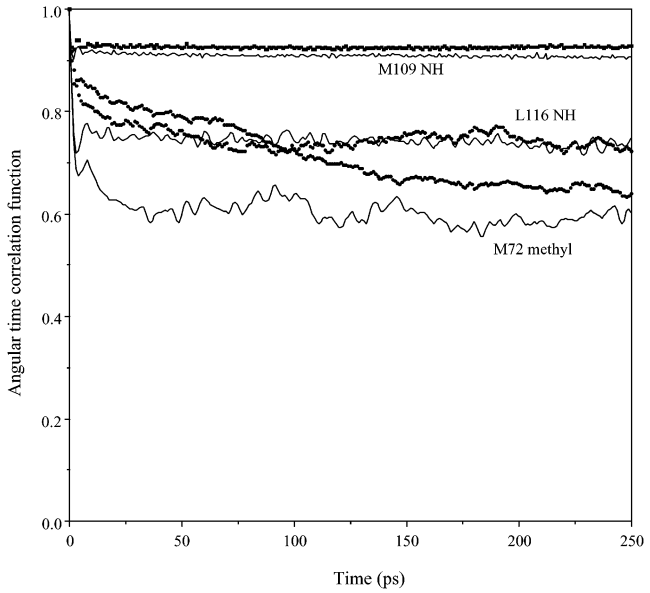

FIGURE 2: Angular time correlation function for a high-O^2 amide (M109 NH), a low-O^2 amide (Leu116 NH), and a midrange-O^2 methyl (M72) at 22 (■) and 73 °C (—).

Table 1: Average Squared Order Parameter Values from Simulations[a]

	22 °C	73 °C
all amides	0.860 ± 0.1 (0.872)	0.839 ± 0.1 (0.837)
helix amides	0.882 ± 0.1	0.856 ± 0.1
loop amides	0.815 ± 0.1	0.805 ± 0.1
all methyls	0.781 ± 0.2 (0.594)	0.733 ± 0.2 (0.481)
methionine methyls	0.475 ± 0.2 (0.472)	0.358 ± 0.2 (0.332)

[a] Experimental data in parentheses.

agreement between the calculated and experimental order parameters is illustrated in Figure 3. For the amides (Figure 3A), there is a clustering of most values around the diagonal line at high order parameter values, indicating good agreement. However, the bulk of the points at 22 °C lie somewhat below the line, indicating a systematic underestimation of order parameters in the simulation, while the systematic deviation at 73 °C is quite small. Overall, there is a small but significant underestimation of the effect of temperature on amide order parameters (Figure 3C). Underestimation of protein amide order parameters has been seen previously in MD simulations, and attributed to ultrafast librational motion of the NH bonds with respect to their peptide plane (*11*). One possibility, systematic deviations in the length of the N−H bond used in extraction of the experimental values (which depends on the sixth power of this length), cannot be ruled out. An MD analysis of the contribution of NH bond stretching to calculated amide order parameters concludes that this is a very small effect (*8*). Another possibility is that the potential functions used in MD are slightly too "stiff" as a function of temperature. For the methyls (Figure 3B), the picture is somewhat different. Here there is a systematic overestimation of the order parameters. Overestimation can result from an insufficiently long simulation, in the region where the time correlation function is still falling (viz. Figure 2 below 50 ps). However, this cannot explain all of the deviation, for two reasons. First, the time correlation function plots indicate that the simulations are in fact fairly well converged. Second, the overestimation is about the same at

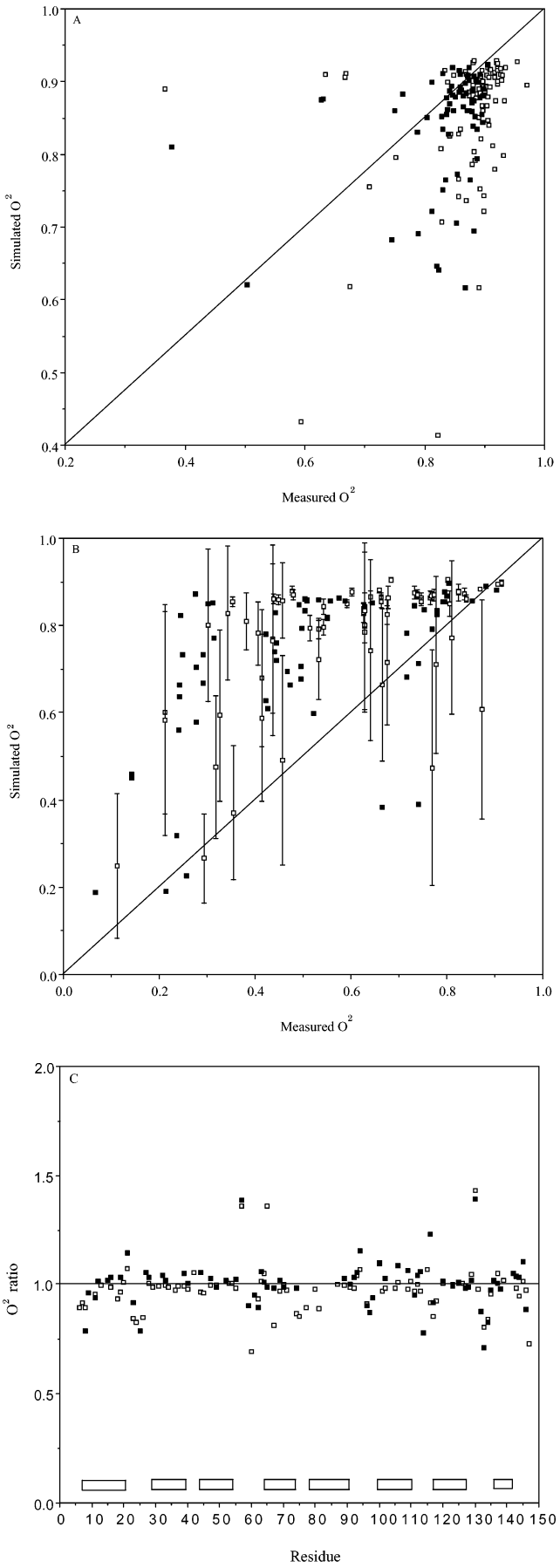

FIGURE 3: Calculated vs experimental O^2: (A) amide, (B) methyl, and (C) ratio of calculated to experimental amide order parameters at (□) 22 and (■) 73 °C.

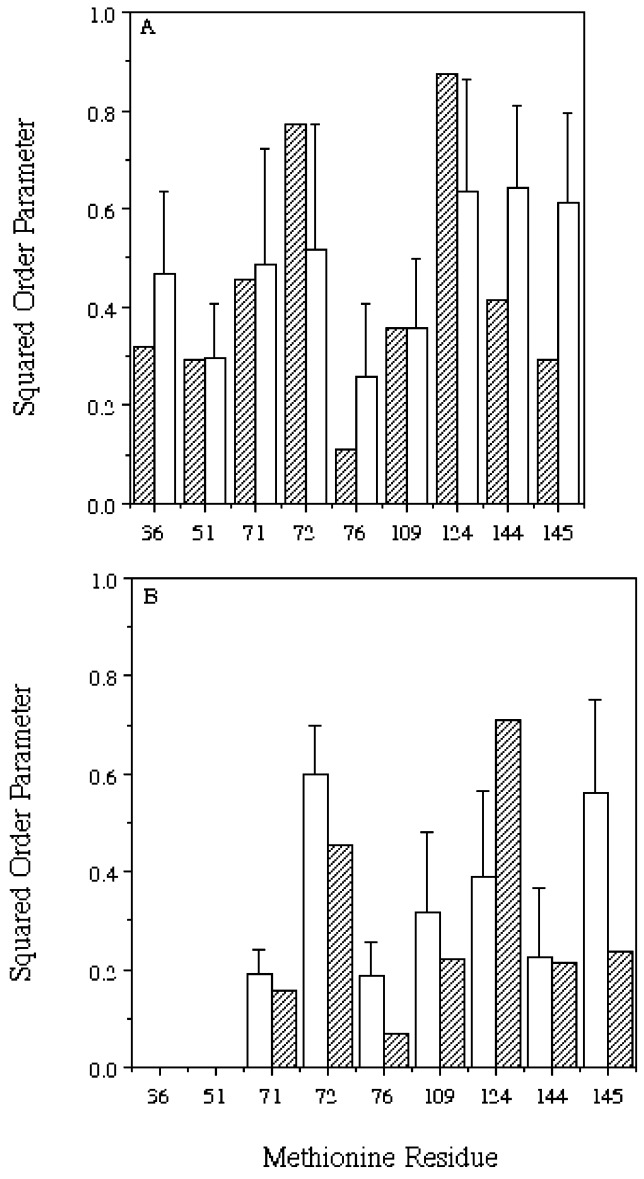

FIGURE 4: Calculated (unshaded bars) vs experimental (shaded) methionine O^2 at (A) 22 and (B) 73 °C.

the high and low temperatures, whereas convergence effects are anticipated to be more marked at 22 °C. The deviation appears to have some other source. Upon comparison of just methionine methyl order parameters, which because of their length should report on the side chain dynamics least coupled to backbone motions and steric effects, the agreement with experiment is very good at both temperatures (Table 1), with just a slight underestimate of the temperature dependence.

Side Chain Dynamics. Calmodulin is unusually rich in methionines, containing nine in fewer than 150 residues. Moreover, they show significantly different order parameter values as a function of position (Figure 4). Methionines also contain the most distal methyl groups from the backbone. In this respect, they are expected to contain the least amount of backbone motion information, i.e., represent the "purest" side chain dynamics (*1*). For these reasons, we focused our analysis of side chain order parameters on the methionines. The pattern of observed high and low O^2 values for methionine is reproduced remarkably well by the simulations

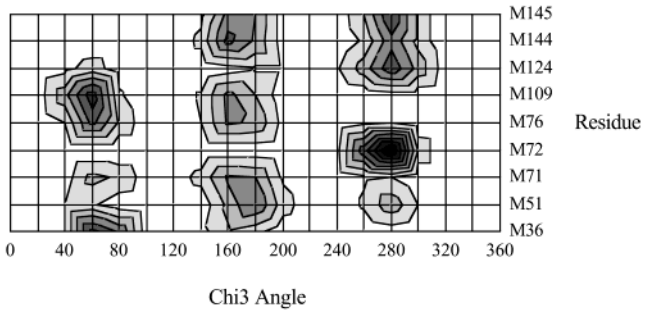

FIGURE 5: Rotamer populations for the methionine CG−SD dihedral (χ_3) at 73 °C.

at both 22 and 73 °C. To interpret the marked differences in order between these methionines in structural terms, we analyzed in detail the methionine motions and environment using the MD trajectories.

Analysis of the rotamer populations of the different methionines revealed a strong correlation between the χ_3 rotamer and the two most ordered Met residues, M72 and M124. These two Met residues are distinct in exhibiting both a small population of the normally most favorable trans rotamer and a large population of the gauche− rotamer (Figure 5), indicating a sterically restrained environment for these two residues. Beyond this was no observable correlation of O^2 with the χ_1 and χ_2 angles for any of the methionines. We also examined the mean time for transitions between rotamers, which ranged from 71 ps at 22 °C to 32 ps at 73 °C, but there was no clear correlation between the transition time and order parameter of individual Met residues, which indicates that the resulting order parameter is determined as much by motions of the terminal methyl within a given $\chi_1,\chi_2\chi_3$ rotamer class as transitions between different rotamers (*3*). Interpretation of side chain order parameters in terms of motion in some kind of constraining potential well is potentially complicated by the occurrence of rotamer transitions, particularly for long side chains such as Met. This is partly because rotamer transitions can be of a longer time scale than motions of the side chain within a single "rotamer" cage, and partly because the motion must now be interpreted in terms of motions in a multiwell potential, rather than a single-well potential (*3*). On the other hand, the terminal methyl of Met lies three side chain dihedrals away from the backbone and the angular range for each dihedral within a given rotamer is close to 120°. Thus, the range of motions available to this methyl within a single $\chi_1,\chi_2\chi_3$ rotamer class is considerable. Thus, in the extraction of entropy changes from the order parameters, it is necessary to consider both single-well and multiwell rotamer-like potentials (see ref *3*).

To interpret the variation in Met order parameters in structural terms, we examined various obvious features, including the mean packing fraction of the methionine side chains, their exposed surface area, the volume accessible to the terminal SCH_3 group, and the number of solvent molecules contacting the side chain. Met76, which has the lowest order parameter, has the highest accessible volume, exposed surface area, and number of contacting solvent molecules. However, none of these structural features correlate with the other methionine order parameter values (Figure 6A). We also classified every atom within different distances of the SCH_3 group of each methionine, and computed the average coordination number for each type of

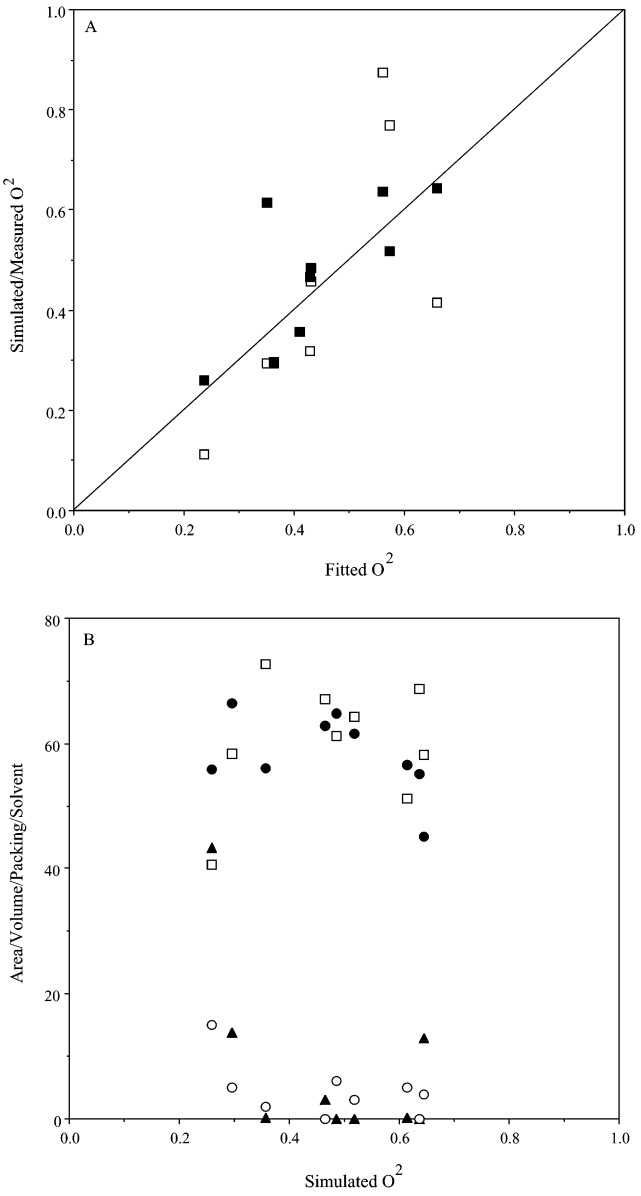

FIGURE 6: (A) Plot of calculated (■) or measured (□) methionine O^2 as a function of that predicted from structural analysis using a function of average neighbor environment (abcissa), $O^2 = 0.474 + 0.05$(mean number of waters) $+ 0.037$(mean number of aromatic residue atoms), where 5% of the variance is explained by the second term and 86% by the third term. (B) Plot of the number of contacting water molecules (○), available side chain volume, in cubic angstroms (●), % packing fraction of the SD and CE groups (□), and solvent accessible area of the SD and CE groups, in square angstroms (▲), vs the calculated methionine O^2 values from MD simulations.

atom. Atoms were first classified as either solvent or protein, and the latter were then subclassified according to amino acid type. Using a multivariate correlation analysis (Origin Software, MicroCal Corp., Northampton, MA), we found that a single factor could account for ~86% of the variance in the simulated order parameter, that factor being the mean number of atoms from side chains of aromatic residues (Phe, Trp, His, and Tyr) within 4 Å (roughly van der Waals contact) of the SCH_3 group. The number of solvent molecules within 4 Å accounts for ~5% of the remaining variance. Figure 6B shows a plot of the simulated and measured methionine O^2 values versus that calculated from the best fit to a two-parameter model using just the mean number of

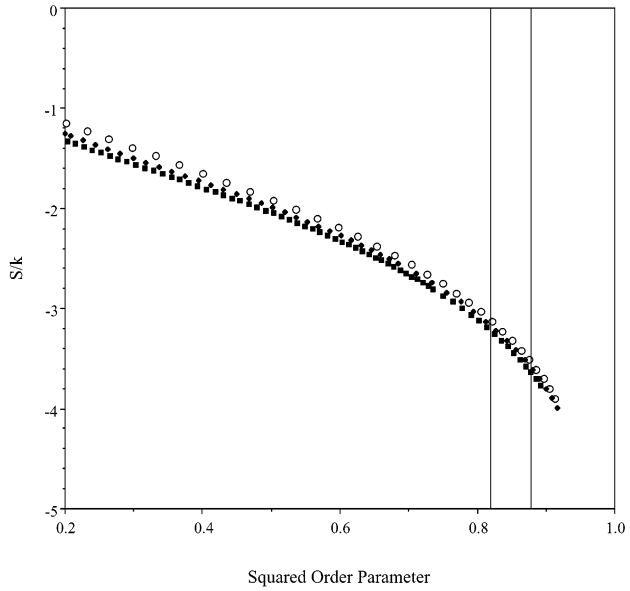

FIGURE 7: Entropy (S/k) vs O^2 for potential well models: (○) harmonic, (◆) quartic, and (■) sixth power. Vertical lines indicate, in increasing order, mean loop and helix amide O^2 values at 22 °C from Table 1.

"neighboring" aromatic side chain and solvent molecules. For the simulated order parameters, with the exception of one outlier, M145, which is very close to the C-terminus, and has an order parameter lower than predicted from its neighbors because of the C-terminal fraying effect, the values essentially lie on the line ($R^2 = 0.93$ excluding M145). The agreement with the experimental values is also remarkable, although not quite as good since there is not perfect agreement between simulated and experimental O^2 values for the methionines.

Configurational Entropy and Correlations. An important use for NMR order parameter measurements is to extract quantitative information about protein dynamics and conformational fluctuations. There have been several studies that have investigated the relationship between O^2 and protein conformational entropy (*3*, *6*, *7*, *17*). The diffusion in a cone model (corresponding to motion in an infinite square well potential) is convenient for analyzing protein motions (*16*) because there is an analytical relationship between entropy and O^2 (*7*). Unfortunately, this model is fundamentally limited in application because it predicts no temperature dependence for O^2, which is in contrast to experiment. We examined the relationship between O^2 and S for several U-shaped potentials, including the harmonic potential, and stiffer potentials which seemed to be appropriate for fairly tightly constrained motions, for example, backbone amide groups in folded proteins (ref *3* and Figure 7). The plots are remarkably similar, which suggests that changes in entropy can be robustly obtained from changes in the order parameter for this class of motions. Proceeding from this assumption, we used the mean values for the loop and helical regions from Table 1, along with the plots of Figure 7 to estimate differences in entropy (the mean loop and helical values are indicated by the lower and upper vertical lines on the figure, respectively). The loops and helices are regions of more and less average flexibility, respectively, and the estimated average difference in backbone entropy between the two regions from any of the models is ca. 1.2 cal mol^{-1} K^{-1}. If

Table 2: Measured Consensus Helix Amide Squared Order Parameters

position in helix[a]	22 °C	73 °C
average over all helices	0.882 ± 0.1	0.856 ± 0.1
1	0.808 ± 0.3	0.635 ± 0.4
2	0.881 ± 0.02	0.838 ± 0.02
3	0.921 ± 0.03	0.866 ± 0.04
4	0.913 ± 0.01	0.874 ± 0.02
N-3	0.914 ± 0.01	nd[b]
N-2	0.905 ± 0.02	0.878 ± 0.01
N-1	0.902 ± 0.02	0.854 ± 0.02
N	0.877 ± 0.02	0.835 ± 0.02

[a] The numbers 1−4 denote the position with respect to the N-terminus of the helix, and N−N-3 denote the position with respect to the C-terminus of the helix. [b] Not determined.

this is assumed to represent the backbone entropy difference per residue, it can be compared to a consensus estimate of 7 cal mol^{-1} K^{-1} per residue for the helix to coil or folding transition (*28*, *29*). The loop regions exhibit an only ∼20% smaller entropy change upon folding than in the more "structured" helical regions. Thus, although the loops are significantly more flexible than the helical regions, they are significantly less flexible than in the unfolded state.

We also used the relationships between O^2 and S to analyze the helix fraying effect. Table 2 contains the amide order parameter data averaged over all equivalent helix positions for the first four and last four positions in the helix. The consensus values clearly show a small but significant fraying effect extending two or three residues into the helix, somewhat larger at the N-terminal end. Using one of the U-shaped single-well potentials in Figure 7 (the harmonic, quartic, and sixth power give almost identical results), we can estimate the increment in entropy at 22 °C going from the terminal residue to the penultimate residue to be ∼0.35 and ∼0.1 cal mol^{-1} K^{-1} at the N- and C-termini, respectively. The magnitude of these effects is rather small compared to the entropy change per residue for a complete helix to coil transition of 7 cal/mol per residue.

A potentially serious limitation for using order parameters to extract conformational entropies is that NMR relaxation techniques currently can report only on the motion of a particular group averaged over the effects of all its neighbors; i.e., they provide no information about correlations between two or more motions. Thus, any model that uses just O^2 to obtain S likely provides the upper limit to the entropy for that model. MD simulations can potentially provide estimates of the extent of correlated motions. Earlier MD studies of amide order parameters found that conformational motions are largely uncorrelated with rotational motions and that different modes of amide motion may be deconvoluted (*8*, *30*, *31*), but correlations between side chains and the side chain backbone have not been studied extensively. Here we use the MD simulations to examine the correlation in motions between side chains, within the backbone, and between the backbone and side chain, in two ways. First, the angular time cross correlation function between various groups was calculated. This function is a generalization of the angular self-correlation function (used to obtain O^2) to the relative motion of two groups (*32*). Figure 8 shows typical plots for methionine−leucine correlations. Flat curves are observed for the cross correlation of the Met36 methyl with nearby (within 4 Å) methyl groups from Leu32 (CD and CG) and

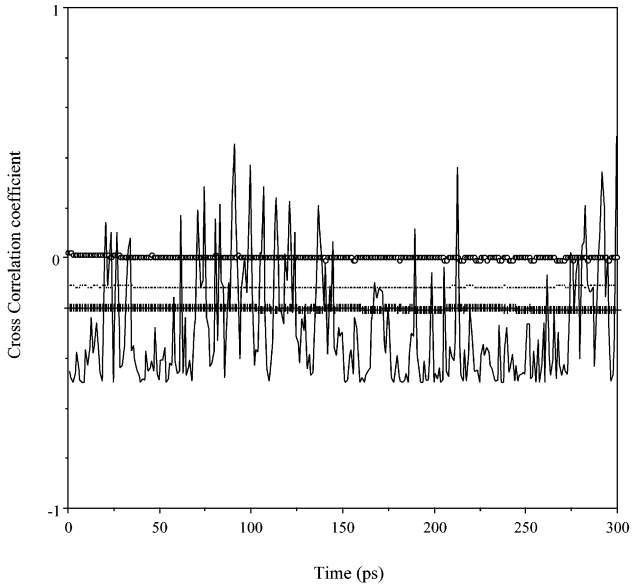

FIGURE 8: Cross correlation function, calculated for the methyl group of Met36 with the methyl (CD) of Leu32 (○), the methyl (CG) of Leu32 (+), or the methyl (CD) of Leu44 (···). Time series of $C(0)$ from the cross correlation function (−).

a more distant methyl group of Leu48. The almost constant value of the cross correlation curves is due to the orthogonality of the periodicity of the spins of the methyl vectors (*32*), indicating that the motions of these pairs of vectors are uncorrelated. The values of this function in these cases are determined by the average relative orientations of the two vectors. An alternative scenario, which would also result in flat correlation curves, is where the motions of two vectors are perfectly correlated. To eliminate this (highly unlikely) possibility, we inspected the $C(0)$ time, which reflects the instantaneous relative orientation of the two vectors along the trajectory. The $C(0)$ for the Met36−Leu32 (CD) correlation is typical, and shown in Figure 8. Indeed, it is seen to fluctuate considerably around the average value, thus eliminating the explanation of perfectly synchronized spins for the two vectors. On the other hand, nonconstant values for the cross correlation functions were observed for trivial cases where correlated motions are expected, e.g., two methyl groups on the same residue (not shown). In general, however, there were no observed significant correlations between the various methyl groups except in the cases where they lie on the same residue.

A second way to examine the correlations is to calculate the coordinate covariance fluctuation matrix $\mathbf{c}_{ij}$ (the elements of which are given by eq 3) for all 246 order parameter groups (amide N and the methyl). The coordinate covariance matrices for both the 22 and 73 °C simulations were computed over all 100 ps trajectory segments. A variety of filters were then applied to the more than 30 000 correlation coefficients to isolate ones of significant interest. First, those between groups whose average separation was >8 Å, and for which $|C_{ij}| < 0.5$, were eliminated, thus retaining only strong correlations between groups that are close enough to interact significantly. All correlations that are expected to occur because of the chain connectivity and peptide geometry, such as those between two methyls of the same side chain, and methyls and amides of the same residue, were then filtered out. Amide−amide correlations between i and

Table 3: Persistent Pairwise Correlations

group 1	group 2	mean distance (Å)	correlation coefficient
Ile5 CG2	Ala6 CB	5.845	0.601
Thr22 CG2	Thr58 CG2	4.661	0.507
Thr24 CG2	Thr58 CG2	4.072	0.528
Thr30 CG2	Val31 CG1	5.235	0.568
Ile81 CG2	Ala84 CB	5.677	0.514
Val117 CG1	Ile121 CD	4.174	0.563
Thr150 CG2	Ala153 CB	5.794	0.504
Ala153 CB	Val154 CG2	5.103	0.567

$i + 1$ residues, which reflect primarily the effects of backbone connectivity (i.e., sequence), were then eliminated. Correlations between neighboring Ala−Ala methyls were discarded for the same reason, since the β-carbon is rigidly connected to the backbone. After the filtering, there were typically fewer than a dozen pairs (<0.04% of the total) left in each 100 ps time frame, which represented putatively significant correlations in motion due to tertiary structure and/or packing. Many of these pairs did not appear in more than one 100 ps time slice or at both 22 and 73 °C, indicating that these correlations resulted from transient motions. These pairs were also eliminated. The ones that remained (Table 3) are all between two side chains, and typically involved the interaction between two β-branched side chains packed closely together, e.g., T26 and T62. In general, β-branched chains appear to be over-represented in this admittedly small sample, indicating that the reduction in side chain motion or rotamer restriction due to backbone steric interactions (*33*) may be a source of significant pairwise correlation in motions. In addition, methyl motions may be reduced by hydroxyl H-bonding with the main chain (*24*), producing correlations. Aside from this β-branch effect, and the immediate neighbor effects of covalent linking, it was concluded that highly correlated pairwise motions due to tertiary interactions are rather rare. Significant three-body and higher-order correlation effects on the entropy cannot be ruled out. However, the number of multibody correlation functions that need to be examined, and the amount of sampling needed, increases exponentially, requiring prohibitively expensive simulation.

DISCUSSION

The extensive NMR relaxation data available for backbone and side chain dynamics in the calmodulin−sMLCKp complex provide a rich and challenging set of data for interpretation by molecular dynamics simulations. We carried out extensive molecular dynamics simulations at lower and upper extremes of the experimental temperature range (22 and 73 °C). In general, the agreement between simulated and experimental amide order parameters is good, particularly at the higher temperature. The latter is particularly encouraging since the CHARMM protein potential function and the TIP3P water potential function are being used nearly 50 °C above the temperature for which they were parametrized, suggesting that they have better transferrability than might be expected. The sequence profile of amide order parameters follows the commonly observed pattern of highs in secondary structures, and lows in loop regions, but in addition, the simulations capture more subtle features of the protein dynamics such as the fraying effect toward the ends of α-helical regions. Overall, the temperature dependence of the amide order parameters is of the right magnitude, but underestimates the experimental variation. On the other hand, the methyl order parameters provide a much more detailed and varied picture of the protein dynamics. The simulations tend to overestimate the methyl order parameters, and underestimate their temperature dependence. However, for methionines, which perhaps represent the purest side chain dynamics, there is good agreement with the magnitude and temperature dependence of the order parameters, suggesting that the general overestimate of the shorter side chain methyl order parameters may require longer simulations for accurate modeling. Further analysis of methyl order parameters is presented elsewhere (*3*). The methionine terminal methyl groups show a surprisingly large range of order parameters, despite all being buried in the hydrophobic protein core. There is also no correlation with the depth of burial, or other simple structural parameters such as packing and surface area exposure. This indicates that the core dynamics are rather heterogeneous. The simulations reproduce the pattern of high and low Met order parameters quite well, and this in addition to the atomic level detail provided by the simulations enabled us to discover a possible major determinant of this dynamic heterogeneity, the presence of aromatic residues. These internally rigid side chains apparently create regions of less dynamic motion in their vicinity. We are currently investigating the generality of this effect by examing whether it is seen in other proteins, and for other side chains such as Ile and Leu. In this light, it is interesting to note that cytochrome c_2 (*34*) and flavodoxin (*35*), which contain large inflexible prosthetic groups, are unusually rigid.

As in previous studies, we have used the order parameter data to relate dynamics to thermodynamic properties, primarily entropy changes. Obtaining *absolute* entropy values from order parameters is problematic unless the exact effective potential in which each group is moving is known. However, by considering a range of similar potential well types, we suggest that reasonable estimates of the *difference* in backbone amide entropy between two states can be obtained, provided the motions in the two states are occurring in a fairly constrained general U-shaped potential well. An example would be amide residues in different parts of a folded protein.

Even with the caveat of similar types of motions, the estimates derived for ΔS are upper bounds as they assume the motion of each group for which an order parameter is measured is uncorrelated with the others. To address this, we used two different types of correlation analysis, from which we conclude that significant pairwise correlations are introduced primarily by chain connectivity (primary structure) effects. These, by their nature, are likely to persist in both folded and unfolded states, and between different folded conformations, and thus, they may not materially devalue the assumption of uncorrelated motions used in relating order parameter changes to entropy differences between such states. Less "trivial" pairwise correlations, i.e., those due to tertiary interactions between groups far apart in sequence, are rather rare probably due to the complex dynamics in the densely packed folded state. We are currently examining ways to quantitate the importance of higher-order correlations for entropy estimates obtained from order parameters (*3*).

REFERENCES

1. Lee, A. L., Kinnear, S. A., and Wand, A. J. (2000) *Nat. Struct. Biol. 7*, 72−77.
2. Wand, A. J. (2001) *Nat. Struct. Biol. 8*, 926−931.
3. Lee, A. L., Sharp, K. A., Kranz, J. K., Song, X., and Wand, A. J. (2002) *Biochemistry* (submitted for publication).
4. Lee, A. L., and Wand, A. J. (2001) *Nature 411*, 501−504.
5. Lipari, G. L., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.
6. Li, Z., Raychaudhuri, S., and Wand, A. J. (1996) *Protein Sci. 5*, 2647−2650.
7. Yang, D., and Kay, L. (1996) *J. Mol. Biol. 263*, 368−382.
8. Buck, M., and Karplus, M. (1999) *J. Am. Chem. Soc. 121*, 9645−9658.
9. Chatfield, D., Szabo, A., and Brooks, B. R. (1998) *J. Am. Chem. Soc. 120*, 5301−5311.
10. Levy, R. M., Dobson, C. M., and Karplus, M. (1982) *Biophys. J. 39*, 107−113.
11. Pfeiffer, S., Fushman, D., and Cowburn, D. (2001) *J. Am. Chem. Soc. 123*, 3021−3036.
12. Philippopoulos, M., Mandel, A. M., Palmer, A. G., and Lim, C. (1997) *Proteins 28*, 481−493.
13. Smith, P., van Schaik, R., Szyperski, T., Wuthrich, K., and van Gunsteren, W. (1995) *J. Mol. Biol. 246*, 356−365.
14. Swint-Kruse, L., Matthews, K. S., Smith, P. E., and Pettitt, B. M. (1998) *Biophys. J. 74*, 413−421.
15. Wong, K. B., and Daggett, V. (1998) *Biochemistry 37*, 11182−11192.
16. Wrabl, J., Shortle, D., and Woolf, T. (2000) *Proteins 38*, 123−133.
17. Akke, M., Bruschweiler, R., and Palmer, A. (1993) *J. Am. Chem. Soc. 115*, 9832−9833.
18. Mercier, P., Spyracopoulos, L., and Sykes, B. D. (2001) *Biochemistry 40*, 10063−10077.
19. Yang, D., Mok, Y., Forman-Kay, J., Farrow, N., and Kay, L. (1997) *J. Mol. Biol. 272*, 790−804.
20. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem. 4*, 187.
21. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) *J. Chem. Phys. 79*, 926.
22. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Wiorkiewicz-Kuczera, J., and Karplus, M. (1992) *FASEB J. 6*, A143.
23. Meador, W., Means, A., and Quiocho, F. (1992) *Science 257*, 1251.
24. LeMaster, D., and Kushlan, D. (1996) *J. Am. Chem. Soc. 118*, 9255−9264.
25. Levy, R., Karplus, M., Kushnick, J., and Perahia, D. (1984) *Macromolecules 17*, 1370−1374.
26. Sridharan, S., Nicholls, A., and Honig, B. (1992) *Biophys. J. 61*, Abstract 995.
27. Richards, F. M. (1974) *J. Mol. Biol. 82*, 1−14.
28. Chakrabbatty, A., and Baldwin, R. (1995) *Adv. Protein Chem. 46*, 141−176.
29. Yang, A.-S., and Honig, B. (1995) *J. Mol. Biol. 252*, 351−365.
30. Prompers, J., and Bruschweiler, R. (2000) *J. Phys. Chem. B 104*, 11416−11424.
31. Prompers, J., and Bruschweiler, R. (2001) *J. Am. Chem. Soc. 123*, 7305−7313.
32. Yang, B.-H., and Sokwoo, R. (2000) *Robot. Autonomous Syst. 30*, 273−281.
33. Dunbrack, R. L., and Karplus, M. (1993) *J. Mol. Biol. 230*, 543−574.
34. Flynn, P. F., Bieber-Urbauer, R. J., Zhang, H., Lee, A. L., and Wand, A. J. (2001) *Biochemistry 40*, 6559−6569.
35. Liu, W., Flynn, P. F., Fuentes, E. J., Kranz, J. K., McCormick, M., and Wand, A. J. (2001) *Biochemistry 40*, 14744−14753.

BI026544Q