JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

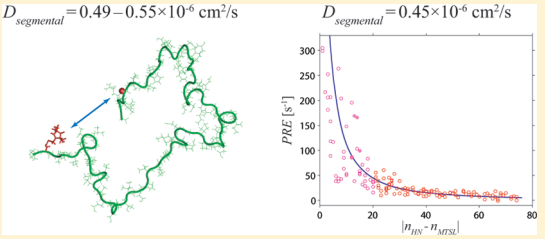
Motion of a Disordered Polypeptide Chain as Studied by Paramagnetic Relaxation Enhancements, ^{15}N Relaxation, and Molecular Dynamics Simulations: How Fast Is Segmental Diffusion in Denatured Ubiquitin?

Yi Xue and Nikolai R. Skrynnikov*

Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, Indiana 47907-2084, United States

Supporting Information

ABSTRACT: Molecular dynamics (MD) simulations have been widely used to analyze dynamic conformational equilibria of folded proteins, especially in relation to NMR observables. However, this approach found little use in the studies of disordered proteins, where the sampling of vast conformational space presents a serious problem. In this paper, we demonstrate that the latest advances in computation technology make it possible to overcome this limitation. The experimentally validated (calibrated) MD models allow for new insights into structure/dynamics of disordered proteins. As a test system, we have chosen denatured ubiquitin in solution with 8 M urea at pH 2. High-temperature MD simulations in implicit solvent have been carried out for the wild-type ubiquitin as well as MTSL-tagged Q2C, D32C, and R74C mutants. To recalibrate the MD data (500 K) in relation to the experimental conditions (278 K, 8 M urea), the time axes of the MD trajectories were rescaled. The scaling factor was adjusted such as to maximize the agreement between the simulated and experimental ^{15}N relaxation rates. The resulting effective length of the trajectories, 311 μs, ensures good convergence properties of the MD model. The constructed MD model was validated against the array of experimental data, including additional ^{15}N relaxation parameters, multiple sets of paramagnetic relaxation enhancements (PREs), and the radius of gyration. In each case, a near-quantitative agreement has been obtained, suggesting that the model is successful. Of note, the MD-based approach rigorously predicts the quantities that are inherently dynamic, i.e., dependent on the motional correlation times. This cannot be accomplished, other than in empirical fashion, on the basis of static structural models (conformational ensembles). The MD model was further used to investigate the relative translational motion of the MTSL label and the individual H^N atoms. The derived segmental diffusion coefficients proved to be nearly uniform along the peptide chain, averaging to $D = 0.49-0.55 \times 10^{-6}$ cm^2/s. This result was verified by direct analysis of the experimental PRE data using the recently proposed Ullman-Podkorytov model. In this model, MTSL and H^N moieties are treated as two tethered spheres undergoing mutual diffusion in a harmonic potential. The fitting of the experimental data involving D as a single adjustable parameter leads to $D = 0.45 \times 10^{-6}$ cm^2/s, in good agreement with the MD-based analyses. This result can be compared with the range of estimates obtained from the resonance energy transfer experiments, $D = 0.2-6.0 \times 10^{-6}$ cm^2/s.



INTRODUCTION

The interplay between order and disorder is an essential feature of any proteome. Those proteins where disorder prevails are known as Intrinsically Disordered Proteins (IDPs). They can be identified with relative ease on the basis of primary sequence—proteins with significant net charge and a low proportion of hydrophobic residues tend to be disordered.[1−3] Additional parameters, such as flexibility and bulkiness of amino acids, can also be incorporated in prediction algorithms.[4] Estimates obtained along these lines suggest that as many as 50% of eukaryotic proteins contain long stretches of disordered residues,[5] and more than 20% can be described as predominantly disordered.[4] A similar proportion of proteins is expected to be disordered *in vivo*, although the effect of macromolecular crowding may influence the proteins' folding status.[6,7] These statistics demonstrate that IDPs cannot be dismissed as a rare quirk of nature—on the contrary, they represent one of the broad and fundamentally important classes of proteins.

Of note, the proportion of IDPs in eukarya is much higher than in archaea or bacteria (in archaea, only 2% of proteins are predicted to contain disordered regions of 30 residues or longer[8]). This is not accidental. In the context of cell signaling and regulation,[9,10] the unfoldedness of IDPs confers a number of functional advantages. IDPs can literally wrap themselves around their binding partners, thus making use of multiple binding sites.[11−15] This mechanism makes it possible for IDPs to bind a wide range of targets. Such "multitasking" helps to

Received: February 20, 2011
Published: August 08, 2011

develop efficient, highly integrated signaling networks, where every protein serves in more than one role. In addition, IDPs offer the benefit of economy since they present large binding interfaces but do not require large supporting scaffolds.[16]

As often is the case with any advanced technology, IDPs are prone to breakdowns.[17] There is a significant link between protein disorder and cancer. An estimated 80% of human cancer-related proteins contain long disordered stretches.[9] It is also widely recognized that disordered proteins are key to most devastating neurodegenerative diseases.[18] It is fairly obvious why natively unfolded proteins tend to malfunction and cause trouble. Various perturbations, such as mutations, impaired posttranslational modifications, interactions with compromised ligands, etc., all can have grave consequences for IDPs, as they lead to proteolytic damage or trigger aggregation. Not surprisingly, IDPs emerge as a major potential target for newly developed drug therapies.[19−22]

Disordered peptides and proteins have been traditionally modeled via conformational ensembles. Conformational ensemble is a natural extension of the fundamental concept of protein structure.[23−33] Its disadvantage, however, is that it represents a collection of "still snapshots", which creates a problem with many important pieces of experimental data that are inherently dynamic. For example, it is difficult to incorporate ^{15}N relaxation data in ensemble calculations, other than in empirical fashion.[32,34,35]

One of the recent and increasingly popular methods for generation of conformational ensembles is constrained ensemble molecular dynamics (MD) simulation. In brief, multiple copies of the protein are simulated concurrently; at any given point in time an array of NMR parameters is calculated across the ensemble. The difference between the calculated and the measured values is converted into a pseudopotential, thus driving the system toward the final state which is consistent with the experimental data. (Alongside with ensemble-averaging schemes, time-averaging schemes have also been implemented, although they tend to be less stable.[36]) This approach has been used in conjunction with nuclear Overhauser effect rates (NOEs), residual dipolar couplings (RDCs), paramagnetic relaxation enhancements (PREs), chemical shifts, relaxation order parameters, and other types of experimental data.[37−45] Note, however, that this method only provides a convenient tool for generating static conformational ensembles: the constrained MD simulations, guided by pseudopotentials, cannot (and have never been claimed to) realistically model protein dynamics.

Here we demonstrate that *unbiased* MD simulations reached the stage where they can be used to model disordered proteins and successfully reproduce an array of experimental observables. Over the past 25 years, molecular dynamics has been used with increasing success to predict motion-dependent NMR parameters in folded proteins.[46] The same strategy has never been tested on unfolded proteins because it was tacitly accepted that relatively short MD trajectories could not adequately sample a vast conformational phase space available to an unfolded protein. In the past few years, the situation has changed. The newest computer hardware allows for microsecond-long MD trajectories to be recorded quickly and at modest cost. The use of implicit solvent and (moderately) elevated temperature further expands the scope of opportunities.

For this pilot study, we have chosen the sample of denatured ubiquitin (8 M urea, pH 2.0) that was shown to be a close approximation to a "random-coil" protein. This classification has been confirmed by the analysis of chemical shifts,[47] as well as scalar couplings,[47,48] auto- and cross-correlated relaxation rates,[49,50] and residual dipolar coupplings.[51] The random-coil-like behavior of unfolded proteins has been sometimes underappreciated since much (deserved) attention has been given to the investigation of residual structure in disordered proteins. It should be emphasized, however, that random-coil proteins provide a fundamentally important point of reference. Indeed, the concept of residual structure can be meaningfully defined only in relation to the random coil.

Real-life proteins never quite conform to the random-coil model. In the case of denatured ubiquitin, there is evidence of the ca. 10% content of the native-like β-hairpin at the N-terminal region (confirmed by detection of scalar couplings across hydrogen bonds).[52] This observation is supported by our PRE data. The main focus of this study, however, is not on the residual structure but rather on the prevailing random-coil characteristics of the denatured ubiquitin.

In this report. we concentrate on three pieces of experimental data and the corresponding MD-based predictions: radius of gyration, backbone ^{15}N relaxation, and paramagnetic relaxation enhancements (PREs). In particular, PREs are of primary interest. In our experiments, we used the conventional cysteine-attached MTSL labels[53] to measure the paramagnetic contribution into $^{1}H^{N}$ relaxation rates. This is arguably one of the most valuable experiments for studies of disordered proteins because: (i) nitroxyl spin labels are small and nonperturbing, (ii) each MTSL-tagged sample produces dozens of valuable pairwise constraints, and (iii) the data allow for rigorous and accurate interpretation.

So far, the PRE data from disordered proteins have been interpreted using the theory originally intended for rigid molecules isotropically tumbling in solution.[54] Recently we developed a number of more satisfactory treatments where the focus is on the relative *translational* motion of the paramagnetic label and the reporter spins ($^{1}H^{N}$).[55] One of the key parameters in these models is the rate of segmental diffusion in the disordered polypeptide coil. In the pioneering work by Gillespie and Shortle the empirical correlation times associated with the PRE rates were determined to lie in the range from 1 to 15 ns, with the average value 4 ns at 32 °C.[54,56] Since then the value of 4 ns has been used as a generic input for PRE analyses.[57−59] Although this estimate is undoubtedly correct by order of magnitude, the exact value is known to be problematic in more than one way.[55] In addition to the NMR-based determination, the time scale of the segmental diffusion in disordered proteins has been also determined using various FRET and FRET-like experiments, as well as electron transfer experiments.[60−68] The survey of these results finds a wide range of values for segmental diffusion coefficients in denatured proteins, from 0.2 to 6 $\times$ 10^{-6} cm^2 s^{-1}. The definition of diffusion coefficient in these studies is also model-dependent and does not necessarily meet the needs of NMR analyses. Here we have addressed this problem using the PRE data from three MTSL-tagged variants of ubiquitin, Q2C-, D32C-, and R74C-MTSL. The data were fitted using the semianalytical Ullman−Podkorytov model (diffusion in harmonic potential),[55,69] resulting in D = 0.45 $\times$ 10^{-6} cm^2 s^{-1} at 5 °C. At the same time, we have determined the values of D on the basis of the MD simulations of denatured ubiquitin. Prior to the analysis, the time axis of each trajectory was rescaled to correct for the effect of elevated temperature in the simulations. The calibrated trajectories were then used to analyze the relative motion of the MTSL moiety and the

individual H^N atoms. The diffusion coefficients obtained in this fashion proved to be surprisingly uniform, with little variation from residue to residue. The average values of D for three different MTSL-tagged mutants were found to lie in the narrow range from 0.49 to 0.55 $\times$ 10^{-6} cm^2 s^{-1}. Good agreement between the MD-based result and the value derived from the direct analysis of the experimental PRE data is encouraging. We anticipate that in the future many important aspects of disordered proteins structure/dynamics will be successfully investigated using properly calibrated MD models.

RESULTS AND DISCUSSION

MD Simulations. Ideally, the system at hand should be modeled using all-atom explicit-solvent simulation (i.e., aqueous solution of 8 M urea, 278 K). The water box should be sufficiently large to accommodate various extended conformations of ubiquitin. Such simulations, however, are problematic on several levels. Because of the large size of the water box, the explicit-solvent computations are expensive; it is, therefore, difficult to obtain satisfactory statistics. In particular, the PRE calculations can be demanding since the PRE rates depend on infrequent "close encounters" between the paramagnetic label and the reporter spins. Furthermore, the proteins in such trajectories often remain too compact or even structured despite the presence of urea molecules.[70]

As an alternative, we carried out the simulations with implicit (aqueous) solvent, using elevated temperature to denature the protein. The computations relying on implicit solvent are much faster. In the case of disordered protein, which requires a large water box, there is an order-of-magnitude speed-up. Use of the elevated temperature further improves the statistics. Clearly, this approach is empirical. The primary justification for this approach is that it proves to be fairly successful in reproducing the array of the experimental data from the system at hand (described below).

In preparing for this study, we have recorded many trial MD trajectories using a variety of force fields,[71−75] MD platforms,[71,74,76,77] and solvation models.[78−83] The general tendency in these simulations is that the protein chain remains too compact even above 500 K. Another unwanted trend is the frequent occurrence of α-helices (experimental chemical shifts suggest that there is little or no α-helical propensity in denatured ubiquitin; see Figure S1C, Supporting Information (SI)). This artifact is not unexpected—excessive helical propensity has been widely observed in MD simulations before.[84−87] On a more subtle level, most of the trial MD simulations showed only moderately good agreement with our ^{15}N relaxation data (discussed below).

After much experimentation, we have settled on the Amber 11 package using an ff99SB force field.[88] A number of recent comparative studies, in particular those based on the experimental NMR data, favor this force field over others.[86,89−92] The trajectories were recorded using Langevin dynamics[93] and the generalized Born solvent involving the so-called pairwise descreening approach (option igb = 1).[80] To ensure that the protein is fully unfolded, the simulation temperature was adjusted to 500 K. Note that high-temperature simulations have been successfully used to study protein unfolding in the past.[94−96] Four trajectories involving wild-type ubiquitin, as well as MTSL-tagged Q2C, D32C, and R74C mutants were recorded using the GPU-based workstation (see Materials and Methods). The force-field parameters for the MTSL moiety were carefully adapted from the recent work by Sezer et al.[97] (SI). The nominal length of each trajectory was 1 μs; all simulations were completed in approximately two weeks time. During the processing of the high-temperature MD trajectories, the "time axis" of each simulation was rescaled to emulate slower dynamics at ambient conditions. After rescaling, the effective length of each trajectory amounts to 311 μs—sufficient to achieve good convergence. As demonstrated below, this MD protocol leads to a good agreement with the experimental data acquired from denatured ubiquitin.

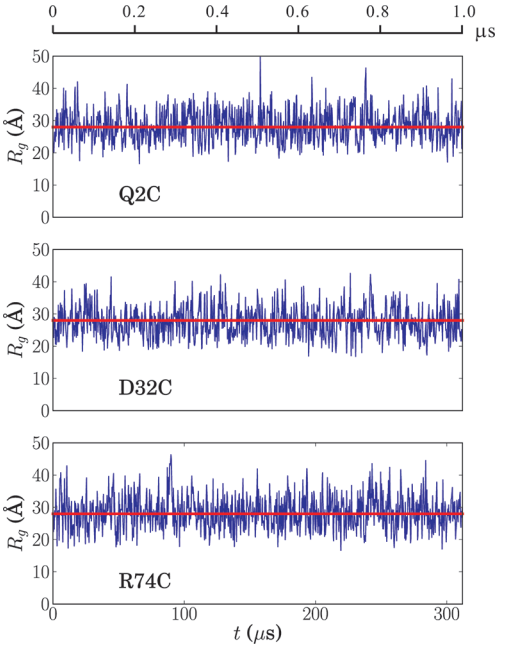


Figure 1. Time variation of R_g in denatured ubiquitin according to MD simulations. The data are shown for three different MTSL-tagged mutants of ubiquitin, as indicated in the respective panels. Red line indicates the experimental result, 28.0 Å.[98] The scale on the top of the plot represents the original (unscaled) time axis of the MD simulations. Of interest, the autocorrelation functions for R_g are monoexponential, with the correlation times in the range from 275 to 360 ns (scaled). This is close to the recently described folding "speed limit".[100,101]

Radius of Gyration. The radius of gyration of denatured ubiquitin was recently determined in the SAXS/SANS study by Gabel et al.[98] The experimental conditions employed—8 M urea in H_2O solvent, pH 2.5—were nearly identical to those used in our NMR measurements. Of note, in their data treatment the authors explicitly considered scattering by urea molecules coordinated to ubiquitin. The effect of hydration shell was also analyzed on the basis of the H_2O/D_2O SANS experiments and deemed insignificant. The reported radius of gyration (28.0 Å) carries a significant uncertainty (±3.5 Å), recognizing the inherent complexity of the measurements which often remains overlooked. This sizable error margin is likely to absorb the modest temperature dependence of R_g (20 °C in Gabel's study vs 5 °C in our work).[99]

Figure 1 shows the fluctuations of R_g during the time course of the MD simulations. The results are shown for three MTSL-tagged mutants of ubiquitin, which are known to be virtually undistinguishable from the wild-type protein (as confirmed by the invariance of chemical shifts). The average R_g values as

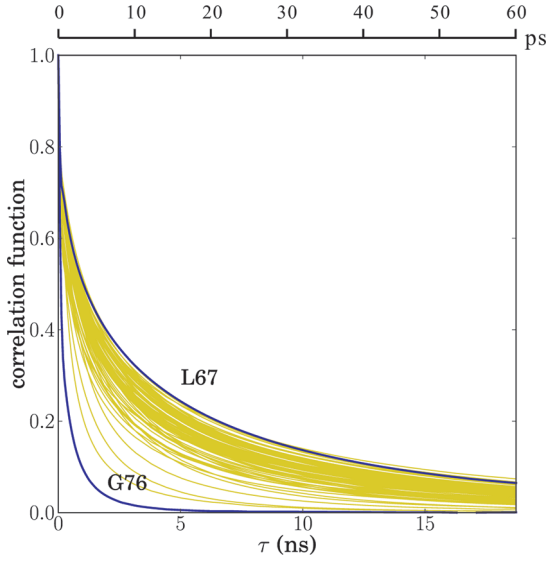


Figure 2. Correlation functions $g_{\text{reorient}}(\tau)$ of $^{1}H^{N}-^{15}N$ vectors from the MD simulations of denatured wt ubiquitin. The time axis in the plot is rescaled, and the original axis from the MD simulation is shown above the graph. The residues showing the fastest and slowest correlation decay are shown with blue curves.

calculated for Q2C-, D32C-, and R74C-MTSL are 28.3, 27.6, and 28.5 Å, respectively—all in good agreement with the experimental value of 28.0 Å (horizontal red line). The simulation involving wild-type ubiquitin produced the average R_g value 28.7 Å (not shown). The nominal time axis of the MD simulation is shown at the top of Figure 1, and the rescaled axes are shown below each panel (see next section for details). In the discussion that follows we always refer to the rescaled time intervals, unless indicated otherwise.

Visual inspection of the plots points toward the presence of ubiquitin conformers that remain relatively stable, in a sense of characteristic R_g, on the time scale of ~10 μs. Against this background, we observe fast excursions to more extended or more compact forms (needle-like spikes in the plot). Note that throughout the course of simulations ubiquitin never comes close to the folded state, R_g = 14.4 Å (which indeed cannot be expected under strong denaturing conditions).

^{15}N Relaxation Rates. Backbone ^{15}N relaxation is a valuable source of information on conformational equilibria in IDPs. In essence, the relaxation data report on the degree of local motional "constrainedness". Some of the variation in ^{15}N relaxation rates in disordered peptide chains can be attributed to steric constraints offered by individual side chains; e.g., glycine-rich segments tend to be more flexible and proline-rich segments less flexible, disulfide bonds impose additional restraints, etc. Consequently, the rates often show a good correlation with bulkiness of side chains, separation of a given residue from a disulfide bond, and other similar parameters.[32,34,102,103] In the case of biologically relevant IDPs (as opposed to random-coil-like models), nitrogen relaxation reveals the presence of transient secondary structure,[104−110] as well as tenuous clusters formed as a result of hydrophobic and possibly electrostatic interactions.[111−117]

The protocol for the MD-based simulation of the ^{15}N relaxation rates in globular proteins is very well established.[46,118,119] In principle, it can be easily adapted for IDPs; however, little progress has been made in this direction because of the concerns about statistical sampling of conformational dynamics in relatively short MD trajectories.[120,121] Here we demonstrate that the standard MD-based methods can actually be used to simulate ^{15}N relaxation rates in disordered proteins. The sampling achieved with the present set of trajectories is fully satisfactory, leading to good convergence properties.

We have started by evaluating the MD correlation functions

$$g_{\text{reorient}}(\tau) = 4\pi \langle Y_{20}(\theta_{\text{HN}}(t+\tau)) \cdot Y_{20}(\theta_{\text{HN}}(t)) \rangle \qquad (1)$$

where θ_{HN} is the time-dependent angle between the $^{1}H^{N}-^{15}N$ vector and the z-axis. The collection of the extracted correlation functions is presented in Figure 2.

The shape of $g_{\text{reorient}}(\tau)$ is familiar—sharp drop caused by fast peptide plane librations, followed by a steady decay which reflects conformational rearrangements of the peptide chain. More thorough analysis identifies three exponential components with the average weights/correlation times of (0.30/44 ps), (0.42/1.4 ns), and (0.28/9.4 ns). The latter two components can be attributed to *local* and *global* conformational transitions (discussed in more detail in the next section). The statistical sampling of the relevant motions is clearly good, as $g_{\text{reorient}}(\tau)$ profiles in Figure 2 are smooth and decay to zero. The results are also broadly consistent with the previously derived parametrizations of $g_{\text{reorient}}(\tau)$. For instance, model-free analysis of ^{15}N relaxation data in the denatured drkN SH3 domain at 14 °C led to biexponential correlation functions, (0.49/106 ps) and (0.51/4.1 ns).[122]

The correlation functions shown in Figure 2 have been used to calculate the relaxation parameters using the standard set of formulas.[123] Recall that in the case of folded proteins the MD trajectory is first processed to remove overall molecular tumbling, and then the tumbling is reintroduced through the multiplicative exponential factor applied to the correlation function.[124] With the current trajectories there is no need for such manipulations. However, to match the simulation results (500 K) with the experimental data (278 K, 8 M urea), it is necessary to rescale the time axis of the MD trajectory. Specifically, we assumed that the interval between the two consecutive MD snapshots is equal to $\alpha \cdot \Delta t$, where Δt = 0.2 ps is the nominal MD time step and α is a floating parameter. The value of α was then optimized in the least-squares sense to maximize the agreement between the simulated and the experimentally measured ^{15}N R_1 and R_2 rates. The outcome of this procedure is presented in Figure 3. The best-fit scaling factor α is 311; the corresponding rescaled time axis is shown at the bottom of the graphs throughout the paper. This rescaling procedure is similar in spirit to the one devised by Prompers and Brüschweiler[121] but significantly more straightforward. (These authors used multiple scaling factors applied to dominant motional modes as found in the MD simulation. While conceptually elegant, this approach redefines protein dynamics and makes it impossible to reconstruct the motions in a form of all-atom trajectory).

Figure 3 shows the experimental and simulated ^{15}N R_1 and R_2 data from the wild-type ubiquitin (magenta) and reduced MTSL-tagged D32C mutant (cyan). The agreement between the experimental and simulated rates is encouraging (cf. left and right portions of the plot). Note that this agreement is achieved using a single global tuning parameter, α. In the case of the R_2 profile, many fine features are reproduced. For example, mobile

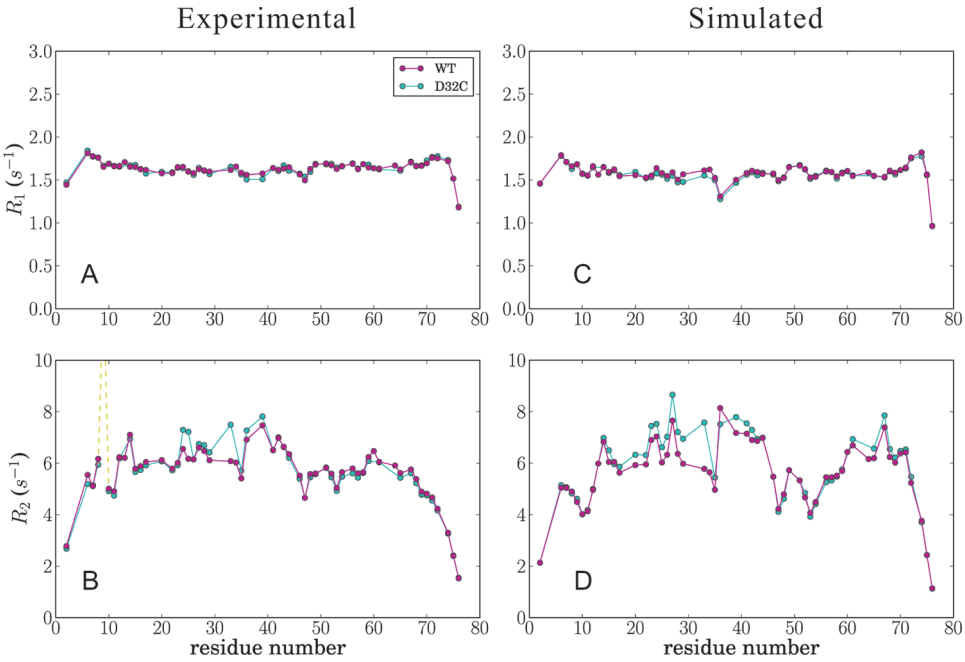


Figure 3. Experimental and simulated ^{15}N relaxation rates in denatured wt ubiquitin (magenta) and reduced MTSL-tagged D32C mutant (cyan). High R_2 rate for residue T9 (dashed line) is due to the R_{ex} contribution.[50] The selection of the points in panels C and D matches that in the experimental data set.

glycine residues, G35, G47, and G53, display lower R_2 rates, whereas the residues adjacent to the inflexible proline pair, P37−P38, display higher rates. The simulations also nicely capture the local increase in R_2 due to the MTSL tagging at the position C32 (cf. the deviation between magenta and cyan profiles).

Note, however, that MD simulations tend to exaggerate the local variability in the R_2 data (e.g., the experimental profile appears to be more flat). In particular, near the C-terminus the simulations predict somewhat higher rates than observed experimentally. This behavior is caused by the appearance of the transient α-helix in the simulations, residues 66−73. Application of the program DSSP[125] to the frames from the MD trajectory finds modest helical propensity in this region, up to 8% (see Figure S1B, SI). Similar observations have been previously made from the analysis of the conformational ensemble generated by Huang and Grzesiek.[42] In both cases, the results are likely to be influenced by excessive helical character of the force fields.[84,85] Analyses of experimental data, both in Figure 3 and elsewhere,[50,52] suggest that the actual helical content in this region is probably very small.

The experimental results in Figure 3A,B highlight the similarity between the wild-type protein and the MTSL-tagged D32C mutant: aside from the residue D32/C32 and a handful of surrounding sites, the relaxation rates from the two systems are virtually identical. A nonperturbing nature of the MTSL label is also confirmed by the near-perfect agreement between the two sets of chemical shifts. This result is reasonably well reproduced in the simulations (Figure 3C,D). Note that the good agreement between the two profiles, magenta and cyan, also demonstrates good convergence properties of the simulations (indeed, these data are derived from two different trajectories). The convergence properties can also be tested in a simpler fashion—by dividing each trajectory into two halves, 155 μs + 155 μs, and comparing the respective relaxation rates. Such tests indicated excellent convergence for all of the simulated ^{15}N relaxation parameters, R_1, R_2, and NOE, with statistical uncertainty well below the experimental error (not shown).

Another convergence-related issue is the choice of the sampling step in the MD trajectory. The nominal step, Δt = 0.2 ps, is transformed into $\alpha \cdot \Delta t$ = 0.62 ps. It is not a priori obvious that this is sufficient to sample the fast-decay portion of the correlation function in Figure 2. In fact, the characteristic time of the fast decay in $g_{reorient}(\tau)$ is, on average, 44 ± 6 ps, so that it may appear that the sampling step is inadequate. As it turns out, however, the chosen step works fine—switching to more frequent sampling, Δt = 0.05 ps, does not change the determined correlation time τ_{fast} or any of the simulated relaxation parameters. The key to this favorable property is our treatment of the correlation functions: $g_{reorient}(\tau)$ are first fitted with multiexponential curves and then converted to spectral densities (see Materials and Methods).[126] As it turns out, this procedure allows for relatively sparse sampling of the trajectory, translating into critical gains in MD performance and storage.

^{1}H^N−^{15}N heteronuclear NOEs in denatured wt ubiquitin were analyzed separately, as a part of the broad validation agenda. The comparison between the experimental and simulated NOE data is presented in Figure 4. On average, the simulated values are somewhat lower than the experimental ones: 0.15 vs 0.28 (for residue range 6−74, corresponding to the plateau in the graph). In particular, the simulations predict significantly lower NOE values around residues G35 and G53 (Figure 4B). In reality, only small dips in the NOE profile are observed around these residues (Figure 4A). The inspection of $g_{reorient}(\tau)$ suggests that ∼1 ns time-scale motions are somewhat exaggerated in the MD simulations for these glycine residues. Generally, the simulations predict somewhat greater amount of site-by-site variability than observed experimentally.

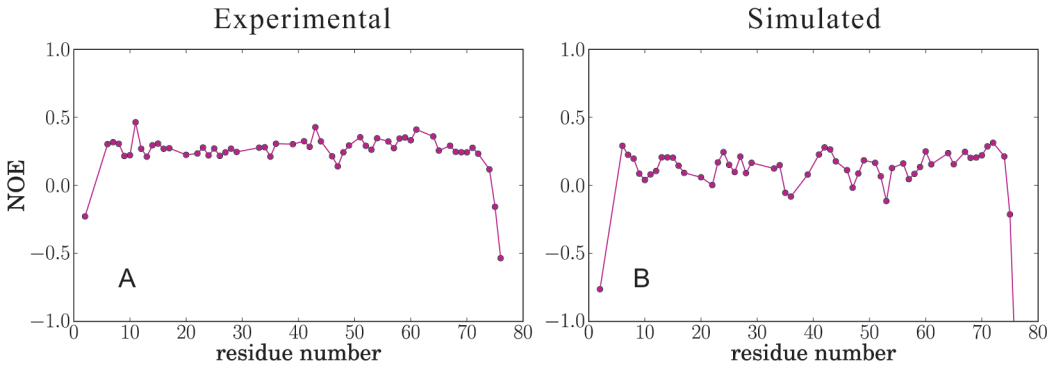


Figure 4. $^{1}H^{N}-^{15}N$ heteronuclear NOE in denatured wt ubiquitin, as obtained from the experimental measurements (A) and MD simulations (B).

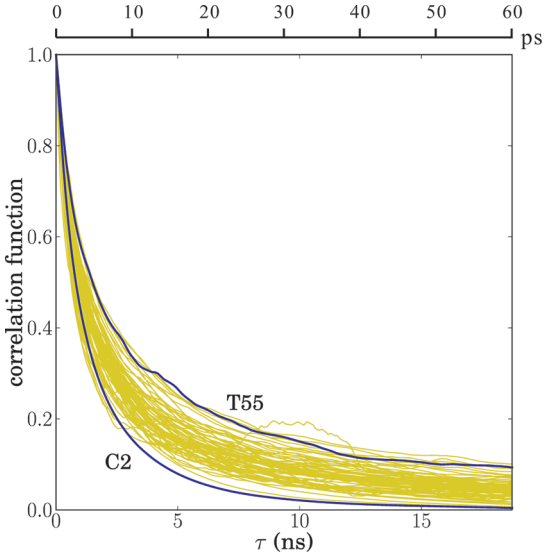


Figure 5. Correlation functions $g(\tau)$ for the dipolar vectors connecting $^{1}H^{N}$ to the paramagnetic center (nitrogen atom of the MTSL ring) as derived from the MD simulation of the MTSL-tagged Q2C mutant of ubiquitin. For the purpose of plotting, $g(\tau)$ are normalized such that $g(0) = 1$. The residues showing the fastest and slowest correlation decay are shown with blue curves.

The NOE experiment samples local dynamics quite differently from R_2 or R_1; furthermore, the MD results were in no way tuned to match the experimental NOE values. Thus, the comparison presented in Figure 4 can be viewed as a largely independent validation of the MD model. The discrepancy in the NOE values on the level of ca. 0.1 unit can be viewed as modest. Given that MD simulations are generally not very good in capturing the motional correlation times,[127,128] the level of agreement observed in Figure 4 can be deemed satisfactory. We conclude that the present MD model reproduces (relaxation-active) picosecond−nanosecond motions with reasonable accuracy, subject to only small biases.

PRE Rates. While ^{15}N relaxation represents the localized probe of IDP structure/dynamics, PREs report on long-range pairwise interactions. Some of the fundamental aspects of the PRE effect in disordered proteins remain poorly understood. The expected PRE profiles for random-coil-like proteins have been presented in the literature.[42,59,129,130] They are based, however, on the rotational relaxation model which is, strictly speaking, not suitable for the strongly disordered systems. Moreover, these predictions have never been thoroughly tested experimentally. This rather uncertain situation makes it difficult to use the PRE data as the indicator of nonrandom behavior (i.e., residual structure).[42,117,131] Developing more accurate theoretical treatments should greatly facilitate the use of PRE data for detection of residual tertiary structure. As a first step in this direction, we recently developed a number of models for the PRE effect in disordered proteins, taking into consideration both *rotational* and *translational* relaxation mechanisms.[55]

The PRE correlation function

$$g(\tau) = 4\pi\left\langle \frac{Y_{20}(\theta(t+\tau))}{r^3(t+\tau)} \cdot \frac{Y_{20}(\theta(t))}{r^3(t)} \right\rangle \qquad (2)$$

refers to the dipolar vector connecting $^{1}H^{N}$ spin with a paramagnetic center assumed to be localized on the nitrogen atom of the MTSL ring.[132] In this correlation function, the memory loss occurs not only through random reorientation of the dipolar vector, $\theta(t)$, but also through modulation of the vector length, $r(t)$. Due to the steep (cubic) dependence of the dipolar interaction strength on the distance, this latter mechanism is very effective.

The set of the PRE correlation functions $g(\tau)$ from the MD trajectory of the MTSL-tagged Q2C mutant of ubiquitin is shown in Figure 5. Note that the curves are reasonably smooth, signifying good convergence (see below). An apparent exception is the "bump" that appears in one of the correlation functions (that of the residue R72) around $\tau = 15$ ns. This feature, however, has a rational explanation: as it turns out, the NH group of R72 briefly forms a hydrogen bond with the NO group in MTSL (see the MD movie, SI). This is a one-of-a-kind event, which means that the corresponding correlation function has less-than-perfect statistical properties. In the subsequent treatment, this statistical effect (the bump) is largely suppressed since $g(\tau)$ profiles are smoothened by means of the six-exponential fitting (see Materials and Methods).

The multiexponential correlation functions in Figure 5 can be to a good approximation described as bimodal. The average weights and correlation times of the two exponential components are (0.74/1.3 ns) and (0.26/14.1 ns). What is the origin of these two components? The inspection of the MD trajectory reveals that most of the conformational dynamics occurs *locally* (see the MD movies, SI). In other words, a change in the backbone dihedral angles of a certain residue usually does not cause any global conformational rearrangement. Instead, it is played out at the local level, i.e., through compensatory motions

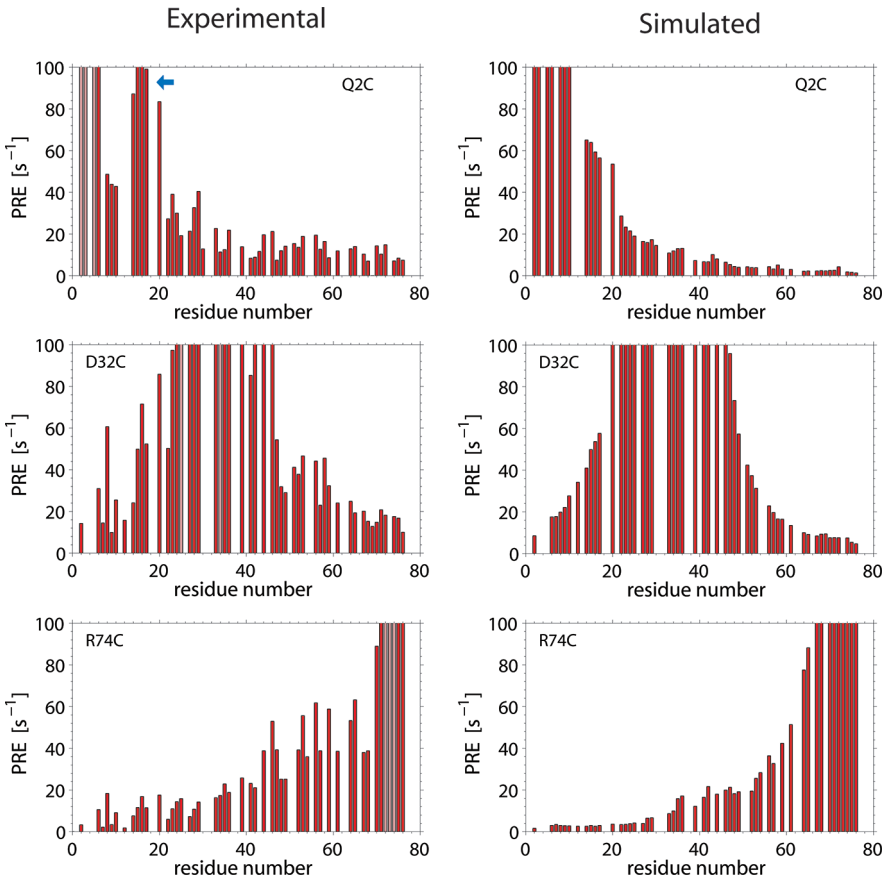


Figure 6. Experimental and simulated $^{1}H^{N}$ PRE rates in three MTSL-tagged mutants of ubiquitin under strong denaturing conditions. The displayed rates are limited to 100 s^{-1} since very high rates generally cannot be measured with sufficient accuracy.[55] Pale (pink) bars correspond to unobservable or extremely weak peaks that cannot be quantified. Blue arrow marks large PRE rates indicative of the residual β-hairpin structure. Note that the comparison presented in this figure involves no tunable parameters (the time scaling factor α has been adjusted on the basis of the ^{15}N relaxation data alone). To verify the convergence, the Q2C trajectory was divided into two halves (i.e., cut at half-length), and each half was used to independently simulate the PRE rates. The results were in agreement with the rms deviation of 3 s^{-1}.

in the adjacent residues. Such local conformational changes, involving short segments of the chain, can be quite effective in modulating the dipolar interaction between the $^{1}H^{N}$ probe and the MTSL label, primarily through modulation of the distance, $r(t)$, eq 2. This local conformational dynamics is the source of the initial decline in the correlation function $g(\tau)$.

Alternatively, there is a possibility that change in one individual dihedral angle would trigger a global conformational change (e.g., opening/closure). This does not happen readily because of the hydrodynamic drag—large displacements of the big fragments of the chain encounter the resistance from solvent. The hydrodynamic drag is the reason why local conformational changes are favored over global. Nevertheless, the global conformation of the peptide chain also changes in due course, leading to effective modulation of the dipolar vectors. These relatively slow global conformational changes are responsible for the long tails of the correlation functions $g(\tau)$, as seen in Figure 5.

In this respect, it is interesting to point out that MD simulations in vacuum show a very different pattern. In the absence of hydrodynamic drag, the changes in individual dihedral angles bring about extensive conformational rearrangements (visualized in the SI of ref 55). The distinction between local and global dynamics is blurred, and the PRE correlation functions are, to a very good approximation, monoexponential. Clearly, the presence of solvent in the MD simulations is critical, resulting in a qualitatively different picture of motion.

It is interesting to find out whether PRE relaxation occurs through orientational, $\theta(t)$, or translational, $r(t)$, degrees of freedom. We are primarily interested in the residues that are sufficiently well separated from the MTSL tag (by 10 residues or more). As it turns out, for such residues the translational mechanism, $r(t)$, completely dominates the PRE effect. As for the residues immediately next to the MTSL site, these are of little practical interest since their spectral peaks are broadened beyond detection. It can be mentioned, however, that for such proximal sites the $r(t)$ mechanism is less efficient (because short fragments of the chain are relatively stiff), and the contribution from $\theta(t)$ is greater.

Finally, it is instructive to compare the characteristic time scales found in $g(\tau)$, 1.3 and 14.1 ns, with those previously observed in $g_{\mathrm{reorient}}(\tau)$, 44 ps, 1.4 ns, and 9.4 ns. As expected, fast (picosecond) angular fluctuations are relevant in the context of ^{15}N relaxation but cannot effectively modulate the long-range PRE interactions. On the other hand, the nanosecond modes representing local and global conformational dynamics are efficient in both cases. The respective correlation times, as extracted from $g(\tau)$ and $g_{\mathrm{reorient}}(\tau)$, are similar.

The correlation functions $g(\tau)$ can be readily converted into the PRE rates using the standard relaxation rate formulas (see also Materials and Methods).[55] In Figure 6 we compare the simulated PRE rates with the experimental results (right and left columns, respectively). Overall, the simulated and experimental profiles are remarkably similar. Yet, there are also noticeable differences—especially at the wings of the PRE profiles (top and bottom panels in Figure 6). Far from the MTSL label, the simulated data show fairly uniform PRE rates on the order of several seconds inverse. In contrast, the experimental rates are far less uniform, ranging from 0 to nearly 20 s^{-1}. The presence of the sizable PRE rates of up to 20 s^{-1} can be interpreted as a manifestation of residual structure. In other words, it can be viewed as the evidence of minor conformational species where certain residues are brought in a relatively close contact with MTSL. These subtle conformational preferences are, apparently, lost in the high-temperature MD simulation, which shows more homogeneous behavior.

Separately, we would like to discuss the evidence of the native-like β-hairpin in the N-terminal part of denatured ubiquitin. Focusing on Q2C-MTSL data (top left panel), we note an interesting local pattern: PRE rates drop to a moderate level in residues 8−10 and then again rise sharply in residues 14−17 (marked by blue arrow in the plot). The origin of this effect becomes clear if we turn to the recent publication by Meier et al.[52] These investigators have demonstrated—in particular, by directly measuring the scalar coupling constants across transient hydrogen bonds—that the conformational ensemble of denatured ubiquitin contains ca. 10% of the species featuring the native-like N-terminal β-hairpin. In this hairpin, L15 and E16 lie in close proximity to the MTSL-labeled Q2C,[133] with L15 hydrogen-bonded to I3.[45,52] These observations nicely explain the PRE pattern observed in our data. Interestingly, the β-hairpin in question makes a brief, ca. 0.3 μs, appearance in the high-temperature MD trajectory of wt ubiquitin. However, this corresponds to only 0.1% of the full trajectory length and thus has no measurable impact on the simulated PRE rates.

While the experimental PRE pattern in the N-terminal region of Q2C-MTSL is most distinctive, the general ruggedness of the PRE profiles suggests the presence of multiple minor species populated at the level of several percent. Capturing these species in the MD simulations is a challenging task, given their transient nature and inherent diversity. However, the principal motif that dictates the shape of the PRE profiles is the prevailing random-coil-like behavior of the denatured ubiquitin. From this perspective, our MD model offers reasonable (near-quantitative) agreement with the experimental data. Note that both experimental data and predictions shown in Figure 6 are significantly different from the empirical dependence that has been previously introduced to describe the PRE data in a random-coil protein.[42,57,129,130,134]

The simulated PRE data shown in Figure 6 can also be used to assess the quality of the standard Gillespie−-Shortle formula.[54] Using this formula, the PRE rates can be converted into distance factors $\langle r^{-6}\rangle$; the results can be subsequently compared with $\langle r^{-6}\rangle_{MD}$ extracted directly from the MD trajectories. Such comparison demonstrates that there is a significant amount of uncertainty associated with the use of Gillespie−Shortle formula (Figure S2, SI). This is in particular true with respect to the empirical correlation time τ_c involved in this formula. The optimization of τ_c in our calculations produced the value 5.1 ns (5 °C). At room temperature, this time is expected to be significantly shorter.[55]

Chemical Shifts. We have used all frames in the MD trajectory of wt ubiquitin as input for the chemical shift calculation program SHIFTX.[135] The calculated $^{13}C^{\alpha}$, $^{13}C^{\beta}$ shifts (indicative of the secondary-structure preferences[136]) were averaged over all frames and then compared to the experimental values. The outcome of this comparison is illustrated in Figure S1A (SI). The obtained Pearson coefficient is $r = 0.99$, with rms deviation of 1.9 ppm. While this level of agreement appears impressive, the result is largely moot—it simply indicates that SHIFTX is capable of reproducing the random-coil shifts with sufficient accuracy. In fact, the agreement is somewhat worse than what has been obtained in the original tests using a set of folded proteins, rmsd < 1.1 ppm.[135]

While successfully reproducing the general random-coil character of denatured ubiquitin, our MD model fails to capture the subtle secondary-structure preferences. Figure S1B (SI) illustrates the results of DSSP[125] analysis of the wt MD trajectory. The content of α-helix in this trajectory is on average 3%, reaching 8% near the C-terminus; the content of β-sheet is negligible. On the other hand, when we used the program SSP[137] to determine the secondary-structure propensity based on the experimental chemical shifts, we found that the protein displays a preference for β-sheet conformations (Figure S1C, SI). Specifically, the average content of β-sheet amounts to ca. 7%, while the average content of α-helix does not exceed ca. 1.5% (see also the discussion in the previous section).

The internal motions in denatured ubiquitin are sufficiently fast, which leads to efficient averaging of the fluctuating chemical shifts. In this situation, the precise values of motional correlation times are unimportant—it is sufficient to generate a representative conformational ensemble. Under these circumstances, the ensemble-generation programs such as ENSEMBLE[35] and ASTEROIDS[33] show good success in predicting chemical shifts. This is not surprising since the ensembles are generated under the control of experimental chemical shifts, residual dipolar couplings, and other similar data. Our model, on the other hand, is derived from the MD simulation which is essentially *unconstrained*. Its unique strength lies with the ability to predict NMR parameters that are significantly dependent on motional correlation times. As for the chemical shifts, future progress in this area depends on the improvements in the MD protocol and in the underlying force field.

Segmental Diffusion Coefficient from MD Simulations. The segmental diffusion coefficient D describes relative motion of the two probes attached to the polypeptide chain. In our case, these are the MTSL tag (or, more specifically, the unpaired electron localized on the nitrogen atom of the MTSL pyrrolinyl ring) and one of the $^{1}H^{N}$ protons. Over the last two decades, there has been intense interest in determination of D in disordered proteins.[60−68] The results obtained in this work offer an excellent opportunity to revisit this problem, which so far lacks the consensus solution.

The MD model generated in this work has been "calibrated" against the backbone ^{15}N relaxation rates. We further reason that if the trajectory correctly reproduces torsional angle dynamics $\varphi(t)$, $\psi(t)$ (as sampled by ^{15}N R_1, R_2) then it would automatically reproduce the global conformational dynamics, including long-range contacts (as sampled by PREs). This assumption is verified by comparing the simulated and experimental PRE rates (Figure 6). On the basis of these results, we suggest that our MD

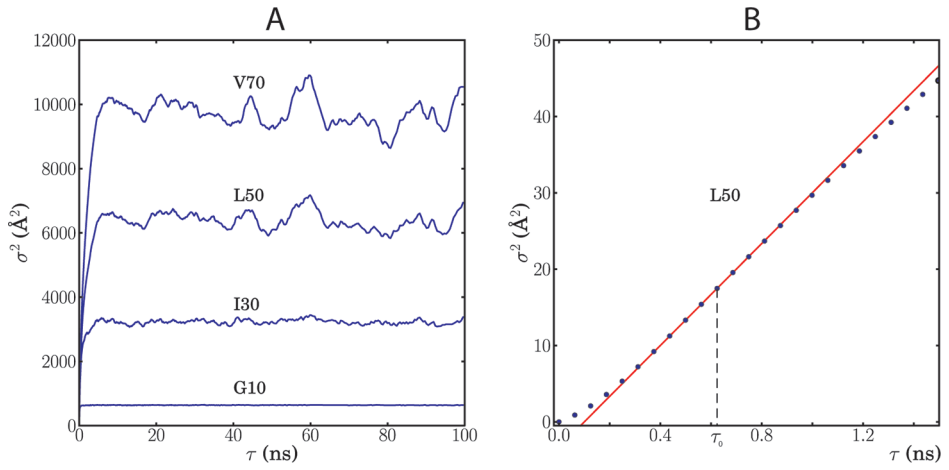


Figure 7. (A) Mean square displacement $\sigma^2(\tau)$ for selected amides in Q2C-MTSL trajectory. The plateau, which is proportional to $<|\vec{r}\,(t)|^2>$, is defined well for G10 but more poorly for distant residues V70 (the statistics is more of a problem for residues far removed from the MTSL site). Note that the plateau is reached after ca. 10 ns, consistent with the characteristic decay time of $g(\tau)$ (Figure 5). (B) Expansion of the initial portion of the $\sigma^2(\tau)$ curve for residue L50. Red line indicates the tangent $d\sigma^2(\tau)/d\tau$ evaluated at the inflection point τ_0.

model can be used to determine other motional parameters of interest.

The diffusion coefficients can be extracted from the MD data using the following formulation of the Einstein relation.[138]

$$D = \frac{1}{6}\frac{d\sigma^2(\tau)}{d\tau}\bigg|_{\tau=\tau_0} \qquad (3)$$

where $\sigma^2(\tau)$ is the mean square displacement of one particle relative to the other, $\sigma^2(\tau) = <|\vec{r}\,(t+\tau) - \vec{r}\,(t)|^2>$, and $\vec{r}\,(t)$ is the vector connecting $^1H^N$ to the paramagnetic center, cf. eq 2. The choice of τ_0 is discussed below (note that in the case of free diffusion the derivative is evaluated at the point where the system is safely in the diffusion regime, $\tau_0 \rightarrow \infty$[139]).

The calculated dependence $\sigma^2(\tau)$ for selected amide sites in Q2C-MTSL is illustrated in Figure 7A. It is easy to see that $\sigma^2(0) = 0$ and $\sigma^2(\infty) = 2<|\vec{r}\,(t)|^2>$. In other words, $\sigma^2(\tau)$ converges to the plateau, proportional to the mean square length of the dipolar vector. We are particularly interested in the initial part of the $\sigma^2(\tau)$ curve; the expansion of the corresponding portion of the graph for the residue L50 is shown in Figure 7B. Analyzing this graph starting from $\tau = 0$ we notice that the system first enters into the so-called free-flight regime, $\sigma^2(\tau) \sim \tau^2$.[140] Next, it progresses to the linear diffusion regime, $\sigma^2(\tau) \sim \tau$, consistent with eq 3. Further down the road, the system starts sensing the constraining effect of the peptide chain linkage. As a result, the curve gradually bends and eventually converges toward the plateau. (Indeed, the mean square distance between $^1H^N$ and MTSL cannot increase indefinitely because they are linked through the finite-length chain).

From the perspective of the D determination, we are interested in the linear regime. Specifically, we choose the point in the middle of the linear region, where the derivative is maximum, for the role of τ_0 (marked in the graph). This is an empirical definition, assuming that at $\tau = \tau_0$ the system fully enters into the diffusion regime but does not yet sense the restraining effect of the linkage. Using this operational definition, we have computed the values of D for all amide sites in the three MTSL-tagged mutants investigated in this study (Figure 8).

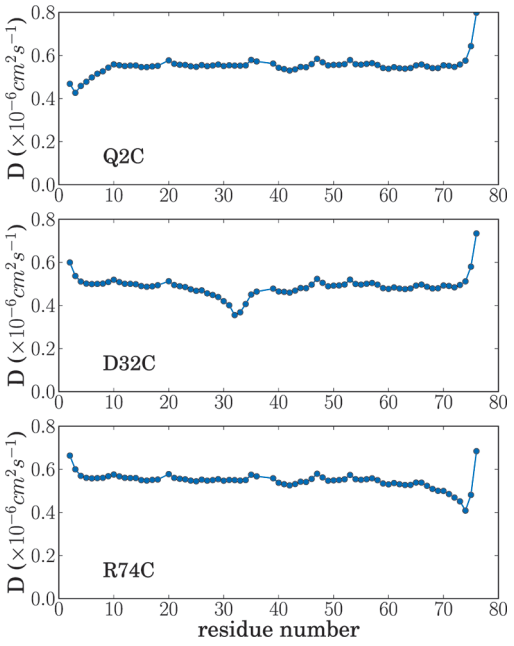


Figure 8. Segmental (mutual) diffusion coefficients for relative motion of the $^1H^N$ spins and the MTSL paramagnetic center, as derived from the MD trajectories of the three MTSL-tagged ubiquitin mutants under denaturing conditions.

The results in Figure 8 represent the coefficients of mutual (segmental) diffusion between the paramagnetic center associated with MTSL and the individual amide protons. The remarkable thing about the profiles in Figure 8 is how flat they are. Some limited variability is observed near the MTSL attachment sites, where the values of D are somewhat lower. This should be attributed to stiffness, which impedes the relative motion of the probes. Toward the termini, the values of D become higher. This clearly reflects the increased motional freedom, especially in the flexible C-terminus. Finally, very slight increases in D are registered around glycine residues G10, G35, G47, and G53. These trends are self-evident and have

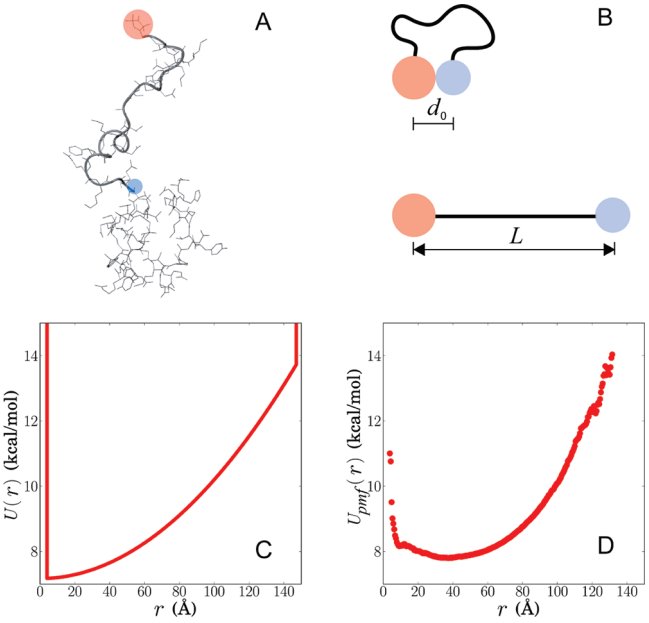


Figure 9. (A) Model of the two spherical particles tethered by a flexible string. (B) The distances of the minimum/maximum separation between the particles. (C) Harmonic potential $U(r)$ with reflecting walls, as calculated for residue T55 in Q2C-MTSL (n_{HN} = 55, n_{MTSL} = 2, other parameters as listed in the text). (D) Mean force potential $U_{pmf}(r)$ for the H^N atom of the residue T55 and the MTSL paramagnetic center (assumed to be localized on the nitrogen atom of the pyrrolinyl ring). Computed on the basis of the MD trajectory of Q2C-MTSL following the protocol described in ref 55. Note that $U(r)$ and $U_{pmf}(r)$ are residue-specific; the example shown here is for the purpose of illustration.

been previously discussed in the resonance energy transfer studies.[141,142]

The D values averaged over all residues in a given mutant amount to 0.55, 0.49, and 0.55 × 10^{-6} cm^2 s^{-1} for Q2C, D32C, and R74C, respectively. These average values are essentially identical to the "plateau" values as seen in Figure 8; i.e., they represent the segmental diffusion coefficients for a vast majority of residues. Note that the average D values in the three mutants follow the same pattern as the average R_g (see Figure 1). Specifically, the mutant with the central placement of the MTSL tag, D32C, turns out to be slightly more compact (lower R_g), with somewhat slower chain dynamics (lower D). The two mutants where MTSL tags are attached to the N- and C-termini display very similar R_g and D values. The fact that D varies so little from residue to residue and from mutant to mutant is rather important. Indeed, if diffusion-based models are used to interpret the PRE data in (strongly) disordered proteins, it is important to know that a single D value can be used in such analyses. This is described in detail in the next section.

Fitting Experimental PREs Using the Ullman−Podkorytov Model. In addition to the MD-based interpretation, PRE rates can also be interpreted using a simple model which has been recently developed for random-coil-like proteins.[55] In brief, the paramagnetic center and $^1H^N$ are modeled as two hard spheres connected by a weightless, inextensible, and perfectly flexible string (Figure 9A). These two spheres diffuse relative to each other subject to the following constraints: (i) reflecting boundary conditions at d_0 and L and (ii) harmonic potential $U(r) = U_0 + k_BT(3r^2/2l^2)$ which enforces Gaussian behavior of the protein chain[143] (see Figure 9B,C). The problem can then be cast in terms of a Smoluchowski equation for a particle diffusing in an effective potential and solved in a "semianalytical" fashion, resulting in a relatively simple and computationally efficient formula for the PRE rates.[55] We refer to this model as the Ullman−Podkorytov model, named after its originators.[55,69]

The parameters involved in the Ullman−Podkorytov model are as follows. l is the root-mean-square distance between the ends of the chain (i.e., in this case between the pair of spins), $l = \sigma(\infty)/\sqrt{2}$. It can be in turn expressed as $l = b(|n_{HN} - n_{MTSL}|+n_{tag})^p$, where b is the effective monomer length; n_{HN} and n_{MTSL} are the residue numbers of the amino acids carrying the $^1H^N$ reporter spin and the MTSL tag; n_{tag}(=3) accounts for the length of the tag per se: and p is the exponent which assumes the value 0.5 for the ideal Gaussian chain.[55] The value of b can be obtained using the relationship $R_g = (b/\sqrt{6})N^p$, where N is a number of residues in the random-coil-like protein. Analysis of the large body of experimental data on chemically denatured proteins led to b = 4.72 Å and p = 0.597;[144,145] in what follows, we use these values to evaluate the Ullman−Podkorytov formula.[55] The term U_0 in the expression for $U(r)$ is introduced with the sole purpose to normalize the pair correlation function, $P(r) = \exp(-U(r)/k_BT)$; this parameter does not enter into the calculation of the PRE rates. The minimum separation distance d_0 has been estimated by means of the conformational modeling using TraDes and Xplor-NIH (CHARMM22)[146] and found to be d_0 = 4 Å.[55] The maximum separation distance is calculated according to the formula $L=(|n_{HN} - n_{MTSL}|\cdot 2.55 + 12)$ Å which describes the fully extended chain and the MTSL tag. The exact value of L is unimportant because fully extended conformations are statistically unlikely, i.e., because $U(r)$ at $r = L$ is prohibitively high (see Figure 9C).

The example of the harmonic potential $U(r)$ with the reflecting boundaries at $r = d_0$ and $r = L$ is shown in Figure 9C. For comparison, we also show the corresponding mean-force potential, $U_{pmf}(r) = -k_BT \ln P(r)$, computed on the basis of one of the microsecond MD trajectories recorded in this study (Figure 9D). Overall, there is a reasonable agreement between the two profiles. The most significant deviation occurs at small separations r. Clearly, this discrepancy reflects the shortcomings of the harmonic potential model, which neglects the excluded volume effects (these effects become increasingly important when spin label and H^N atom come close to each other). This is the price for the computational simplicity afforded by the Ullman−Podkorytov model. (Note that we have also developed a numeric method for calculation of PREs where the harmonic potential $U(r)$ is replaced with $U_{pmf}(r)$.[55]).

The Ullman−Podkorytov model parametrized as described above contains a single floating parameter—coefficient of segmental diffusion D. We have used this opportunity to fit the experimental PRE data directly to the Ullman−Podkorytov model and thus determine the value of D. For this purpose, we have combined the PRE data from Q2C-, D32C-, and R74C-MTSL samples and plotted them as a function of $|n_{HN} - n_{MTSL}|$ (Figure 10). The fitting was restricted to the points for which $|n_{HN} - n_{MTSL}| > 20$ (red circles in the plot). Indeed, the model is formulated for the ideal fully flexible chain and is supposed to work well only when there is a sufficient number of monomeric

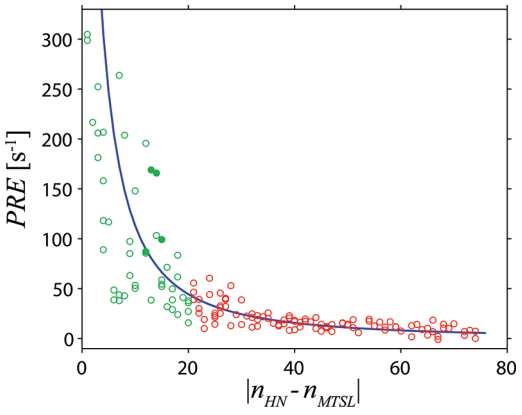


Figure 10. Experimental and theoretical PRE rates in denatured ubiquitin as a function of the primary sequence separation between the MTSL and $^1H^N$ sites. The experimental data are from Q2C-, D32C-, and R74C-MTSL samples, represented by red circles (for sites with $|n_{HN} - n_{MTSL}| \leq 20$) and green circles ($|n_{HN} - n_{MTSL}| \leq 20$); filled circles correspond specifically to residues 14−17 in Q2C-MTSL (see text). Blue curve represents the fitting of the experimental data (red points only) to the Ullman−Podkorytov model using the mutual diffusion coefficient D as a sole fitting parameters. Other parameters are assigned values as indicated in the text.

units (amino acids) between the MTSL site and $^1H^N$ spin. The best-fit curve obtained with the Ullman−Podkorytov model is shown in blue. The fitted value of the segmental diffusion coefficient D is 0.45×10^{-6} cm^2 s^{-1}, in good agreement with the MD-based values (average D in the range from 0.49 to 0.55 $\times 10^{-6}$ cm^2 s^{-1}, depending on the placement of the MTSL label). Note that all of these results pertain to the temperature of 5 °C.

The quality of the Ullman−Podkorytov fit in Figure 10 is good, especially considering that there is only one adjustable parameter involved. The tail of the dependence (red circles) is reproduced very well. Also, the initial portion which is not a part of the fit (green circles) shows reasonable overall agreement. With regard to the initial part of the graph, one should keep in mind that the results are affected by stiffness of the polypeptide chain—it prevents the MTSL tag and H^N atom from coming close to each other, thus undercutting the PRE rates. The Ullman−Podkorytov model does not take the stiffness effect into consideration, assuming instead that the chain is infinitely flexible. Note also that some of the experimental PREs are higher than expected, e.g., the PREs from residues 14−17 in Q2C-MTSL which reflect the presence of the transiently populated β-hairpin (filled green circles in Figure 10; see above for discussion). One should also bear in mind that the PRE rates in excess of 100 s^{-1} suffer from systematic errors and are less accurate.[55]

It is important to realize that the constant, coordinate-independent D is essentially a model concept.[147] Furthermore, the accuracy of D is limited by the uncertainties in the model parameters. For instance, the outcome of the analysis depends on the distance of the nearest approach d_0. Note that the exact localization of the unpaired electron spin remains a bit of an open question. In vacuum, the unpaired electron is distributed along the nitroxide NO bond; however, in the polar solvent (water) it is strongly shifted toward the nitrogen atom.[132,148−152] If we assume for a moment that the unpaired electron is localized in the middle of the NO bond instead of the nitrogen atom, then estimated d_0 drops to 3.5 Å. Consequently, the value of D determined by the Ullman−Podkorytov analyses increases to 0.53×10^{-6} cm^2 s^{-1}.

Also the choice of $|n_{HN} - n_{MTSL}| > 20$ cutoff in the above analysis is somewhat arbitrary. It is clear that the cutoff should be greater than the persistence length λ, i.e., the characteristic length over which the polypeptide chain retains some stiffness. For denatured proteins, λ has been estimated to fall in the range from 5 to 10 amino acids.[32,153,154] Accordingly, we tested the $|n_{HN} - n_{MTSL}|$ cutoff values between 10 and 30 residues. It was found that the choice of cutoff has relatively little effect on the outcome of the PRE fitting, with D value fluctuating between 0.42 and 0.50×10^{-6} cm^2 s^{-1}. This and the previous example give an idea about the uncertainty that is built into the current interpretation procedure.

CONCLUSION

MD simulations have been widely used to analyze dynamic conformational equilibria of folded proteins, especially in relation to the NMR observables. In contrast, little has been done in the area of disordered proteins, where the conformational phase space is larger and the sampling presents a serious problem. In this paper, we demonstrate that the latest advances in computation technology make it possible to overcome this limitation.

To test the new approach, we have chosen a model system which is known to be highly disordered: ubiquitin in solution with 8 M urea, pH 2. To create a MD model for denatured ubiquitin, we recorded four high-temperature trajectories in implicit solvent. The nominal length of each trajectory, representing the wt protein and three MTSL-tagged mutants, was 1 μs. To recalibrate the model with regard to temperature (MD simulations, 500 K; experimental measurements, 278 K) we applied the scaling factor to the MD time axis. The effective length of each trajectory thereby increased to 311 μs, demonstrating the benefits of high-temperature simulations for statistical sampling.

The above approach to the construction of the MD model is clearly empirical. A rigorous criticism can be directed at it on several levels. First, thermal denaturation (500 K) is not necessarily equivalent to chemical denaturation (278 K, 8 M urea). Second, general-use force fields and solvation models are parametrized to reproduce the properties of folded proteins at room temperature. Strictly speaking, they are not designed for use with disordered proteins at high temperature. Third, the application of the uniform time scaling factor ($\alpha = 311$) seems like a naive idea. Indeed, one may expect that different motional modes scale differently with temperature.

While admitting that these objections are fundamentally sound, we maintain that the present model is nevertheless useful and promising. First and foremost, the justification is provided by the extensive comparison with experimental data. Toward this end, we have compared a number of the simulated and experimental parameters, including radii of gyration, ^{15}N relaxation rates, and multiple sets of paramagnetic relaxation enhancement data. In each case a near-quantitative agreement has been observed, thus suggesting that the MD-based model is reasonable. So long as the model successfully reproduces a (sufficiently broad and diverse) set of experimental data, such a model can be deemed worthwhile irrespective of its humble origins.

Our calibration procedure is aimed primarily at conformational dynamics dominated by large-amplitude jumps $\phi(t)$ and $\psi(t)$. These jumps ultimately control both local and global conformational dynamics and therefore determine ^{15}N relaxation rates, as well as the PREs. However, other experimental parameters, such as $^{1}H^{N}-^{15}N$ heteronuclear NOE, are also sensitive to very fast forms of local dynamics, which deserve a special discussion.

Ultrafast local motions are prominent in MD simulations of various proteins using different force fields.[155,156] These motions mainly come from two sources: in- and out-of-plane bending of the NH bond and fast fluctuations involving ϕ, ψ. Ultrafast motions are responsible for the initial steep drop in the $^{1}H^{N}-^{15}N$ correlation functions; the corresponding correlation time is ca. 0.15 ps and does not change with temperature. After rescaling, this correlation time translates into ca. 44 ps (see Figure 2 and associated discussion). How reasonable is this value? As it turns out, model-free analyses of ^{15}N relaxation data, including data taken at multiple fields, consistently produce τ_{fast} values in the range from 10 to 100 ps.[157−159] Hence, our rescaled MD model is consistent with the Lipari−Szabo description of fast local dynamics. In turn, this leads to fairly accurate predictions of the heteronuclear NOEs (see Figure 4). Keep in mind, however, that τ_{fast} values derived from the Lipari−Szabo treatment should be viewed as effective values that do not necessarily reproduce the details of the actual correlation functions.[155]

High-temperature simulations have been used on many occasions before to study protein unfolding and, in particular, chemical denaturation.[160−163] While this approach is largely empirical, it is considerably more realistic than many simple models that remain in wide use in protein science. So long as the limitations of high-temperature simulations are clearly recognized, this approach remains a legitimate and useful research tool.[95] In the future, the methodology described in this paper will be used in conjunction with more realistic versions of MD simulations. In fact, the computational resources already exist that allow one to generate ∼100 μs explicit-solvent trajectories of small unfolded proteins at room temperature.[164]

The properly validated MD models can be used to gain insight into those aspects of molecular structure/dynamics that cannot be readily accessed experimentally. In particular, we focused on the coefficient of segmental diffusion D, characterizing the relative translational motion of the two segments of flexible polypeptide chains (cf. end-to-end dynamics). The calculations using MD data showed that D values are nearly uniform across the peptide chain, averaging to $0.49-0.55 \times 10^{-6}$ cm^2 s^{-1}. To validate this result, we directly analyzed the set of experimental PRE rates obtained from three MTSL-tagged mutants of ubiquitin. The analysis was conducted using the Ullman−Podkorytov model, which treats the relevant dynamics as diffusion of two particles in a harmonic potential. The extracted value of the segmental diffusion coefficient, 0.45×10^{-6} cm^2 s^{-1}, is in good agreement with the MD results. We feel that this determination of D is more definitive than the previous estimations using resonance energy transfer methods, 0.2 to 6×10^{-6} cm^2 s^{-1}.

As indicated by this example, long MD trajectories validated against NMR and other experimental data provide a valuable tool for studies of disordered proteins. In the future, such trajectories should efficiently augment the existing structural models, i.e., static conformational ensembles.

MATERIALS AND METHODS

Sample Preparation. The plasmid of human ubiquitin was supplied by Rachel Klevit laboratory[165] through Addgene repository (Addgene plasmid 12647). The expression and purification protocol was adapted from the work by Lazar et al.[166] Three single-cystein mutants, Q2C, D32C, and R74C, were prepared and labeled with MTSL as described earlier.[55] In addition, MTSL labeling of the freshly prepared samples without the DTT pretreatment also proved to be successful. The sample conditions were 0.33 mM protein, 8 M urea, 90% H_2O−10% D_2O, pH 2.0. The internal standard, ^{15}N *N*-acetyl−glycine, has been added to the solution to match the intensity of the signals from the oxidized and reduced samples. As described in our previous report, the accuracy of the PRE measurements tends to be compromised by the presence of diamagnetic species in the presumed paramagnetic sample.[55] The content of these unwanted species can be estimated directly from the HSQC spectra—specifically from the intensity of weak peaks corresponding to the residues close to the MTSL site, $|n_{HN} - n_{MTSL}| \leq 2$. In our previous drkN and ubiquitin preparations, the content of diamagnetic species was on the order of 10%.[55] In this study, we were able to reduce it to approximately 1−2%, although currently we do not have an explanation as to the source of this improvement.

NMR Measurements. All NMR measurements were conducted at 5 °C using a 600 MHz Varian Inova spectrometer. The spectral assignment was obtained from the publication of Peti et al.[47] and confirmed by the HNCACB experiment.[167] The ^{15}N relaxation measurements were carried out using the updated versions of the standard relaxation experiments,[168−170] including the recently corrected heteronuclear NOE sequence.[171] To determine the PRE rates, we have employed the conventional (unenhanced) HSQC sequence with WATERGATE water suppression scheme.[172] The use of low-power rectangular water flip-back pulses ensures that $^{1}H^{N}$ magnetization remains in the transverse plane and therefore experiences the PRE effect throughout the duration of the final INEPT element. The PRE rates were obtained from the ratio of the peak volumes from the oxidized and reduced samples.[55]

Molecular Dynamics Simulations. MD simulations were performed using the Amber11 ff99SB package which supports the GPU computing.[88,173] The initial random conformation of ubiquitin was generated using the facilities of Amber. Asp and Glu residues were replaced with their protonated equivalents, ASH and GLH, as appropriate under the conditions of our experimental study, pH 2.[174] The chain was subsequently energy-minimized using steepest descent and conjugate gradient methods (500 steps each). The simulations were conducted at 500 K using Langevin dynamics with the collision frequency 1 ps^{-1} (low collision frequency appears to be appropriate for high-temperature simulations;[175] ultimately, the validity of this choice is confirmed by the comparison with experimental data). The implicit solvent was represented by the optimized version of the pairwise generalized Born model,[80] igb = 1, with modified Bondi radii.[176] During the simulations, bonds involving hydrogen atoms were constrained by SHAKE algorithm with tolerance 0.00001 Å; the nonbonded cutoff was set to 999 Å. The integration step was 1 fs. The first 1 ns of the trajectory was treated as the equilibration stage; subsequently, the protein coordinates were saved every 0.2 ps (sufficient for accurate PRE and ^{15}N relaxation rate calculations). The force-field parameters of the MTSL moiety were taken from the recent study by Sezer et al.[97] and translated for use with Amber using the program CHAMBER (see SI for detailed information).[177] The correctness of the translation was checked by comparison of the individual energy terms computed using CHARMM[178] and Amber. The simulations were performed on a GPU workstation equipped with four GeForce GTX480 cards using the CUDA version of the pmemd program

(assembled by Electronics Nexus). The production rate was 72 ns per day on a single GPU card. The trajectories with nominal length 1 μs were recorded for wild-type ubiquitin and three MTSL-tagged mutants, Q2C-, D32C-, and R74C-MTSL.

Data Analysis of MD Simulations. The radius of gyration, ^{15}N relaxation rates, and PREs were calculated using the C++/Matlab package written in-house (available upon request).[55,127] The correlation functions $g_{\mathrm{reorient}}(\tau)$ and $g(\tau)$, as well as the mean square displacement $\sigma(\tau)$, were evaluated using the previously described nonlinear sampling scheme (fine sampling at short τ, sparse sampling at long τ).[127] The correlation functions were subsequently fitted with six-exponential curves, as originally proposed by Bremi et al.[126] For a handful of residues, the fits included low-amplitude components with very long decay times (e.g., amplitude 0.003%, decay time 2.5 ms). Such terms are clearly artifactual, yet they can significantly contribute to the spectral density $J(0)$. To suppress these components, we multiplied $g_{\mathrm{reorient}}(\tau)$ and $g(\tau)$ by the filtering function $\exp(-\tau/\tau_{\mathrm{mask}})$, with τ_{mask} set to 311 ns. Since τ_{mask} is much longer than any of the actual correlation times observed in the system (see Figures 2 and 5), this step does not interfere with the computation of the physically meaningful R_2 and PRE rates.

In calculating the ^{15}N relaxation rates, we have assumed that the proton−nitrogen distance r_{NH} is 1.02 Å, the magnitude of the nitrogen CSA $\Delta\sigma_{\mathrm{N}}$ is −172 ppm; and the unique axis of the CSA tensor coincides with the NH bond. It has been recently determined that the effective r_{NH} value averaged over the bond stretching is 1.015 Å. On the other hand, the r_{NH} value averaged over both stretching *and* zero-point librations is 1.04 Å.[179] The former value, which is close to the standard setting 1.02 Å, is appropriate for the regular MD simulations, whereas the latter can be recommended for the TAD MD simulations. Note finally that the assumption of the uniform CSA value is especially well justified in the case of the denatured protein.

■ ASSOCIATED CONTENT

Supporting Information. Excerpts from the MD trajectories, rendered in a form of .avi movies: Q2C- and D32C-MTSL simulations (1% of the full trajectory, 10 000 frames per movie), expanded view of the Q2C-MTSL simulation illustrating the formation of a hydrogen bond between MTSL and the amide group of R72 (0.04% of the full trajectory, 2000 frames). Correlation between experimental and calculated $^{13}C^{\alpha}$, $^{13}C^{\beta}$ chemical shifts. Content of transient secondary structure in denatured ubiquitin. Validation of the Gillespie−Shortle treatment of the PRE data. Force-field and topology parameters for residue CYS-MTSL[97] translated for use with Amber. Complete refs 71, 173, and 178. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
nikolai@purdue.edu

■ ACKNOWLEDGMENT

This study was funded through the NSF grant MCB-044563. We thank Ivan Podkorytov for reading the manuscript of this paper.

■ REFERENCES

(1) Williams, R. J. *Biochem. Soc. Trans.* **1978**, *6*, 1123.
(2) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins: Struct. Funct. Genet.* **2000**, *41*, 415.
(3) Romero, P.; Obradovic, Z.; Li, X. H.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins: Struct. Funct. Genet.* **2001**, *42*, 38.
(4) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Brown, C. J.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2005**, *44*, 1989.
(5) Uversky, V. N.; Dunker, A. K. *Science* **2008**, *322*, 1340.
(6) van den Berg, B.; Ellis, R. J.; Dobson, C. M. *EMBO J.* **1999**, *18*, 6927.
(7) Dedmon, M. M.; Patel, C. N.; Young, G. B.; Pielak, G. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12681.
(8) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* **2004**, *337*, 635.
(9) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. *J. Mol. Biol.* **2002**, *323*, 573.
(10) Xie, H. B.; Vucetic, S.; Iakoucheva, L. M.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N.; Obradovic, Z. *J. Proteome Res.* **2007**, *6*, 1882.
(11) Dames, S. A.; Martinez-Yamout, M.; De Guzman, R. N.; Dyson, H. J.; Wright, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5271.
(12) Bochkareva, E.; Kaustov, L.; Ayed, A.; Yi, G. S.; Lu, Y.; Pineda-Lucena, A.; Liao, J. C. C.; Okorokov, A. L.; Milner, J.; Arrowsmith, C. H.; Bochkarev, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15412.
(13) Mohan, A.; Oldfield, C. J.; Radivojac, P.; Vacic, V.; Cortese, M. S.; Dunker, A. K.; Uversky, V. N. *J. Mol. Biol.* **2006**, *362*, 1043.
(14) Dancheck, B.; Nairn, A. C.; Peti, W. *Biochemistry* **2008**, *47*, 12346.
(15) Ragusa, M. J.; Dancheck, B.; Critton, D. A.; Nairn, A. C.; Page, R.; Peti, W. *Nat. Struct. Mol. Biol.* **2010**, *17*, 459.
(16) Gunasekaran, K.; Tsai, C. J.; Kumar, S.; Zanuy, D.; Nussinov, R. *Trends Biochem. Sci.* **2003**, *28*, 81.
(17) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys.* **2008**, *37*, 215.
(18) Uversky, V. N. *Front. Biosci.* **2009**, *14*, 5188.
(19) Cheng, Y.; LeGall, T.; Oldfield, C. J.; Mueller, J. P.; Van, Y. Y. J.; Romero, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *Trends Biotechnol.* **2006**, *24*, 435.
(20) Necula, M.; Kayed, R.; Milton, S.; Glabe, C. G. *J. Biol. Chem.* **2007**, *282*, 10311.
(21) Crowe, A.; Ballatore, C.; Hyde, E.; Trojanowski, J. Q.; Lee, V. M. Y. *Biochem. Biophys. Res. Commun.* **2007**, *358*, 1.
(22) Erkizan, H. V.; Kong, Y. L.; Merchant, M.; Schlottmann, S.; Barber-Rotenberg, J. S.; Yuan, L. S.; Abaan, O. D.; Chou, T. H.; Dakshanamurthy, S.; Brown, M. L.; Uren, A.; Toretsky, J. A. *Nat. Med.* **2009**, *15*, 750.
(23) Brüschweiler, R.; Blackledge, M.; Ernst, R. R. *J. Biomol. NMR* **1991**, *1*, 3.
(24) Groth, M.; Malicka, J.; Czaplewski, C.; Oldziej, S.; Lankiewicz, L.; Wiczk, W.; Liwo, A. *J. Biomol. NMR* **1999**, *15*, 315.
(25) Nikiforovich, G. V.; Prakash, O.; Gehrig, C. A.; Hruby, V. J. *J. Am. Chem. Soc.* **1993**, *115*, 3399.
(26) Gippert, G. P.; Wright, P. E.; Case, D. A. *J. Biomol. NMR* **1998**, *11*, 241.
(27) Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J. Phys. Chem.* **1996**, *100*, 2661.
(28) Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002.
(29) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099.
(30) Choy, W. Y.; Forman-Kay, J. D. *J. Mol. Biol.* **2001**, *308*, 1011.
(31) Song, J.; Guo, L. W.; Muradov, H.; Artemyev, N. O.; Ruoho, A. E.; Markley, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1505.
(32) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977.
(33) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G. W.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 8407.
(34) Schwarzinger, S.; Wright, P. E.; Dyson, H. J. *Biochemistry* **2002**, *41*, 12681.
(35) Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391*, 359.
(36) Torda, A. E.; Brunne, R. M.; Huber, T.; Kessler, H.; van Gunsteren, W. F. *J. Biomol. NMR* **1993**, *3*, 55.

(37) Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. *J. Biomol. NMR* **1994**, *4*, 143.
(38) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923.
(39) Tang, C.; Schwieters, C. D.; Clore, G. M. *Nature* **2007**, *449*, 1078.
(40) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128.
(41) Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 18314.
(42) Huang, J. R.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 694.
(43) Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. *Structure* **2010**, *18*, 923.
(44) Lange, O. F.; Lakomek, N. A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471.
(45) Esteban-Martin, S.; Fenwick, R. B.; Salvatella, X. *J. Am. Chem. Soc.* **2010**, *132*, 4626.
(46) Case, D. A. *Acc. Chem. Res.* **2002**, *35*, 325.
(47) Peti, W.; Smith, L. J.; Redfield, C.; Schwalbe, H. *J. Biomol. NMR* **2001**, *19*, 153.
(48) Wirmer, J.; Schwalbe, H. *J. Biomol. NMR* **2002**, *23*, 47.
(49) Peti, W.; Hennig, M.; Smith, L. J.; Schwalbe, H. *J. Am. Chem. Soc.* **2000**, *122*, 12017.
(50) Wirmer, J.; Peti, W.; Schwalbe, H. *J. Biomol. NMR* **2006**, *35*, 175.
(51) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799.
(52) Meier, S.; Strohmeier, M.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2007**, *129*, 754.
(53) Berliner, L. J.; Grunwald, J.; Hankovszky, H. O.; Hideg, K. *Anal. Biochem.* **1982**, *119*, 450.
(54) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 158.
(55) Xue, Y.; Podkorytov, I. S.; Rao, D. K.; Benjamin, N.; Sun, H. L.; Skrynnikov, N. R. *Protein Sci.* **2009**, *18*, 1401.
(56) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170.
(57) Teilum, K.; Kragelund, B. B.; Poulsen, F. M. *J. Mol. Biol.* **2002**, *324*, 349.
(58) Lindorff-Larsen, K.; Kristjansdottir, S.; Teilum, K.; Fieber, W.; Dobson, C. M.; Poulsen, F. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 3291.
(59) Felitsky, D. J.; Lietzow, M. A.; Dyson, H. J.; Wright, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6278.
(60) Buckler, D. R.; Haas, E.; Scheraga, H. A. *Biochemistry* **1995**, *34*, 15965.
(61) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11615.
(62) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220.
(63) Moglich, A.; Joder, K.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12394.
(64) Möglich, A.; Joder, K.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6787.
(65) Soranno, A.; Longhi, R.; Bellini, T.; Buscaglia, M. *Biophys. J.* **2009**, *96*, 1515.
(66) Lee, J. C.; Gray, H. B.; Winkler, J. R. *J. Am. Chem. Soc.* **2005**, *127*, 16388.
(67) Nettels, D.; Gopich, I. V.; Hoffmann, A.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2655.
(68) Singh, V. R.; Kopka, M.; Chen, Y.; Wedemeyer, W. J.; Lapidus, L. J. *Biochemistry* **2007**, *46*, 10046.
(69) Ullman, R. *J. Chem. Phys.* **1965**, *43*, 3161.
(70) Caflisch, A.; Karplus, M. *Struct. Fold. Des.* **1999**, *7*, 477.
(71) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586.
(72) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400.
(73) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(74) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(75) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596.
(76) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701.
(77) Pappu, R. V.; Hart, R. K.; Ponder, J. W. *J. Phys. Chem. B* **1998**, *102*, 9725.
(78) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
(79) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578.
(80) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.
(81) Im, W. P.; Lee, M. S.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1691.
(82) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227.
(83) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct. Funct. Genet.* **2002**, *46*, 24.
(84) Chen, J. H.; Im, W. P.; Brooks, C. L. *J. Am. Chem. Soc.* **2006**, *128*, 3728.
(85) Best, R. B.; Buchete, N. V.; Hummer, G. *Biophys. J.* **2008**, *95*, L7.
(86) Penev, E.; Ireta, J.; Shea, J. E. *J. Phys. Chem. B* **2008**, *112*, 6872.
(87) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004.
(88) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
(89) Showalter, S. A.; Bruschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961.
(90) Aliev, A. E.; Courtier-Murias, D. *J. Phys. Chem. B* **2010**, *114*, 12358.
(91) Cerutti, D. S.; Freddolino, P. L.; Duke, R. E.; Case, D. A. *J. Phys. Chem. B* **2010**, *114*, 12811.
(92) Li, D. W.; Bruschweiler, R. *J. Phys. Chem. Lett.* **2010**, *1*, 246.
(93) Wu, X. W.; Brooks, B. R. *Chem. Phys. Lett.* **2003**, *381*, 512.
(94) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863.
(95) Day, R.; Bennion, B. J.; Ham, S.; Daggett, V. *J. Mol. Biol.* **2002**, *322*, 189.
(96) Voelz, V. A.; Singh, V. R.; Wedemeyer, W. J.; Lapidus, L. J.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 4702.
(97) Sezer, D.; Freed, J. H.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 5755.
(98) Gabel, F.; Jensen, M. R.; Zaccai, G.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 8769.
(99) Nettels, D.; Muller-Spath, S.; Kuster, F.; Hofmann, H.; Haenni, D.; Ruegger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B.; Gast, K.; Best, R. B.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 20740.
(100) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76.
(101) Neuweiler, H.; Johnson, C. M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18569.
(102) Nishimura, C.; Lietzow, M. A.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2005**, *351*, 383.
(103) Cho, M. K.; Kim, H. Y.; Bernado, P.; Fernandez, C. O.; Blackledge, M.; Zweckstetter, M. *J. Am. Chem. Soc.* **2007**, *129*, 3032.
(104) Alexandrescu, A. T.; Shortle, D. *J. Mol. Biol.* **1994**, *242*, 527.
(105) Yao, J.; Chung, J.; Eliezer, D.; Wright, P. E.; Dyson, H. J. *Biochemistry* **2001**, *40*, 3561.
(106) Ochsenbein, F.; Guerois, R.; Neumann, J. M.; Sanson, A.; Guittet, E.; van Heijenoort, C. *J. Biomol. NMR* **2001**, *19*, 3.
(107) Bussell, R.; Eliezer, D. *J. Biol. Chem.* **2001**, *276*, 45996.
(108) Cao, W.; Bracken, C.; Kallenbach, N. R.; Lu, M. *Protein Sci.* **2004**, *13*, 177.
(109) Zhang, X. C.; Xu, Y. Q.; Zhang, J. H.; Wu, J. H.; Shi, Y. Y. *Biochemistry* **2005**, *44*, 8117.
(110) Shojania, S.; O'Neil, J. D. *J. Biol. Chem.* **2006**, *281*, 8347.
(111) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719.

(112) Le Duff, C. S.; Whittaker, S. B. M.; Radford, S. E.; Moore, G. R. *J. Mol. Biol.* **2006**, *364*, 824.
(113) Kumar, D.; Chugh, J.; Sharma, S.; Hosur, R. V. *Proteins* **2009**, *76*, 387.
(114) Mishima, T.; Ohkuri, T.; Monji, A.; Kanemaru, T.; Abe, Y.; Ueda, T. *J. Mol. Biol.* **2009**, *392*, 1033.
(115) Pujato, M.; Bracken, C.; Mancusso, R.; Cataldi, M.; Tasayco, M. L. *Biophys. J.* **2005**, *89*, 3293.
(116) Cho, J. H.; Raleigh, D. P. *J. Mol. Biol.* **2005**, *353*, 174.
(117) Bertoncini, C. W.; Jung, Y. S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1430.
(118) Palmer, A. G.; Case, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 9059.
(119) Nicholas, M. P.; Eryilmaz, E.; Ferrage, F.; Cowburn, D.; Ghose, R. *Prog. NMR Spectrosc.* **2010**, *57*, 111.
(120) Prompers, J. J.; Bruschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 7305.
(121) Prompers, J. J.; Bruschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522.
(122) Farrow, N. A.; Zhang, O. W.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1997**, *36*, 2390.
(123) Palmer, A. G. *Annu. Rev. Biophys. Biomol.* **2001**, *30*, 129.
(124) Brüschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. *J. Am. Chem. Soc.* **1992**, *114*, 2289.
(125) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(126) Bremi, T.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1997**, *119*, 4272.
(127) Xue, Y.; Pavlova, M. S.; Ryabov, Y. E.; Reif, B.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2007**, *129*, 6827.
(128) Showalter, S. A.; Johnson, E.; Rance, M.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 14146.
(129) Lietzow, M. A.; Jamin, M.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2002**, *322*, 655.
(130) Sung, Y. H.; Eliezer, D. *J. Mol. Biol.* **2007**, *372*, 689.
(131) Yi, Q.; Scalley-Kim, M. L.; Alm, E. J.; Baker, D. *J. Mol. Biol.* **2000**, *299*, 1341.
(132) Berliner, L. J. *Spin labeling: theory and applications*; Academic Press: New York, 1976.
(133) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.
(134) Kristjansdottir, S.; Lindorff-Larsen, K.; Fieber, W.; Dobson, C. M.; Vendruscolo, M.; Poulsen, F. M. *J. Mol. Biol.* **2005**, *347*, 1053.
(135) Neal, S.; Nip, A. M.; Zhang, H. Y.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215.
(136) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311.
(137) Marsh, J. A.; Singh, V. K.; Jia, Z. C.; Forman-Kay, J. D. *Protein Sci.* **2006**, *15*, 2795.
(138) Haile, J. M. *Molecular Dynamics Simulation*; John Wiley & Sons, Inc.: New York, 1992.
(139) Meier, K.; Laesecke, A.; Kabelac, S. *Int. J. Thermophys.* **2001**, *22*, 161.
(140) Sullivan, D. C.; Kuntz, I. D. *J. Phys. Chem. B* **2002**, *106*, 3255.
(141) Fierz, B.; Kiefhaber, T. *J. Am. Chem. Soc.* **2007**, *129*, 672.
(142) Urie, K. G.; Angulo, D.; Lee, J. C.; Kozak, J. J.; Gray, H. B.; Winkler, J. R. *J. Phys. Chem. B* **2009**, *113*, 522.
(143) Doi, M.; Edwards, S. F. *The theory of polymer dynamics*; Clarendon Press: Oxford, 1986.
(144) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12491.
(145) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14475.
(146) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65.
(147) Yang, S. C.; Onuchic, J. N.; Levine, H. *J. Chem. Phys.* 2006, *125*, 054910.
(148) Plato, M.; Steinhoff, H. J.; Wegener, C.; Torring, J. T.; Savitsky, A.; Mobius, K. *Mol. Phys.* **2002**, *100*, 3711.
(149) Improta, R.; Barone, V. *Chem. Rev.* **2004**, *104*, 1231.
(150) Rastrelli, F.; Bagno, A. *Chem.—Eur. J.* **2009**, *15*, 7990.
(151) Ikryannikova, L. N.; Ustynyuk, L. Y.; Tikhonov, A. N. *Magn. Reson. Chem.* **2010**, *48*, 337.
(152) Klug, C. S.; Feix, J. B. *Biophysical Tools for Biologists: In Vitro Techniques*; Elsevier Academic Press, Inc.: San Diego, CA, 2008; Vol. 84, p 617.
(153) Gast, K.; Damaschun, H.; Eckert, K.; Schulzeforster, K.; Maurer, H. R.; Mullerfrohne, M.; Zirwer, D.; Czarnecki, J.; Damaschun, G. *Biochemistry* **1995**, *34*, 13211.
(154) Tran, H. T.; Wang, X. L.; Pappu, R. V. *Biochemistry* **2005**, *44*, 11369.
(155) Pfeiffer, S.; Fushman, D.; Cowburn, D. *J. Am. Chem. Soc.* **2001**, *123*, 3021.
(156) Nederveen, A. J.; Bonvin, A. M. J. J. *J. Chem. Theory Comput.* **2005**, *1*, 363.
(157) Butterwick, J. A.; Loria, J. P.; Astrof, N. S.; Kroenke, C. D.; Cole, R.; Rance, M.; Palmer, A. G. *J. Mol. Biol.* **2004**, *339*, 855.
(158) Ye, J. Q.; Mayer, K. L.; Mayer, M. R.; Stone, M. J. *Biochemistry* **2001**, *40*, 7820.
(159) Vugmeyster, L.; Trott, O.; McKnight, C. J.; Raleigh, D. P.; Palmer, A. G. *J. Mol. Biol.* **2002**, *320*, 841.
(160) Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5142.
(161) Rocco, A. G.; Mollica, L.; Ricchiuto, P.; Baptista, A. M.; Gianazza, E.; Eberini, I. *Biophys. J.* **2008**, *94*, 2241.
(162) Hua, L.; Zhou, R. H.; Thirumalai, D.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16928.
(163) Canchi, D. R.; Paschek, D.; Garcia, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 2338.
(164) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; Wriggers, W. *Science* **2010**, *330*, 341.
(165) Brzovic, P. S.; Lissounov, A.; Christensen, D. E.; Hoyt, D. W.; Klevit, R. E. *Mol. Cell* **2006**, *21*, 873.
(166) Lazar, G. A.; Desjarlais, J. R.; Handel, T. M. *Protein Sci.* **1997**, *6*, 1167.
(167) Muhandiram, D. R.; Kay, L. E. *J. Magn. Reson. Ser. B* **1994**, *103*, 203.
(168) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Formankay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984.
(169) Korzhnev, D. M.; Skrynnikov, N. R.; Millet, O.; Torchia, D. A.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 10743.
(170) Hansen, D. F.; Kay, L. E. *J. Biomol. NMR* **2007**, *37*, 245.
(171) Ferrage, F.; Cowburn, D.; Ghose, R. *J. Am. Chem. Soc.* **2009**, *131*, 6048.
(172) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661.
(173) Case, D. A. et al. *Amber 11 Users' Manual*; University of California, San Francisco, 2010.
(174) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins* **2008**, *73*, 765.
(175) Hamelberg, D.; Shen, T. Y.; McCammon, J. A. *J. Chem. Phys.* 2006, *125*, 094905.
(176) Tsui, V.; Case, D. A. *Biopolymers* **2000**, *56*, 275.
(177) Crowley, M. F.; Williamson, M. J.; Walker, R. C. *Int. J. Quantum Chem.* **2009**, *109*, 3767.
(178) Brooks, B. R.; et al. *J. Comput. Chem.* **2009**, *30*, 1545.
(179) Yao, L. S.; Grishaev, A.; Cornilescu, G.; Bax, A. *J. Am. Chem. Soc.* **2010**, *132*, 4295.

Supporting Information

Motion of disordered polypeptide chain as studied by PREs, ^{15}N relaxation, and MD simulations: how fast is segmental diffusion in denatured ubiquitin?

Yi Xue and Nikolai Skrynnikov*

submitted 20 February 2011

in revised form 30 April 2011

Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette IN 47907-2084, USA

* Corresponding author (nikolai@purdue.edu)

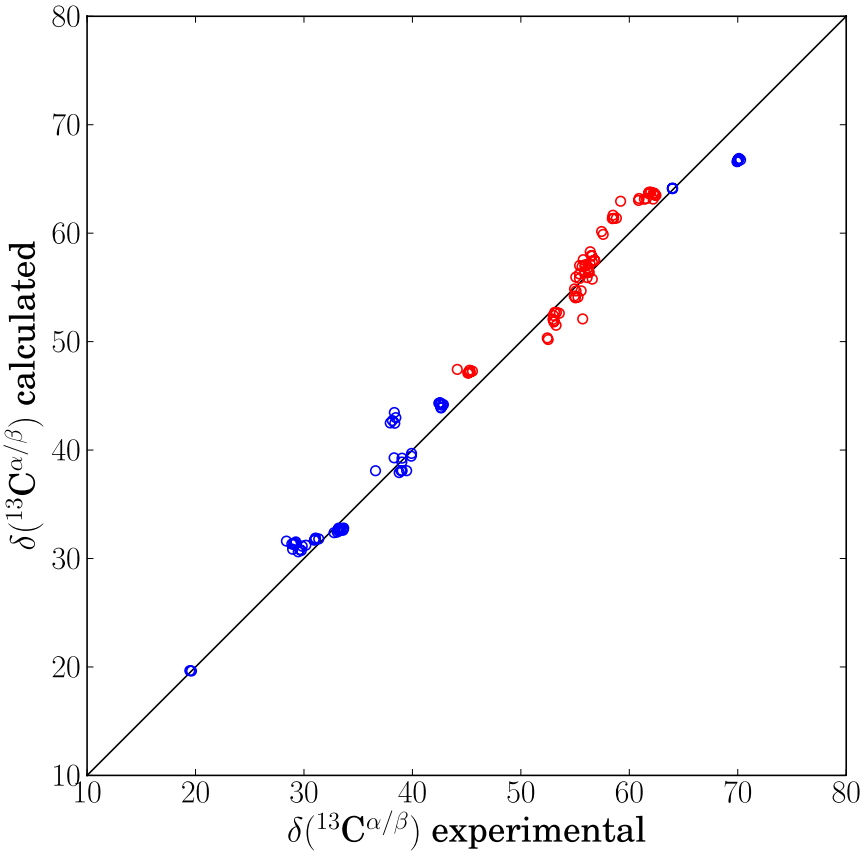


Fig. S1A. Predicted vs. experimental carbon chemical shifts ($^{13}C^{\alpha}$ – red symbols, $^{13}C^{\beta}$ – blue symbols). The experimental results are from HNCACB experiment[1] conducted under the conditions of this study (8 M urea, pH 2.0, 5°C). The calculated values are from application of the program SHIFTX[2] to all frames in the MD trajectory of the unfolded wt ubiquitin. Both calculated and experimental shifts have been subjected to re-referencing procedure, as implemented in the program SSP.[3]

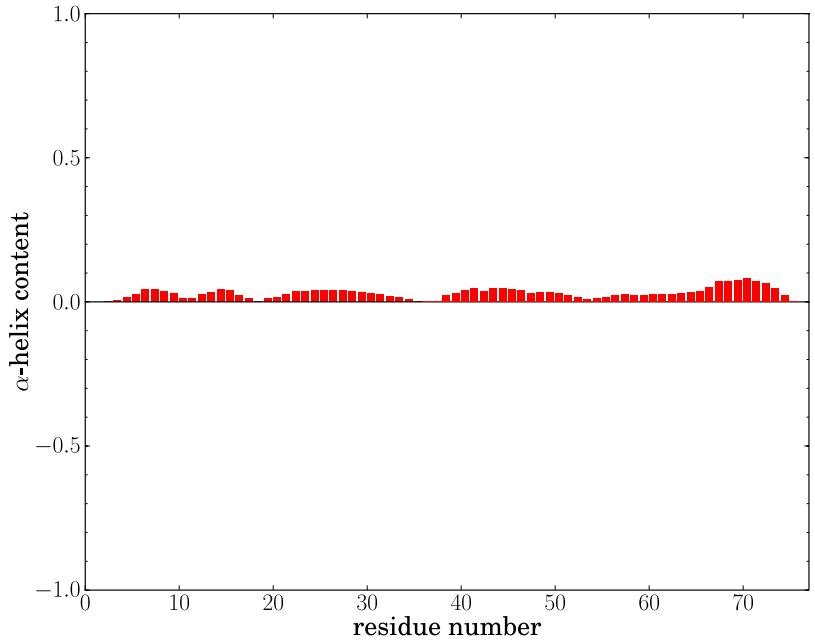


Fig. S1B. Proportion of the transient α-helical structure in the MD trajectory of unfolded wt ubiquitin. Determined by application of the program DSSP[4] to all frames in the MD trajectory.

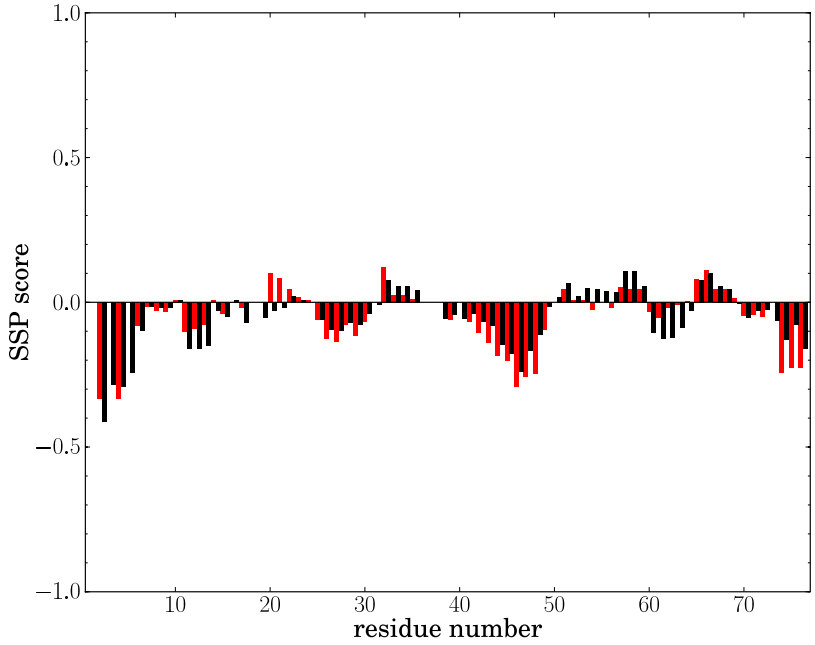


Fig. S1C. Local secondary-structure propensity in denatured wt ubiquitin. Determined by application of the program SSP[3] to the experimental $^{13}C^{\alpha}$, $^{13}C^{\beta}$ chemical shifts, as obtained by us at 5 °C (red bars) or, alternatively, by Peti *et al.*[5] at 30 °C (black bars). The SSP score of +1 (-1) corresponds to pure α-helix (β-sheet). All shifts have been subjected to re-referencing procedure, as implemented in the program SSP.

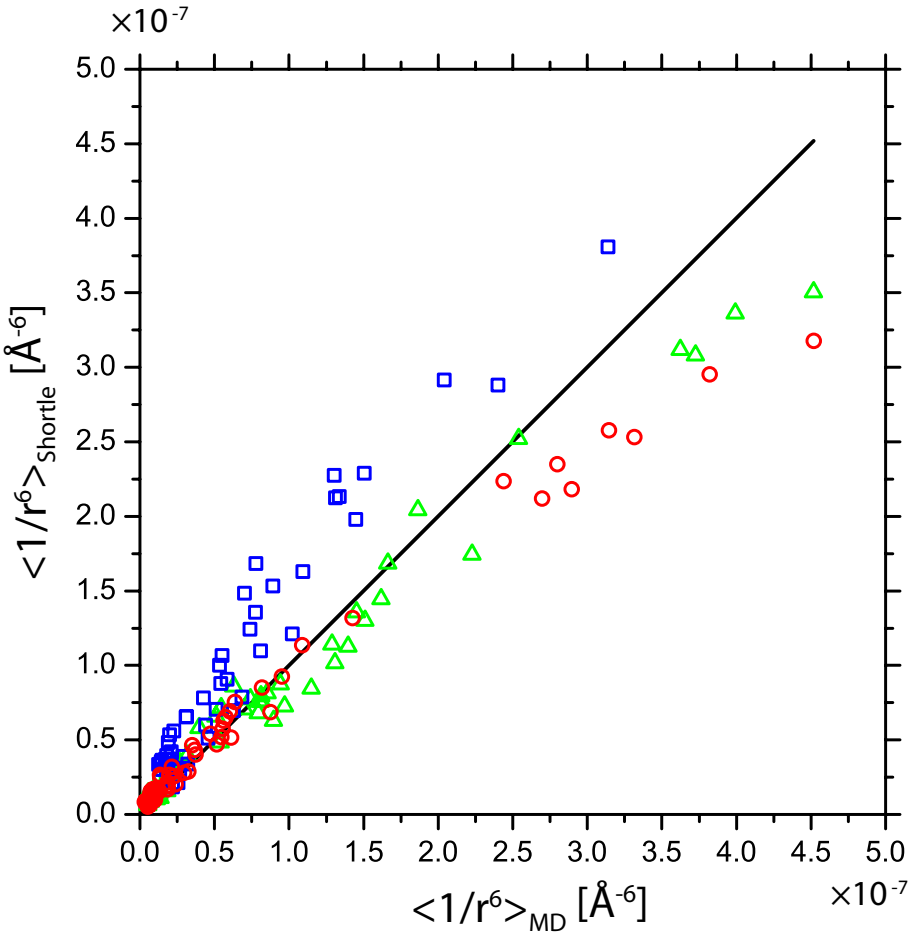


Fig. S2. Average distance factor $< r^{-6} >$ for distances r between the paramagnetic center and H^N atoms in Q2C-, D32C- and R74C-MTSL samples (red circles, blue squares, and green triangles, respectively). Values on the x axis are derived by direct averaging of r^{-6} over all frames in the respective MD trajectories. Values on the y axis are derived by analyzing the simulated PRE rates, Fig. 6, using Gillespie-Shortle formula.[6] The analyses were restricted to those residues for which the simulated PRE rates did not exceed 100 s^{-1}. The calculations using Gillespie-Shortle formula employed $\omega_{0H} / 2\pi = 600$ MHz and the generic correlation time, $\tau_c = 5.1$ ns (which has been adjusted such as to minimize the χ^2 deviation for the quantities plotted in this graph).

Transferring parameters of the MTSL label from CHARMM to AMBER.

Building a new residue, CML

Fig. S3 shows the MTSL-tagged cystein residue, termed CML. The cystein atoms are from the standard AMBER topology file; the MTSL atoms are from the CHARMM topology file due to Sezer and co-workers [7], with the atom types renamed to meet AMBER's specifications.

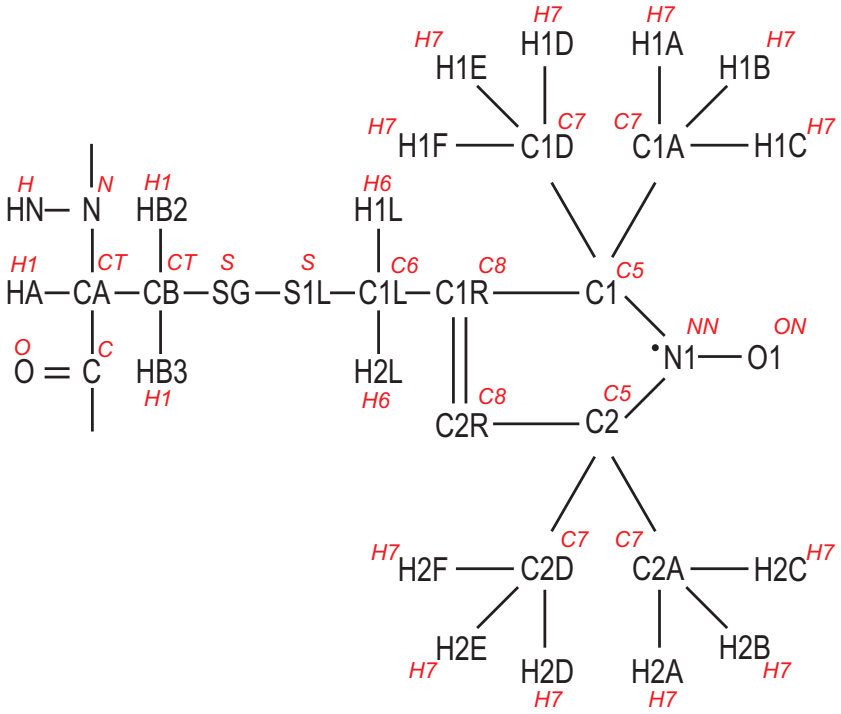


Fig. S3. Structure of CYS-MTSL (CML) residue. Atom types are indicated in red.

Charge values are obtained from the two sources: (*i*) CHARMM parameters for MTSL moiety as developed by Sezer *et al.*[7] and (*ii*) AMBER parameters for disulfide-bonded cystein residue, CYX. The two sets of charges differ only marginally in CB, HB1, HB2 and SG positions, see Fig. 2. For these atoms, the preference is given to the CHARMM values. To maintain the neutrality of the residue, we introduced a very small correction, -0.005 *e*, into the original charge of the CA atom. The complete list of the CML charges is presented in Tab. S1

Tab. S1. CML residue: partial atom charges

Name	Type	Charge
N	N	-0.4157
HN	H	0.2719
CA	CT	0.0378
HA	H1	0.0766
CB	CT	-0.1000
HB2	H1	0.0900
HB3	H1	0.0900
SG	S	-0.0800
S1L	S	-0.0800
C1L	C6	-0.1000
H1L	H6	0.0900
H1M	H6	0.0900
C1R	C8	-0.0030
C2R	C8	-0.3400
H2R	H8	0.1620
C1	C5	0.3340
C2	C5	0.3290
N1	NN	0.2200
O1	ON	-0.4380
C1A	C7	-0.3370
H1A	H7	0.0900
H1B	H7	0.0900
H1C	H7	0.0900
C2A	C7	-0.3350
H2A	H7	0.0900
H2B	H7	0.0900
H2C	H7	0.0900
C1D	C7	-0.3370
H1D	H7	0.0900
H1E	H7	0.0900
H1F	H7	0.0900
C2D	C7	-0.3350
H2D	H7	0.0900
H2E	H7	0.0900
H2F	H7	0.0900
C	C	0.5973
O	O	-0.5679

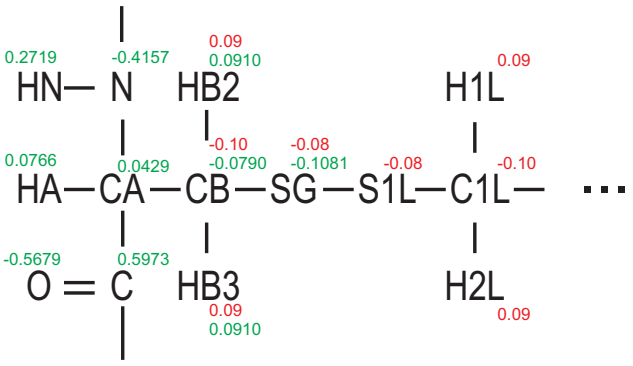


Fig. S4. Partial atom charges in the CML residue. Charge values typeset in green are from the CYX residue in AMBER; charge values typeset in red are from the MTSL moiety as parametrized by Sezer and co-workers.[7]

Translating the force-field parameters

The improved set of the CHARMM force-field parameters for the MTSL label has been made available by Sezer *et al.*[7] To translate these parameters for use with Amber, we took advantage of the new tool in the AMBER 11 distribution, CHAMBER.[8] Initially, the ubiquitin mutant D32C was built under CHARMM (v32b2) using the standard topology and parameters files, top_all27_prot_lipid.rtf and par_all27_prot_lipid.prm, together with the Sezer's MTSL patch. The generated psf and crd files were then converted into AMBER prmtop file with the help of CHAMBER. The BOND, ANGL, DIHE and IMPR force-field parameters for all possible combinations of protein atoms have been extracted from the prmtop file by means of the rdparm program which is also supplied as a part of the AMBER 11 distribution. Finally, an in-house python script was used to collect all non-repetitive force-field parameter items.

The NONBON parameters were directly translated from Sezer's parameter file and par_all27_prot_lipid.prm, keeping in mind the functional form of the van der Waals potential:

$$E_{\mathrm{vdw}} = \varepsilon\left[\left(\frac{R_{\mathrm{min},ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\mathrm{min},ij}}{r_{ij}}\right)^{6}\right].$$

To verify the correctness of the translated force-field parameters, we resorted to direct comparison of the individual energy terms. The procedure can be described as follows: (*i*) a short peptide GLN-CML-LYS, containing the centrally located MTSL tag, was built and energy-minimized under CHARMM v32b2; (*ii*) the resulting structure was used to evaluate the individual energy terms using CHARMM with non-bonded cutoff set to 999 Å and

CMAP module disabled;[*] (*iii*) only the energy terms associated with the MTSL moiety were retained using the ANALysis command; (*iv*) the coordinates of the peptide from step (*i*) were loaded into AMBER using the crd file for extra precision, and the corresponding AMBER topology file was built; (*v*) a special python module, based on readparm.py, has been developed to evaluate the Amber energy terms pertaining to the selected atom sets; it has been tested by comparing the results with the total energies as reported by Amber. The comparison of the CHARMM and AMBER individual energy terms demonstrated that the two sets of values are in excellent agreement, with only minimal deviations (less than 0.03 kcal/mol). These minor deviations should be attributed to the 'interfacing' of the CHARMM MTSL parameters with the AMBER backbone parameters across the CA-CB bond. AMBER force-field parameter and topology files for the CML residue, frcmod.cml and cml.lib, are supplied as a part of the Supporting Information.

References

(1) Muhandiram, D. R.; Kay, L. E. *J. Magn. Reson. Ser. B* **1994,** *103*, 203.
(2) Neal, S.; Nip, A. M.; Zhang, H. Y.; Wishart, D. S. *J. Biomol. NMR* **2003,** *26*, 215.
(3) Marsh, J. A.; Singh, V. K.; Jia, Z. C.; Forman-Kay, J. D. *Protein Sci.* **2006,** *15*, 2795.
(4) Kabsch, W.; Sander, C. *Biopolymers* **1983,** *22*, 2577.
(5) Peti, W.; Smith, L. J.; Redfield, C.; Schwalbe, H. *J. Biomol. NMR* **2001,** *19*, 153.
(6) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997,** *268*, 158.
(7) Sezer, D.; Freed, J. H.; Roux, B. *Journal of Physical Chemistry B* **2008,** *112*, 5755.
(8) Crowley, M. F.; Williamson, M. J.; Walker, R. C. *International Journal of Quantum Chemistry* **2009,** *109*, 3767.

[*] Note that CMAP dihedral energy corrections have been derived for MTSL,[7] but we have not made an attempt to formulate the corresponding corrections for AMBER.

Complete references, as found in the main text

(71) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998,** *102*, 3586.

(173) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber 11 Users' Manual*. University of California, San Francisco **2010**.

(178) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009,** *30*, 1545.