THE PROTEIN SOCIETY

Ensemble MD simulations restrained via crystallographic data: Accurate structure leads to accurate dynamics

Yi Xue[1] and Nikolai R. Skrynnikov[1,2]*

[1]Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, Indiana 47907-2084, USA
[2]Laboratory of Biomolecular NMR, St. Petersburg State University, St. Petersburg 199034, Russia

Received 8 July 2013; Revised 6 January 2014; Accepted 18 January 2014
DOI: 10.1002/pro.2433
Published online 22 January 2014 proteinscience.org

Abstract: Currently, the best existing molecular dynamics (MD) force fields cannot accurately reproduce the global free-energy minimum which realizes the experimental protein structure. As a result, long MD trajectories tend to drift away from the starting coordinates (e.g., crystallographic structures). To address this problem, we have devised a new simulation strategy aimed at protein crystals. An MD simulation of protein crystal is essentially an ensemble simulation involving multiple protein molecules in a crystal unit cell (or a block of unit cells). To ensure that average protein coordinates remain correct during the simulation, we introduced crystallography-based restraints into the MD protocol. Because these restraints are aimed at the ensemble-average structure, they have only minimal impact on conformational dynamics of the individual protein molecules. So long as the average structure remains reasonable, the proteins move in a native-like fashion as dictated by the original force field. To validate this approach, we have used the data from solid-state NMR spectroscopy, which is the orthogonal experimental technique uniquely sensitive to protein local dynamics. The new method has been tested on the well-established model protein, ubiquitin. The ensemble-restrained MD simulations produced lower crystallographic *R* factors than conventional simulations; they also led to more accurate predictions for crystallographic temperature factors, solid-state chemical shifts, and backbone order parameters. The predictions for ^{15}N R_1 relaxation rates are at least as accurate as those obtained from conventional simulations. Taken together, these results suggest that the presented trajectories may be among the most realistic protein MD simulations ever reported. In this context, the ensemble restraints based on high-resolution crystallographic data can be viewed as protein-specific empirical corrections to the standard force fields.

Keywords: protein structure and dynamics; Molecular Dynamics simulations; force fields; solid-state NMR; protein crystallography; chemical shifts; crystallographic *R* factors; crystallographic *B* factors; order parameters; ^{15}N relaxation; ubiquitin

Introduction

Molecular dynamics (MD) is a powerful tool for modeling protein conformational dynamics, with particular emphasis on functionally relevant motions. Importantly, MD simulations can reconstruct the picture of motion in its entirety, including those aspects that cannot be easily probed experimentally. Unfortunately, current MD trajectories tend to drift away from the starting coordinates (e.g., crystallographic

Additional Supporting Information may be found in the online version of this article.

Yi Xue's current address is Department of Chemistry & Biophysics, University of Michigan, 930 North University Avenue, Ann Arbor, MI, 48109-1055, USA.

Grant sponsor: NSF grant; Grant number: MCB 1158347.

*Correspondence to: Nikolai Skrynnikov; Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47907-2084, USA. E-mail: nikolai@purdue.edu

structures) during the course of the simulation. This fact has been brought into spotlight by a very recent work of Shaw and coworkers.[1] In their study, a number of ultralong (at least 40 μs) MD trajectories have been recorded using state-of-the-art force fields. In all cases, it was found that the simulated structures "moved away" from the true coordinates; in most cases, the structures continued to deteriorate throughout the course of the simulation (sometimes to a substantial degree). This helps to explain the notable lack of successes in many previous attempts to refine protein models by means of unconstrained MD (uMD) simulations in explicit solvent. With the exception of some small, tightly packed proteins,[2–4] the uMD approach generally fails to improve the models in the range 1–10 Å from the target structure.[5–9] Initially, this situation was blamed on short uMD trajectories that could not adequately sample the conformational phase space. However, the latest results suggest that "the structure that realizes the global free-energy minimum for the force field employed is not the X-ray or NMR structure,"[1] i.e. that the force field itself is to blame.

This is a disappointing result which casts a long shadow on the future of conventional protein MD simulations. Clearly, there is need for systematic work on development and redesign of force fields, which remains a major challenge for the foreseeable future. To illustrate the complexity of this challenge, we will mention that the most advanced polarizable force field, AMOEBA, currently fails to maintain the integrity of certain protein structures for more than several nanoseconds.[10] As an alternative to such ground-up redesign work, the existing MD force fields can be amended based on experimental data; the emerging trend is to optimize force-field parameters based directly on the data from protein studies.[11–13]

Here, we propose a more pointed strategy, where protein-specific restraints are introduced directly into the MD simulation. Our motivation is to eliminate the bias in the force field that causes protein structures to drift. Toward this goal, we use the crystallography-based restraints, which are far more complete and accurate than any other experimental data insofar as protein structure is concerned. As crystallographic data pertain to the mean protein structure (averaged over dynamic fluctuations), the corresponding restraints should be applied in a form of ensemble average. In this manner, the simulated protein ensemble remains consistent with X-ray diffraction data (i.e., maintains the correct average structure), whereas the individual protein molecules retain their native-like internal dynamics.

Hydrated protein crystals are uniquely suited to implement this strategy. MD simulations of protein crystals have been the area of interest,[14–16] with emerging applications to solid-state NMR (ssNMR) spectroscopy.[17,18] An MD simulation of a protein crystal is intrinsically an ensemble simulation, as it involves multiple protein molecules in a crystal unit cell (or a block of unit cells). Therefore, it is straightforward to incorporate the crystallography-based ensemble restraints into the standard MD protocol. In addition to X-ray diffraction, protein crystals offer access to another incredibly rich source of experimental information—ssNMR data. These two types of data are largely orthogonal, as ssNMR can probe internal protein dynamics at the level of detail that is not available to X-ray crystallography. This creates an opportunity for rigorous cross-validation of the obtained results. Briefly, the proposed ensemble-restrained MD (erMD) strategy relies on X-ray data to ensure that the average protein structure remains correct during the course of the simulation, while ssNMR data are used to verify that the resulting trajectories accurately reproduce protein dynamics.

To establish a feasibility of our approach, we have focused on crystalline ubiquitin. Ubiquitin is one of a handful of proteins for which major efforts have been made to characterize protein structure and dynamics by means of ssNMR[19–24] and, furthermore, to establish a connection between the NMR and crystallographic samples.[25] Implementing ensemble restraints eliminated structural drift in the trajectory of crystalline ubiquitin, while preserving the dynamics of individual ubiquitin molecules. We have found that erMD trajectories produced significantly lower crystallographic R factors than comparable uMD simulations. Furthermore, the erMD simulations were more successful in predicting ssNMR chemical shifts. We have also observed improvements in crystallographic temperature factors and backbone order parameters S^2_{NH}. Finally, erMD was at least as accurate as uMD in predicting ^{15}N R_1 rates. Taken together, these results suggest that erMD simulations provide a uniquely accurate model of protein structure and dynamics.

Methods

Figure 1 shows the crystal unit cell of ubiquitin based on the recent crystallographic structure 3ONS (six protein molecules per unit cell, one protein molecule per asymmetric unit). Using these coordinates, we have recorded a 1-μs unrestrained MD trajectory of hydrated ubiquitin crystal. The effect of crystal lattice in this simulation is modeled via the periodic-boundary conditions. As it turns out, the average protein coordinates obtained from this MD trajectory deviate by 0.52 Å (backbone rmsd) from the original crystallographic coordinates. The deviation of this magnitude is beyond the uncertainty of high-resolution crystallographic structure. In fact, rmsd becomes progressively worse during the course of the simulation, climbing to 0.7 Å toward the end of

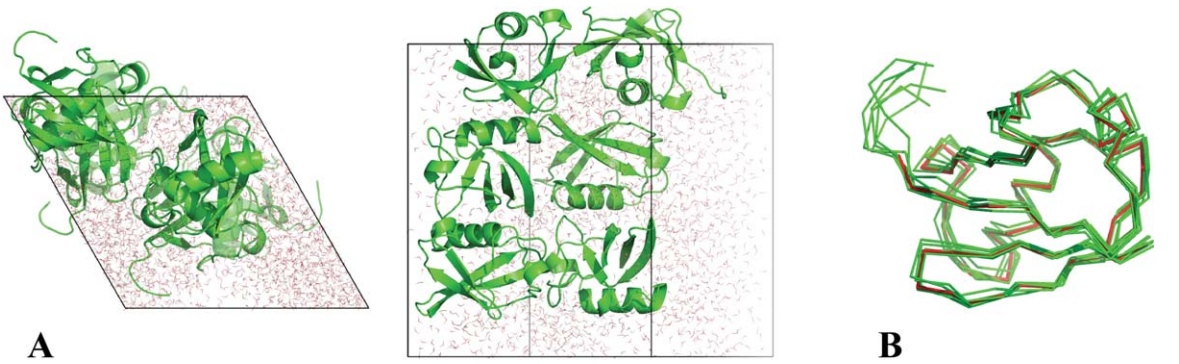

Figure 1. (A) The snapshot from erMD simulation of ubiquitin showing periodic-boundary box (corresponding to the single crystal unit cell, 1U). The unit cell with the primitive trigonal space group $P3_221$ is based on the crystallographic structure 3ONS. The reported dimensions of the cell, $a=b$ and c, are all increased by a factor 1.016 to account for thermal expansion of the protein crystal on transition from 100 K (temperature at which 3ONS was solved) to 301 K (temperature at which ssNMR data were taken).[27] Shown are the top view and side view of the unit cell. The MD trajectory was recorded with $k_0=0.1$ kcal mol^{-1} Å^{-2}; the displayed snapshot represents the time point 150 ns. The areas with apparent low water density arise from the periodic-boundary images of ubiquitin molecules. (B) Six ubiquitin molecules from the MD frame, panel A, superimposed according to Eq. (2.1) (green Cα traces). Also shown is the crystallographic structure 3ONS centered according to Eq. (2.2) (red Cα trace). Such superpositions are used to calculate the instantaneous value of $U_{\text{restraint}}$, Eq. (1). Since protein molecules are superimposed via the symmetry transformations rather than least-square fitting, $U_{\text{restraint}}$ proves to be sensitive to small reorientations of proteins in the simulated unit cell.

the trajectory, see Figure 2(a). The simulated diffraction data also suggest that the quality of the protein structure becomes degraded in the MD simulation, as manifested by the increased R factor.[26]

This behavior is the manifestation of the coordinate drift, caused by a subtle bias in the MD force field. In addition, one should bear in mind that MD trajectories cannot easily accommodate some of the experimental conditions, such as the presence of protein species with different charges due to titratable side-chain sites. To address this situation, we implemented the MD restraints seeking to ensure that the ensemble of protein molecules contained in the simulation box is on average consistent with the crystallographic structure. In this manner, we use the X-ray crystallography data as ensemble restraints, while retaining the (orthogonal) ssNMR data for the purpose of validation.

Generally speaking, it is desirable to restrain crystal MD simulation directly against the crystallographic diffraction data. Indeed, diffraction data contain the entirety of experimental information, including certain amount of information about the conformational diversity in the system. We have implemented this strategy programmatically and found it unsatisfactory: as it turns out, diffraction-based ensemble restraints are incompatible with *bona fide* MD simulations. The reasons for this failure are discussed in Supporting Information. In this situation, we pursue a more practical solution, using crystallographic coordinates of a protein to generate ensemble restraints. Specifically, we seek to ensure that the average protein structure, as calculated over the MD ensemble, remains close to the X-ray structure. This is accomplished by introducing the following pseudopotential:

$$U_{\text{restraint}} = k \sum_{i=1}^{N\text{atom}} |\overline{\tilde{\mathbf{x}}_i^{(q)\text{MD}}} - \tilde{\mathbf{x}}_i^{\text{cryst}}|^2 \qquad (1)$$

The pseudopotential $U_{\text{restraint}}$ is harmonic with the force constant $k=k_0N_{\text{prot}}$, where N_{prot} is the number of protein molecules in the simulation. With this choice of k, the pseudoforce acting on an individual atom does not depend on the size of the simulated system. $\mathbf{x}_i^{\text{MD}}$ is the vector representing current coordinates of the i-th heavy atom in the MD trajectory and $\mathbf{x}_i^{\text{cryst}}$ represents the corresponding coordinates in the crystallographic structure. The summation in Eq. (1) is over all atoms contained in the crystallographic structure (typically these are heavy atoms). Because there are multiple protein molecules in the simulated unit cell(s), Figure 1(a), they need to be superimposed prior to comparison with the crystallographic coordinates. This is accomplished by applying symmetry operators as appropriate for the given crystal space group:

$$\tilde{\mathbf{x}}_i^{(q)\text{MD}} = \hat{\mathbf{R}}^{(q)} \left(\mathbf{x}_i^{(q)\text{MD}} - \mathbf{v}^{(q)\text{MD}} \right) \qquad (2.1)$$

Here, index q enumerates protein molecules in the MD frame, vector $\mathbf{v}^{(q)\text{MD}}$ translates the center of mass of the particular protein molecule to the origin of the coordinate frame, and $\hat{\mathbf{R}}^{(q)}$ represents the symmetry rotation matrix.[28] Subsequent to this manipulation, the coordinates of all protein

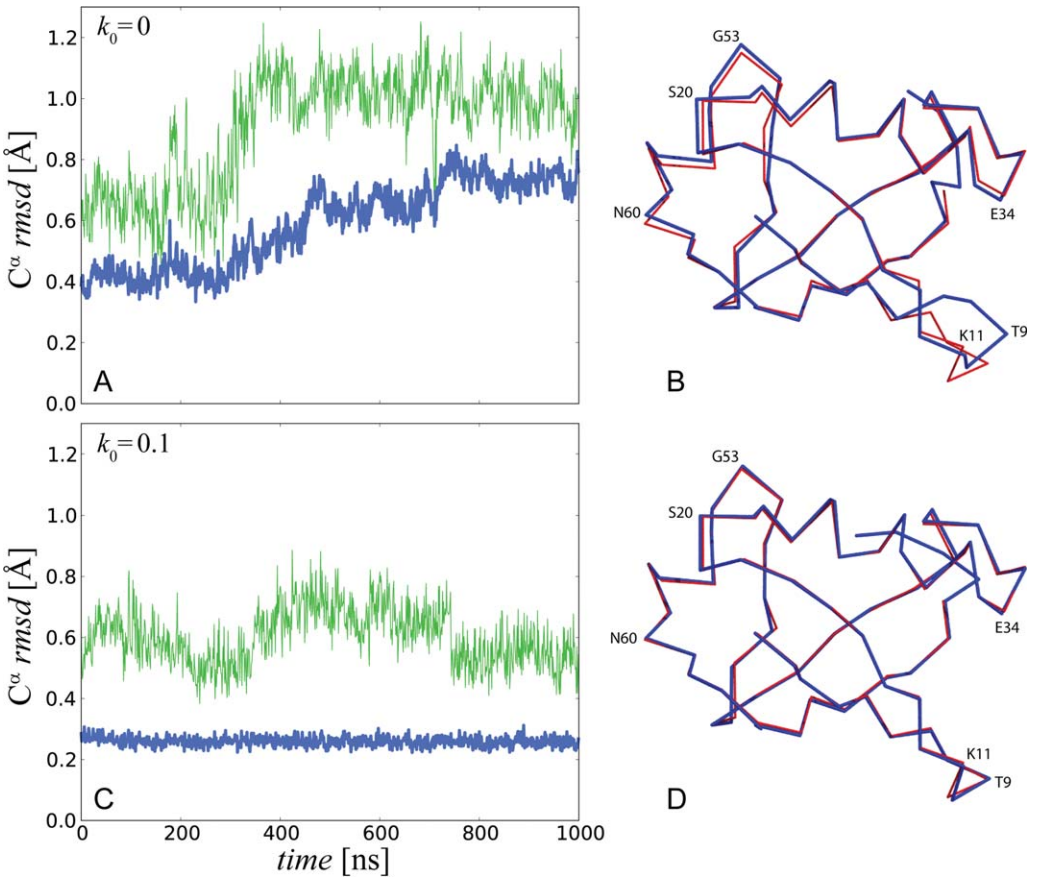

Figure 2. (A) Time course of C^α rmsd for uMD ($k_0=0$) trajectory of crystalline ubiquitin. The simulation models a single crystal unit cell (1U) with six ubiquitin molecules. Blue profile represents the ensemble-average rmsd, where protein coordinates $\mathbf{x}^{(q)\mathrm{MD}}$ are overlaid according to Eq. (2.1) and then averaged before calculating the rms deviation from $\mathbf{x}^{\mathrm{cryst}}$. Green profile represents the rmsd of one individual ubiquitin molecule randomly selected from the ensemble of six. The sampling step is 1 ns. The increase in rmsd observed in this graph does not necessarily mean that the structure will continue further degrading if the simulation is extended. Ubiquitin is one of those small proteins where the structure can be relatively well maintained in the MD simulations. In the recent ultralong solution trajectory, ubiquitin remained within 0.5-1.0 Å of the crystal structure.[29] It remains to be seen what is the magnitude of structural drift in long crystal trajectories. (B) Average protein coordinates calculated from the final 100 ns of the uMD trajectory (blue trace) superimposed onto the crystal coordinates (red trace). Structural deviations are found in the area which is known for its plasticity and serves as a ligand-binding interface (loop β1–β2 and C-terminal end the strand β5, lower right part of the molecule). The opposite side of the molecule (β2–α1 loop interacting with the β-turn at the site G53) is affected as well. (C, D) Same as (A) and (B), respectively, for erMD ($k_0=0.1$) trajectory.

molecules in the MD frame are averaged, $\overline{\tilde{\mathbf{x}}_i^{(q)\mathrm{MD}}}$. In turn, the crystallographic structure is also translated to the origin:

$$\tilde{\mathbf{x}}_i^{\mathrm{cryst}} = \mathbf{x}_i^{\mathrm{cryst}} - \mathbf{v}^{\mathrm{cryst}} \tag{2.2}$$

Finally, the deviation between the ensemble-average MD structure and crystallographic structure is used to generate the correcting force according to Eq. (1) (see also Supporting Information). The superposition of multiple protein structures used in calculating $U_{\mathrm{restraint}}$ is illustrated in Figure 1(b).

Clearly, the restraints Eq. (1) are sensitive to the internal dynamics of protein molecules. In addition, they are also sensitive to rigid-body reorientational dynamics (i.e., small-amplitude rocking motion of protein molecules embedded in the crystal lattice). The only motional mode which is left out is translation—the restraints are insensitive to translational displacements of ubiquitin molecules in the crystal lattice.

The restraints set up in this fashion have only mild effect on each individual ubiquitin molecule. In essence, individual molecules are free to move as dictated by the original force field, so long as the ensemble average remains close to the crystallographic structure. When a difference emerges between the ensemble average $\overline{\tilde{\mathbf{x}}_i^{(q)\mathrm{MD}}}$ and the crystallographic structure $\tilde{\mathbf{x}}_i^{\mathrm{cryst}}$, a small correcting force is applied across the ensemble [one and the same force, derived from the pseudopotential Eq. (1), acts on the i-th atom in all ubiquitin molecules]. Assuming that MD simulation includes a sufficiently large number of protein molecules, this

approach should remedy the average structure without stifling the dynamics. As it turns out, our method can actually improve the modeling of dynamics (see below).

The pseudopotential $U_{\text{restraint}}$ was incorporated into Amber ff99SB*-ILDN force field in Amber 11 MD simulation package.[30] This is one of the most successful protein force fields which includes the backbone helical propensity corrections[12] and the ILDN side-chain corrections.[31] The initial coordinates for the MD simulations were derived from the recent crystallographic structure of ubiquitin 3ONS, as illustrated in Figure 1. This structure has been solved with the explicit goal to obtain a crystallographic model suitable for the analyses of the ssNMR data.[25] Importantly, the sample has been crystallized in the same crystal form as used in the ssNMR experiments.

Prior to the start of the MD trajectory, we have extended the peptide chain of ubiquitin by adding residues 73–76, for which crystallographic coordinates are unavailable. The protein structure was then protonated; the protonation status of Asp and Glu was determined according to the PROPKA[32] calculations for crystallization conditions pH 4.2. The results were generally consistent with the estimations using solution pK_a,[33] except for several residues experiencing the effect of crystal contacts. The unit crystal cell was hydrated using SPC/E water, which has been recommended for protein crystal simulations with Amber ff99SB force field.[34] In doing so, the crystallographic water molecules have been retained in their original positions. Finally, the system was neutralized by adding counter ions and equilibrated before the production run. The simulations were conducted at 301 K, which is the temperature used in ssNMR measurements, using the NPT ensemble. The simulated systems ranged from a 1U to a block of four crystal unit cells (4U). In the latter case, the simulations involved 24 ubiquitin molecules and about 8770 water molecules, for the total of 56,240 atoms. For this system, the production rate using NVIDIA GeForce GTX580 cards was 9 ns per card per day. The complete MD protocol is described in the Supporting Information.

Results

C^α rmsd

The data for C^α rms deviation between the different ubiquitin models and the target structure 3ONS are summarized in the first column in Table I. The widely used crystal structure of ubiquitin 1UBQ[42] belongs to a different space group than 3ONS. This is manifested in substantial C^α rmsd between the two sets of coordinates, 0.43 Å. The solution-state conformational ensemble 2KOX[43] displays a similar level of agreement. In the case of unrestrained solution MD trajectory, the deviation rises to 0.86 Å. The crystal simulation appears to fare better, with average C^α rmsd of 0.52 Å (unrestrained simulation, $k_0=0$; here, and in what follows we cite the results from 1U trajectories unless indicated otherwise). One should bear in mind though that the quality of the structure gradually deteriorates through the course of this simulation, Figure 2(a). Ultimately, during the final 100-ns segment of the trajectory average C^α rmsd amounts to 0.71 Å (not including the disordered C-terminus). This is well beyond the intrinsic uncertainty of the crystallographic structure 3ONS. Indeed, the reported resolution of 3ONS is 1.8 Å. At this level of resolution, the accuracy of backbone coordinates is expected to be near 0.2 Å[44,45]. It is most likely that the elevated rmsd is due to subtle biases in the force-field parameters, as well as the approximate character of the MD setup.

To address this problem, we have implemented the erMD protocol, as described above. Already the use of very weak restraints, $k_0=0.1$ kcal mol^{-1} Å^{-2}, promptly brings rmsd down to the level of 0.22 Å. Bear in mind that this rmsd value describes the deviation between the ensemble-average MD coordinates and the target, the conformational diversity of ubiquitin across the ensemble is retained. This is illustrated in Figure 2(c). Although ensemble-average ubiquitin structure remains within 0.2–0.3 Å of the reference X-ray coordinates (blue trace in the plot), one single ubiquitin molecule which is a part of the ensemble shows much larger deviations (green trace). Furthermore, this one molecule undergoes significant conformational fluctuations. In doing so, it samples certain conformational states that turn out to be sufficiently long-lived (on the order of hundreds of nanoseconds). This behavior demonstrates that individual protein molecules largely retain their native-like internal dynamics in our erMD simulations.

The simulation results obtained with $k_0=0.1$ kcal mol^{-1} Å^{-2} are justified by the accuracy of the X-ray structure. It is reasonable to expect that ensemble- and time-averaged MD coordinates fall within ca. 0.2 Å of the X-ray structure, because the uncertainty margin of the crystallographic coordinates is ca. 0.2 Å. Increasing the force constant from 0.1 to 1.0 kcal mol^{-1} Å^{-2} reduces C^α rmsd to 0.10 Å (see Table I). When the restraints are strengthened even further, to 10.0, the rmsd drops to 0.05 Å. The latter situation should be viewed as "over-restraining" as the limited accuracy of the crystal coordinates does not justify excessive tightening of the (average) structure.

Crystallographic R factors

The standard structure-calculation protocol in X-ray crystallography accounts for local protein dynamics via adjustable per-atom *B* factors. Conversely, if MD trajectory is used as a structural model to interpret

Table I. *Comparison Between the Experimental Data from Crystalline Ubiquitin and the Predictions Using Different Structural / MD Models*

	rmsd to 3ONS[a] (Å)	R factor[b]		$E_{\text{restraint}}$ per aa[c] (kcal/mol)	rmsd (δ_{calc}, δ_{exptl})[d] (ppm)			rmsd (S^2_{calc}, S^2_{exptl})[e]
		R_{work}	R_{free}		^{15}N	^{13}Cα	^{13}Cβ	
3ONS	0	0.30	0.31	–	2.39	0.75	1.11	–
1UBQ	0.43	0.44	0.41	–	2.77	0.85	1.29	–
2KOX	0.36	0.37	0.35	–	2.89	0.83	1.23	0.056
Solution MD, k_0=0 (1 μs)	0.86	0.41	0.39	–	3.02	0.97	1.26	0.048
Solid MD, k_0=0, 1U (1 μs)	0.52	0.41	0.39	–	2.96	0.92	1.15	0.056
Solid MD, k_0=0, 4U (200 ns)	0.37	0.37	0.35	–	2.91	0.92	1.12	0.062
Solid MD, k_0=0.1, 1U (1 μs)	0.22	0.31	0.29	0.21	2.72	0.90	1.09	0.043
Solid MD, k_0=0.1, 4U (200 ns)	0.21	0.32	0.31	0.18	2.73	0.90	1.09	0.040
Solid MD, k_0=1, 1U (1 μs)	0.10	0.29	0.28	0.48	2.68	0.89	1.11	0.047
Solid MD, k_0=1, 4U (200 ns)	0.09	0.32	0.29	0.36	2.68	0.89	1.11	0.046
Solid MD, k_0=10, 1U (1 μs)	0.05	0.37	0.36	0.76	2.64	0.85	1.12	0.041
Solid MD, k_0=10, 4U (200 ns)	0.05	0.32	0.29	0.54	2.65	0.85	1.12	0.041

The shaded rows correspond to the recommended k_0 setting.

[a] C^{α} rmsd relative to the crystallographic structure. In the case of crystal MD simulations, protein coordinates $\mathbf{x}^{(q)\text{MD}}$ are overlaid according to Eq. (2.1) and then averaged over the entire trajectory; the average coordinates are superimposed onto 3ONS in the least-square sense (via C^{α} atoms) before calculating the rms deviation from $\mathbf{x}^{\text{cryst}}$. In other cases, protein coordinates are superimposed onto 3ONS, averaged if necessary, and then used to calculate the rmsd.

[b] In calculating crystallographic R, all per-atom B factors have been omitted. This was done to facilitate the comparison between MD models (which encode local dynamics) and static structures (which are dynamics-free). Furthermore, no attempt was made to calculate reflections from explicit water molecules. In the case of crystal MD trajectories, each protein molecule was first transformed according to Eq. (2.1). Then, structure factors $F_{\text{calc}}(h,k,l)$ were computed using the *fmodel* tool in PHENIX.[35] In doing so, the flat bulk-solvent contribution was included with k_{sol}=0.35 e Å^{-3} and B_{sol}=46 Å^2, as recommended by Fokine and Urzhumtsev.[36] The obtained values $F^{(q)}_{\text{calc}}(h,k,l)$ from individual ubiquitin molecules have been averaged (with phases) to determine the intensities of reflections, $|\overline{F^{(q)}_{\text{calc}}(h,k,l)}|^2$, which were in turn averaged over the entire trajectory. The result was then subjected to the overall scaling to account for the effect of lattice vibrations (translational movement of the protein molecules).[37] The degree of overall anisotropy, as reported in 3ONS, is modest; therefore, we chose to use the isotropic scaling whereby a single B_{iso} value was optimized using a designated script. Finally, the results were correlated to $F_{\text{obs}}(h,k,l)$ and the crystallographic R factor was calculated in a standard manner. When calculating R_{work} and R_{free}, we used the same subsets of reflections as listed for 3ONS. For structural models other than crystal MD trajectories, the protein coordinates were first superimposed onto 3ONS in the least-square sense (via C^{α} atoms); the remaining calculations followed the same procedure as described above.

[c] The restraint energy per residue, $E_{\text{restraint}}=\langle U_{\text{restraint}}\rangle/N_{\text{res}}$, where $U_{\text{restraint}}$ is calculated according to Eq. (1) and subsequently averaged over all snapshots in the trajectory and N_{res} is the number of residues for which crystallographic restraints are available, N_{res}=72.

[d] Chemical shifts were calculated using the program SHIFTX2 version 1.07.[38] A customized version of the program, where ubiquitin was excluded from the training set to avoid biasing the results, was kindly provided by B. Han. The program was used on static structures as well as MD frames, processing one protein structure at a time (disregarding small shifts across protein-protein interface, e.g., due to ring current shifts). Taking intermolecular effects into consideration leads to a slight improvement in δ_{calc} (e.g., by ca. 0.05 ppm for ^{15}N nuclei). In the case of MD data, every 10-th snapshot was included in the chemical shift calculations, corresponding to 50-ps sampling step. The control calculations using 5-ps sampling step produced the results that were virtually identical. The experimental data were obtained from the studies by Igumenova *et al*.[39] (^{13}C) and Schanda *et al*.[23] (^{15}N); we found that there was no need to re-reference these chemical shifts.

[e] ^{15}N-^{1}H$^{\text{N}}$ dipolar order parameters for crystal trajectories were computed using the following protocol. First, symmetry transformations Eq. (2.1) have been applied to all ubiquitin molecules in the periodic boundary box. Then, ^{15}N-^{1}H$^{\text{N}}$ vectors were extracted from the transformed coordinates; the vectors pertaining to each individual residue were arranged in a long array. The array had an effective length of 6 × 1 = 6 μs in the case of 1U simulations and 0.2 × 24 = 4.8 μs in the case of 4U simulations. Finally, the standard Brüschweiler's formula[40] has been applied to these arrays to calculate $S^2_{i,\text{calc}}$ values. The experimental data $S^2_{i,\text{exptl}}$ are from the recent solid-state NMR experiments by Haller and Schanda,[41] which is the revision of the earlier work by Schanda *et al*.[23] Additionally, the table includes the results from solution-state ensemble 2KOX and 1 μs-long solution simulation. For these models, $S^2_{i,\text{calc}}$ values were obtained by straightforward application of the Brüschweiler's formula.

X-ray diffraction data then local protein dynamics is taken into consideration explicitly. These two approaches to local dynamics are significantly different, which potentially complicates the comparison between the respective models. To simplify the analyses, we excluded per atom B factors from further consideration (more precisely, for each model we employed a single adjustable B_{iso} value which was meant to capture the effect of lattice vibrations). This puts different models in Table I on the same footing, allowing for a clear-cut comparison of the R values.

Another simplification that we have made in our analyses is the neglect of explicit water. The coordinate set 3ONS includes 91 crystallographic water molecules. Conversely, the MD models include on the order of several thousand water molecules, some of which belong to the protein hydration shell, whereas others are classified as bulk solvent. Once again, the situation is asymmetric. To simplify the treatment, we have chosen to ignore the explicit water and instead use flat bulk solvent correction for the portion of space that is not occupied by protein molecules.[36,37]

Clearly, the above simplifications degrade the performance of the original crystallographic model. The original deposition 3ONS reports $R_{\mathrm{work}} = 0.18$ and $R_{\mathrm{free}} = 0.21$. With our simplified protocol, these values rise to 0.30 and 0.31, respectively. The importance of this result is that it provides the point of reference for further comparative analyses. In particular, the unrestrained MD simulation of the ubiquitin crystal produces the R factors 0.41 and 0.39, which is significantly worse than the static crystallographic structure. This means that uMD trajectory provides an inferior structural model, as judged on the basis of the experimental diffraction data. When restraints are turned on, the situation is improved. Both k_0=0.1 and 1.0 trajectories of a single unit cell (1U) produce R values that are essentially the same as in the case of 3ONS. Thus, the erMD simulation can at least match the quality of the crystallographic model in this rubric, if not surpass it. Further strengthening of the restraints, k_0=10.0, can make the results worse. As it appears, the excessive force leads to MD artifacts—specifically, the individual ubiquitin molecules in 1U trajectory become slightly reoriented (while the ensemble-average structure remains near-perfect). As already indicated, the value k_0= 10.0 corresponds to over-restraining and thus should be rejected.

The lowest R factors obtained in the erMD simulations are seemingly unimpressive, about 0.30. Note, however, that including explicit water should significantly reduce this value. Also bear in mind that low R factors, about 0.20, that are customary for high-resolution X-ray crystallography are obtained with the help of per-atom B factors, which effectively create a very large number of fitting parameters. In the current treatment, these fitting parameters have been eliminated. Interestingly, MD trajectories listed in Table I display the values of R_{free} that tend to be slightly lower than R_{work}. As it turns out, this is a statistical effect which depends on the specific subset of reflections used to calculate R_{free}. Additional calculations using randomly chosen subsets of $F_{\mathrm{obs}}(h,k,l)$ led us to conclude that R_{work} and R_{free} are equal within the statistical error. Of note, this situation is different from crystallographic refinement where R_{work} is subject to minimization and thus tend to be somewhat lower than R_{free}.

Restraint energy

Listed in Table I are the average restraint energies as registered in the series of erMD simulations (per mole of ubiquitin per residue). The lowest energies, ~0.2 kcal mol^{-1} per residue, are found in k_0=0.1 trajectories. In the strongly restrained simulations, the energies increase by threefold to fourfold. The value 0.2 kcal mol^{-1} is comparable to intrinsic uncertainties of the existing force fields. For instance, the accuracy of MD-based calculations for hydration free energies of amino-acid side chains is no better than about 1 kcal mol^{-1}.[46] Similarly, the MD-based predictions for change in protein thermal stability upon point mutations $\Delta\Delta G$ are accurate only to within ca. 1 kcal mol^{-1}.[47] Thus, it can be assumed that erMD restraints serve as a (partial) correction for small errors inherent in the standard force fields, rather than produce an unreasonably large new energy term.

In this connection, it is also instructive to compare erMD method to other types of restrained simulations. In the erMD protocol, the pseudoforce acting on an individual heavy atom in a given protein molecule is proportional to $2k_0|\overline{\tilde{\mathbf{x}}_i^{(q)MD}} - \tilde{\mathbf{x}}_i^{\mathrm{cryst}}|$ (see Supporting Information). Hence, the value of k_0 is directly comparable to the force constants associated with NOE restraints in the context of protein structure refinement. In the explicit-solvent refinement protocols, k_{NOE} is typically set to 30–50 kcal mol^{-1} Å^{-2},[48] which is much higher than the setting k_0=0.1 kcal mol^{-1} Å^{-2} advocated in this work. It is also important to keep in mind that in our approach the force is only generated when the average coordinates $\overline{\tilde{\mathbf{x}}_i^{(q)\mathrm{MD}}}$ deviate from the crystallographic template. A structural fluctuation in one individual protein molecule generates very little force. From this perspective, $U_{\mathrm{restraint}}$ implemented in the erMD algorithm should be viewed as a "gentle" version of distance restraint.

At this point, we reaffirm the choice of k_0=0.1 as the recommended setting for the erMD simulations. This choice leads to the reasonable value of rms deviation between the ensemble-average protein coordinates and the target crystallographic structure. It also yields a relatively low value of crystallographic R factor. Other things being equal, we favor the low value of $E_{\mathrm{restraint}}$ as found in the erMD simulations with k_0=0.1; low restraint energy ensures that the simulated system retains its native-like dynamics. In what follows, we validate the erMD (k_0=0.1) approach, primarily focusing on comparison with the traditional uMD simulations.

Chemical shifts

Chemical shifts were computed by processing protein coordinates using the prediction program SHIFTX2.[38] In this program, the module SHIFTX+ deals with the conformational dependence of chemical shifts while SHIFTY+ relies on sequence homology. To elucidate the dependence of chemical shifts on protein structure / dynamics, we limited the analyses to SHIFTX+. In the case of conformational ensembles and MD trajectories, the results are averaged over multiple conformers or MD frames.

When the static high-resolution structure 3ONS is used to predict chemical shifts, the rms deviations from the experimental ssNMR shifts amount to 2.39, 0.75, and 1.11 ppm for ^{15}N, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$, respectively. This is very much in line with the typical performance demonstrated by SHIFTX+.[38] Of note, when 1UBQ is used as a structural model, the quality of predictions clearly deteriorates (see Table I). This is a significant result—it provides an independent confirmation that 3ONS is indeed a superior model for analyses of ssNMR data.

When solution-state MD trajectory is used as an input for chemical shift calculations, the quality of the predictions proves to be rather poor. Turning to unrestrained solid-state MD trajectory, $k_0=0$, improves the situation somewhat. Further improvement is obtained using a weakly restrained solid-state trajectory, $k_0=0.1$. At this stage, the quality of the predictions is comparable to that obtained with the static structure 1UBQ. Strengthening the restraints to $k_0=1$ and then to $k_0=10$ leads to further incremental improvements. The comparison of ^{15}N chemical shifts on per-residue basis is illustrated in Supporting Information, Figure S1—there is good overall agreement between δ_{calc} and δ_{exptl}, with several residues showing distinct improvement in going from $k_0=0$ to $k_0=0.1$. We conclude that our erMD strategy leads to a better, more realistic representation of the protein crystal, most likely reflecting the improvements in the average protein structure (cf. first column in Table I).

Interestingly, even though the erMD simulations lead to the average protein coordinates in close agreement with 3ONS (rms deviation 0.2 Å or less), the quality of δ_{calc} still falls somewhat short of what is obtained using the original static crystallographic structure. Naively, one may expect just the opposite—indeed, not only the average coordinates are faithfully reproduced in the erMD simulations, but also the local dynamics is successfully modeled (see below). What is the reason for this less-than-perfect outcome?

SHIFTX2, just like other chemical shift prediction programs, has been trained on static crystallographic structures and solution chemical shifts. Here, we apply SHIFTX2 to the snapshots from MD trajectories with the goal to reproduce solid-state chemical shifts. Thus, strictly speaking, the program is used outside its domain of validity. We believe that this explains the relative underperformance of the prediction algorithm.

Generally, the prediction program which is trained on high-resolution crystallographic structures would likely produce the best results when applied to another high-resolution crystallographic structure. In doing so, the atomic fluctuations, that are strongly structure-dependent,[49] are likely taken into consideration in implicit fashion. From this perspective, the use of an MD model as an input for chemical shift prediction programs probably leads to double counting of the local protein dynamics. As a consequence, the MD models can match the level of δ_{calc} accuracy demonstrated by high-quality crystallographic structures, but cannot significantly outperform them.[50,51]*

Order parameters

Finally, let us turn to discussion of the dipolar order parameters, S^2, as listed in the right-most column of Table I. These parameters have been computed using the orientational dependence of ^{15}N-$^{1}H^{N}$ vectors as extracted from the MD trajectories. The ubiquitin coordinates were used "as is," subject only to crystal symmetry transformations. In this manner, the extracted S^2 values reflect both local protein dynamics and small-amplitude rocking motion of the protein as a whole (with protein molecules embedded in the crystal lattice).[54] The inspection of the data in Table I shows that unrestrained MD trajectory leads to $S^2_{i,calc}$ values that are appreciably different from $S^2_{i,exptl}$, as manifested by rmsd 0.056. The situation is to a certain degree improved in the erMD simulation using $k_0=0.1$, rmsd 0.043. Strengthening of the restraints does not offer any significant improvement. The meaning and the importance of these results are discussed in the next sections.

Dipolar correlation functions

The survey of Table I suggests that most promising results are obtained in the simulations using weak restraints, $k_0=0.1$. As already discussed, further strengthening the restraints brings ensemble-

*These realizations led to development of the next generation of chemical shift predictors which are trained on MD trajectories and intended for use with MD trajectories.[52,53] We have tested one of these newer predictors, PPM,[53] on all trajectories listed in Table I. As one may expect, PPM-based predictions using static coordinates 3ONS turn out to be poor. Conversely, the predictions using uMD and erMD $k_0=0.1$ trajectories are of similar overall quality to those obtained via SHIFTX2. More specifically, PPM performs somewhat better for $^{1}H^{N}$ chemical shifts, somewhat worse for $^{13}C^{\beta}$ chemical shifts, and on par with SHIFTX2 for ^{15}N and $^{13}C^{\alpha}$ chemical shifts.

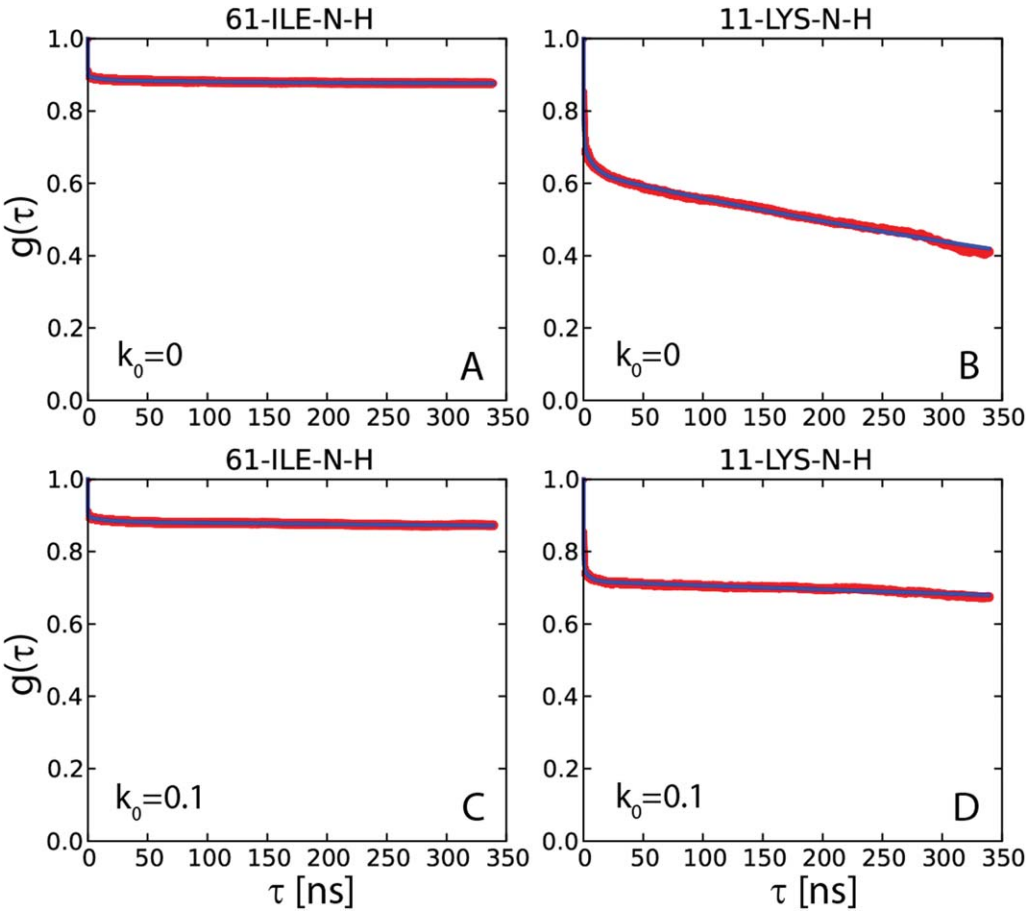

Figure 3. ^{15}N-^{1}H^{N} dipolar correlation functions from two 400-ns-long simulations of crystalline ubiquitin: (a,b) 4U, $k_0=0$ uMD simulation and (c,d) 4U, $k_0=0.1$ erMD simulation. Red profiles represent the numerically calculated MD correlation functions $g_i(\tau)$ (after averaging over 24 ubiquitin molecules found in 4U periodic-boundary box). Blue curves are the result of 4-exponential fitting $g_i^{\mathrm{fit}}(\tau)$, as conducted over the interval from 0 to 85% of the total simulation length. The residue I61 shows typical convergence behavior as observed in the uMD simulation (its convergence parameter Δ corresponds to the median value in the list comprising the simulated data for residues 1–72). The residue K11 shows the worst convergence behavior in the uMD simulation (highest Δ value). All of the obtained correlation functions are remarkably smooth, which reflects good statistical properties of the simulations containing 24 ubiquitin molecules.

average coordinates to within 0.05–0.1 Å of the target crystallographic structure, which is not justified by the accuracy of the crystallographic model. Other measures of quality do not show any significant improvement beyond what is achieved with $k_0=0.1$. In addition, we expect that 4U setup should be preferable to 1U. Conceptually, the erMD method is better suited for large molecular ensembles, where the average coordinates are statistically well-defined. Furthermore, the 4U model should be less vulnerable to potential artifacts associated with periodic-boundary conditions. There are certain indications that this indeed may be the case; in particular, 4U simulations consistently produce lower restraint energies, compared to Table I. Based on all of these observations, we choose to focus on $k_0=0.1$, 4U erMD simulation, comparing it with the conventional 4U uMD simulation. To obtain a better grasp on the issue of convergence, both of these trajectories have been extended from 200 to 400 ns.

Figure 3(a) shows the typical ^{15}N-^{1}H^{N} dipolar correlation function as derived from 400-ns-long 4U uMD simulation. Red curve in the plot represents $g_i(\tau)$ for residue I61 as extracted directly from the MD data (after averaging over 24 ubiquitin molecules contained in 4U periodic-boundary box). The blue curve is the result of least-square fitting using four-exponential function, $g_i^{\mathrm{fit}}(\tau)$. Note that the specifics of the best-fit curve are inconsequential so long as it nicely reproduces the shape of the original correlation function.[55] The plateau of the correlation function is identified with dipolar order parameter. This paves the way for an alternative definition of the order parameter, i.e. it can be equated with the value of g_i^{fit} at the time point corresponding to the full length of the trajectory, $S^2_{i,\mathrm{calc\ alt}}=g_i^{\mathrm{fit}}(t_{\mathrm{traj}})$. This definition is clearly empirical, but we find it useful in the context of the following discussion.

To address the issue of convergence, we have introduced the parameter $\Delta=g_i^{\mathrm{fit}}(t_{\mathrm{traj}})-g_i^{\mathrm{fit}}(2t_{\mathrm{traj}})$.

For those correlation functions that show a well-established plateau, Δ is close to zero. For example, the correlation function shown in Figure 3(a) is characterized by Δ=0.005. This result is representative of the uMD trajectory where most of the correlation functions are well-converged. Specifically, half of the residues in this trajectory display even better convergence properties than I61 (i.e., smaller Δ values).

At the same time, there are several residues in uMD trajectory which lack convergence. The correlation function with the worst convergence properties belongs to residue K11 [shown in Fig. 3(b), Δ=0.15]. The failure to converge is due to rare conformational transitions involving the loop β1–β2 and, to a certain degree, also due to the "structural drift" affecting this region (see Fig. 2). Under these circumstances, the extracted value of the order parameter should be viewed merely as an estimate. The problem cannot be easily resolved—in particular, doubling the length of the trajectory does not help; for instance, using the first 200 ns of the uMD trajectory, we obtain the order parameter $S^2_{i,\mathrm{calc\ alt}}$ = 0.37 for residue K11, whereas using the full-length 400 ns trajectory the value is 0.39. In both calculations, $g_i(\tau)$ fails to reach a plateau [cf. Fig. 3(b)]. A considerably longer simulation would be needed to achieve good convergence for this residue.

It is worth noting, however, that the correlation-function-based order parameters $S^2_{i,\mathrm{calc\ alt}}$ are consistent with $S^2_{i,\mathrm{calc}}$ calculated with the help of Brüschweiler's formula (see footnote in Table I). For example, in the case of residue K11, the calculation based on full-length uMD trajectory yields $S^2_{i,\mathrm{calc}}$ =0.37, consistent with $S^2_{i,\mathrm{calc\ alt}}$ results discussed above. Similar good agreement is found throughout the protein sequence. In this situation, we choose to use $S^2_{i,\mathrm{calc}}$ data for the purpose of further analysis, while relying on parameter Δ to indicate convergence.

Let us now turn to the results in Figure 3(c,d) that illustrate the effect from introducing soft ensemble restraints, k_0=0.1 kcal mol^{-1} Å^{-2}. Characteristically, the correlation function of residue I61 remains unchanged. The order parameter determined for this residue is near-identical to the one previously found in the uMD simulation (in fact it turns out to be slightly lower, 0.86 vs. 0.87). This is generally the case for most residues in ubiquitin, where uMD and erMD simulations produce identical or near-identical results. Conversely, the behavior of residue K11 has undergone a significant change, cf. Fig. 3(b) and 3(d). Although the order parameter remains relatively low, $S^2_{i,\mathrm{calc}}$ =0.68, the slowly decaying component of the correlation function is less pronounced, Δ=0.04. In general, the picture emerging from Figure 3(d) is that of a mobile loop with motions mostly on subnanosecond time scale, plus presumably a certain limited amount of μs-time-scale dynamics (cf. the remaining downward trend in $g_i(\tau)$, as seen in the plot). As it turns out, this picture is largely consistent with the available experimental evidence (discussed below).

Order parameters (continued)

The survey of the results in Table I suggests that 4U, k_0=0.1 erMD simulation achieves a better agreement with experimental order parameters compared to the equivalent uMD simulation (rmsd 0.040 vs. 0.062). Extending both trajectories from 200 to 400 ns does not change this result (rmsd 0.039 vs. 0.065). To appreciate the significance of these improvements, let us compare the S^2 values on per-residue basis. Figure 4(a) shows $S^2_{i,\mathrm{calc}}$ data as obtained from the crystal uMD simulation k_0=0 (blue symbols) in comparison with the recent experimental results by Haller and Schanda[41] (red symbols).

Generally, good agreement is observed on per-residue basis, although computed values tend to be slightly higher than the experimental ones. However, the plot also reveals one major problem area, loop β1–β2, where molecular dynamics seriously exaggerates the amount of backbone motion. Other areas with significant discrepancies are the boundary between α2 and β3, the turn following β4, and the terminal residue in β5. Of note, all the affected regions coincide with the areas of dynamic instability. The residues following glycines, for example, K11 and R54, are especially problematic. The corresponding correlation functions tend to be poorly converged [cyan circles in Fig. 4(a)], which is indicative of μs-time-scale motions. Experimentally, all these sites stand out, featuring elevated R_2 rates and in some cases direct evidence of millisecond dynamics.[24,41]

Introducing ensemble restraints which act on the average protein structure leads to better overall agreement with the experiment, Figure 4(b) (k_0=0.1). Importantly, most of the calculated order parameters remain virtually unchanged. Specifically, for 40 residues the $S^2_{i,\mathrm{calc}}$ values derived from erMD and uMD simulations fall within 0.01 of each other. Furthermore, for 29 residues the order parameters derived from the ensemble-restrained trajectory are actually slightly lower than their uMD counterparts. Hence, we conclude that the native-like local dynamics is largely preserved in the erMD simulations.

For those sites where uMD simulation shows poor agreement with the experiment, the erMD achieves a significant improvement. The most pronounced improvement is observed for β1–β2 loop, specifically for residues G10 and T12. With regard to K11, one has to keep in mind that (i) ssNMR relaxation dispersion measurements showed that K11 signal is broadened by an exchange process on the time scale <100 μs;[24] (ii) K11 is one of those rare residues where solid-state $S^2_{i,\mathrm{exptl}}$ is significantly

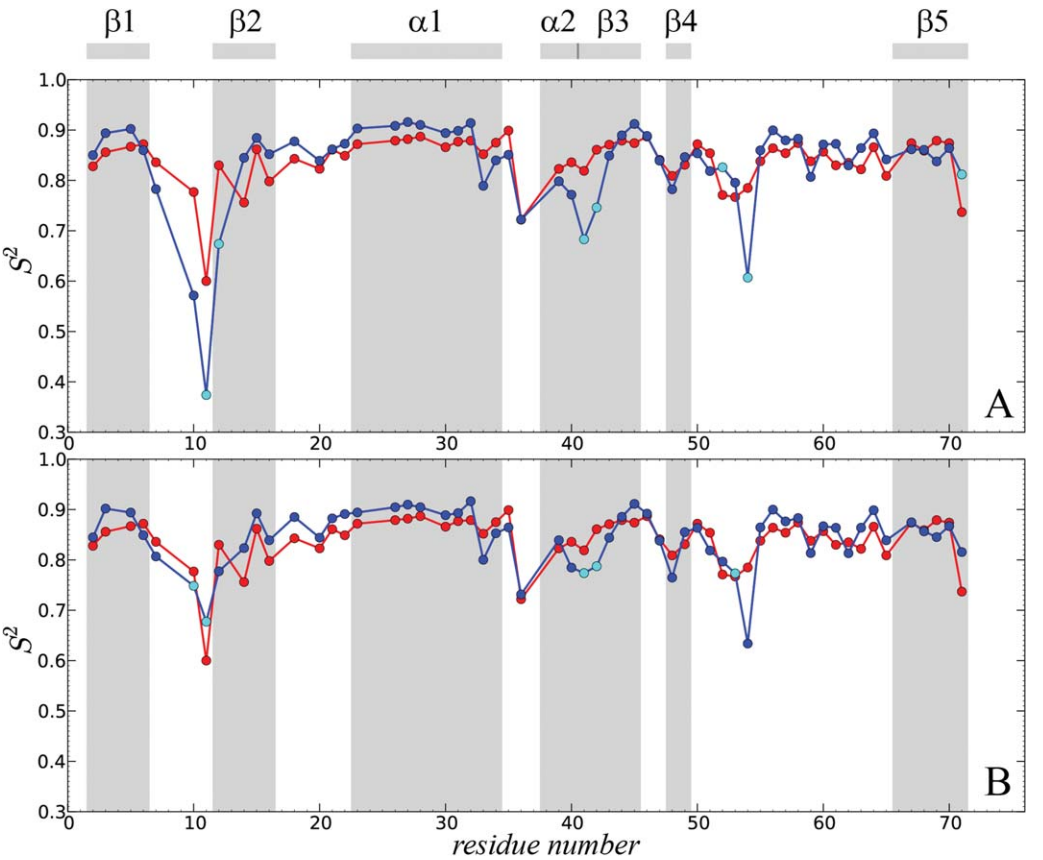

Figure 4. Comparison of the experimental and predicted ^{15}N-$^{1}H^{N}$ dipolar order parameters in crystalline ubiquitin. Experimental data (red symbols) are from Haller and Schanda.[41] The simulated data (blue symbols) are from: (A) the uMD simulation, $k_0=0$, and (B) the erMD simulation, $k_0=0.1$ kcal mol^{-1} Å^{-2}. Each MD trajectory involves a block of four crystal unit cells (4U, 24 ubiquitin molecules) and has a total duration of 400 ns. The residues for which the correlation function lacks convergence, $\Delta > 0.03$, are indicated by cyan filled circles. Of note, the MD-derived correlation functions for residues 72–76 also lack convergence; for these residues we have no experimental data since their signals are absent from the ssNMR spectra (presumably due to slow motions). The secondary-structure regions are represented by the shaded areas and labeled at the top of the plot.

lower than solution-state $S^2_{i,\mathrm{exptl}}$;[41] (iii) similarly, RDC-based $S^2_{i,\mathrm{exptl}}$ for K11 in solution is substantially lower than the relaxation-based $S^2_{i,\mathrm{exptl}}$;[56] (iv) the adjacent residues L8 and T9 are both unobservable in the ssNMR experiment due to exchange broadening.[23] In agreement with all these observations, the erMD correlation function for K11 contains a slowly-decaying component [characteristic time about 6 μs, see Fig. 3(d)]. To obtain a better handle on microsecond motions involving K11, one would need to record a considerably longer erMD trajectory.[57] It is likely that such extended simulation would lead to even better agreement with the experimental result.

Another area where erMD simulation produces partial improvement is the stretch of residues 52–54 which interconverts between type II and type I β-turn conformation. Severe line broadening due to μs time scale conformational exchange has been observed in residue G53 in solution, while T55 displays a moderate amount of broadening both in solution and in solid.[24,58,59] We have scanned the trajectory for the evidence of transitions between type II and type I conformations (the indicative angles are ψ in D52 and ϕ in G53[59]). Although the current simulation is relatively short, 400 ns, it contains 24 ubiquitin molecules, thus offering respectable statistics. In the erMD trajectory, we have found four transitions between type II and type I conformations.[†] These transitions are responsible for the slowly-decaying component in the correlation function of G53, which has characteristic time of about 6 μs [cf. Fig. 4(b), where this residue is classified as lacking convergence]. The presence of μs dynamics at this site is consistent with the experimental data.

Of note, erMD simulation produces small but appreciable decrease in the order parameters for residues D52 and G53, along with a small increase for

[†]The uMD trajectory features no such transitions, although one of the molecules converts into type I conformation during the equilibration stage.

R54, resulting in better agreement with experiment. This is an instructive example which demonstrates that ensemble restraints do not necessarily reduce the amount of motion in the system; on the contrary, sometimes the amount of dynamics is increased. This can be readily understood from a thermodynamic perspective. For intrinsically unstable regions, such as the discussed β turn in ubiquitin, small imperfections in the force field (on the order of 1 kcal mol^{-1}) can significantly alter the population balance between two or more local conformations, resulting in underestimation or overestimation of the order parameters. This is partially corrected by the ensemble restraints, which effectively play the role of empirical force-field corrections.

The findings presented in this section are nontrivial. The restraints implemented in our study are aimed at the average structure of the multiple ubiquitin molecules in the crystal unit cell(s). *A priori*, it is not clear what may be the effect of these restraints on local protein dynamics. In the worst-case scenario, the dynamics may be "stifled," resulting in exceedingly high S^2_{calc} values. Contrary to any such expectations, the modeling of local dynamics is actually preserved and even improved. This result can be viewed as a strong validation of the erMD strategy—the method which relies on structural restraints is validated by the "orthogonal" dynamics data.

In this context, it is also interesting to discuss the relationship between solid- and solution-state order parameters. We have previously compared the two sets of order parameters for α-spc SH3 domain, demonstrating a high degree of correlation on per-residue basis.[17] Here, we present a similar comparison for ubiquitin, Supporting Information, Figure S2. The agreement on per-residue basis proves to be very good, with low rmsd of 0.035. Thus, the solution $S^2_{i,\mathrm{exptl}}$ data provide a strong endorsement for their solid-state counterparts. These results also shed additional light on the role of the so-called supra-τ_c dynamics, that is, internal protein motions on the time scale longer than the protein tumbling time.[56] The comparison of solid- and solution-state data from α-spc SH3 previously led us to conclude that supra-τ_c motions are relatively rare and localize in loop regions or near termini, whereas the structured elements of the protein scaffold remain unaffected.[17] The results from the other small globular protein, ubiquitin, are consistent with this view (see Supporting Information, Fig. S2).

Crystallographic B factors

An additional opportunity to validate the results of MD simulations is provided by crystallographic *B* factors. *B* factors are in a certain sense complementary to dipolar order parameters as they are sensitive to translational displacements of the individual atoms. To compute *B* factors, all protein molecules in the MD trajectory are superimposed via symmetry transformation and then centered at origin according to Eq. (2.1). As a next step, the average coordinates of each protein atom are calculated, $\tilde{\mathbf{x}}_{i,\mathrm{av}} = <\tilde{\mathbf{x}}_i^{(q)\mathrm{MD}}>$, where overbar denotes averaging over N_{prot} protein molecules and angular brackets indicate the averaging over all frames in the trajectory. Finally, the *B* factors are calculated via mean square fluctuation of the atomic coordinates:

$$B = \frac{8\pi^2}{3}\left\langle \overline{\left(\tilde{\mathbf{x}}_i^{(q)\mathrm{MD}} - \tilde{\mathbf{x}}_{i,\mathrm{av}}\right)^2} \right\rangle \tag{3}$$

The *B* factors calculated in this fashion can be compared with the experimental values as contained in the crystallographic coordinate set 3ONS. One should bear in mind, however, that such comparison is at best semiquantitative. There are several reasons for this:

i. The approximate character of the procedure used to derive *B* factors during the refinement of crystallographic structures. At moderately high level of resolution (1.8 Å in the case of 3ONS), it is standard to assume that atomic fluctuations are isotropic and harmonic, corresponding to the Gaussian probability density. Clearly, these assumptions are crude; in particular, they do not hold well for mobile loops on the surface of the protein and side chains undergoing rotameric jumps.[44,60] The general trend is that the reported *B* factors underestimate the mobility at such sites. Furthermore, various heuristic strategies are used to optimize the *B* factors (e.g., group atomic displacement parameters, similarity restraints, motional models such as TLS and normal mode analyses, etc.[61–64]). This makes the reported *B* factors dependent on the details of the refinement protocol.
ii. In our protocol for calculating the *B* factors (see above), we subtract out the effect of small translational displacements of the protein relative to the unit crystal cell. The vibrations of the crystal lattice are also disregarded. As a result, one can expect that the calculated *B* factors are underestimated. It is safe to assume that the two suppressed motional modes are harmonic. Hence their contributions to the *B* factors should be additive. Thus, one may expect that the *B* factors obtained from the MD trajectory are subject to a certain constant offset, making them systematically underestimated.
iii. Finally, one should keep in mind that all MD simulations have been conducted at the temperature 301 K, whereas the X-ray diffraction data were collected at 100 K. Assuming that the motion is harmonic, *B* factors should scale

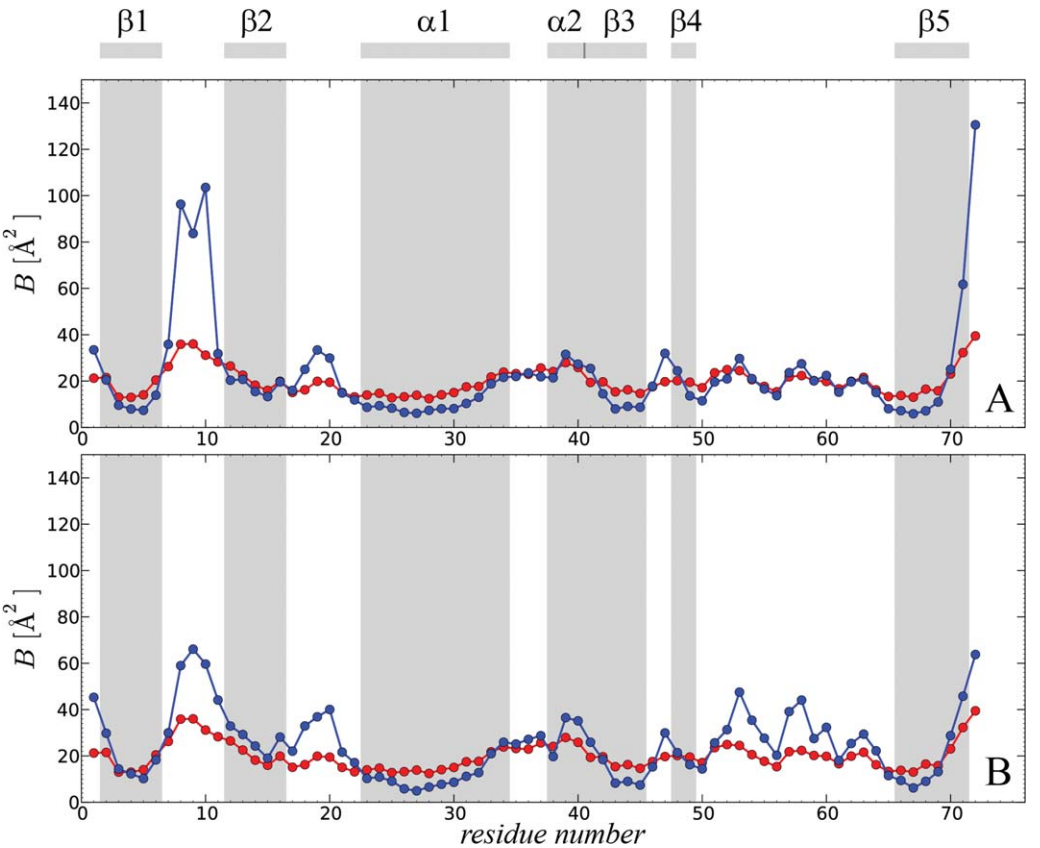

Figure 5. Comparison of the experimental and predicted *B* factors in crystalline ubiquitin. Experimental data (red symbols) are as reported in the coordinate set 3ONS.[25] The simulated data (blue symbols) are from: (A) the uMD simulation, $k_0=0$, and (B) the erMD simulation, $k_0=0.1$ kcal mol^{-1} Å^{-2}. Each MD trajectory involves a block of four crystal unit cells (4U, 24 ubiquitin molecules) and has a total duration of 400 ns.

linearly with temperature.[65,66] There are also examples of crystals where the dependence of *B* factors on temperature is piecewise linear with a transition point.[67–69] Numerous pairs of X-ray structures can be found in the Protein Data Bank where the coordinates of the same protein have been determined at 100 K as well as at ambient temperature, for example, 1U06 and 2NUZ,[70] 1GZR and 1GZZ,[71] and so forth. As expected, at room temperature the *B* factors display a systematic shift toward higher values. By the same token, it can be expected that the *B* factors obtained from the MD trajectory are systematically overestimated. This effect is the opposite of what has been described above, (i) and (ii).

Given all these complications, it is difficult to expect a quantitative agreement between the predicted and experimental *B* factors. Nevertheless, a semiquantitative agreement can usually be obtained.[15] In Figure 5, we present the *B* factors from the crystal structure 3ONS (red symbols) together with the results from uMD and erMD ($k_0=0.1$) simulations (blue symbols). The *B* factors shown in this plot have not been in any way corrected—the values are taken directly from the coordinate set 3ONS or calculated using Eq. (3).

The simulations clearly reproduce the trends seen in the crystallographic study. However, the uMD simulation predicts unreasonably high mobility in the area of β1–β2 loop as well as C-terminal residues 71–72[‡] [see Fig. 5(a)]. In erMD simulation, the amount of motion in these regions is reduced, in line with the experimental data [see Fig. 5(b)]. This change leads to a substantial improvement in the rms deviation between the simulated and experimental data, from 18 to 11 Å^2. Given all reservations about *B* factors expressed above, this result should not be overinterpreted. Nevertheless, it is clear that erMD strategy is broadly successful in reproducing the crystallographic *B* factors. The emerging picture is similar to the one previously obtained from the analysis of S^2 data, leading us to conclude that

[‡]Note that crystallographic coordinates are unavailable for residues 73–76 and ssNMR data are unavailable for residues 72–76.

erMD approach offers an improved description of the local protein dynamics.

Of interest, outside the area of β1–β2 loop and C-terminus, the *B* factors obtained from the erMD simulation tend to be somewhat higher than their uMD counterparts [cf. Fig. 5(a,b)]. As it turns out, this is the consequence of small-amplitude rotational dynamics (rocking motion) which is somewhat more pronounced in the erMD simulation. To quantify this effect, we recalculated the *B* factors such as to eliminate the effect of rotational fluctuations[§]. The results of these alternative calculations are shown in Supporting Information, Figure S3. This latter graph demonstrates a very good agreement between the *B* factors derived from uMD and erMD simulations, except in the area of β1–β2 loop and C-terminus where erMD achieves big improvements and two other sites where minor improvements are obtained. Furthermore, the erMD predictions are in very good agreement with the experiment (up to a scaling factor). Thus, the internal protein dynamics is indeed faithfully captured by the erMD simulation.

The calculation illustrated in Supporting Information, Figure S3 also provides an insight into the amount of orientational disorder in the uMD and erMD trajectories. The mean amplitude of orientational fluctuations experienced by ubiquitin molecules in these two trajectories equals 4.1° and 4.5°, respectively. These are small rotations that have virtually no effect on ssNMR order parameters. However, they can generate up to ca. 1 Å linear displacements for certain protein atoms and thus produce appreciable contributions to *B* factors. Given the limitations (i–iii) discussed above, it is difficult to further clarify the extent of orientational disorder in this system.

^{15}N R_1 rates

Both order parameters and *B* factors are a measure of motional amplitudes. In contrast, ^{15}N spin relaxation rates depend not only on amplitudes, but also on motional time scales. It is generally more challenging for MD simulations to correctly reproduce motional correlation times than it is to recover the amplitudes. When simulating ^{15}N relaxation rates in solution, it is customary to adjust protein overall tumbling time τ_{rot} by setting it equal to the experimentally determined value. This ensures a good level of agreement between the simulated and the experimental rates. In solids—where ^{15}N relaxation is controlled by local motions—there is no such readily available option. Furthermore, it is not known *a priori* if erMD simulations preserve the time scale of local dynamics. One may imagine that restraints lead to stiffening of the system, thus causing a shift toward faster motions. To test this aspect of the erMD model, we turn to the analysis of ^{15}N relaxation data.

The ^{15}N R_1 and $R_{1\rho}$ relaxation rates in crystalline ubiquitin (same form as 3ONS) have been measured at multiple fields by Schanda *et al*. [23] The $R_{1\rho}$ data are not well-suited for the purpose of comparative analysis. Indeed, transverse relaxation rates are a function of the spectral density at zero frequency and thus are highly sensitive to slowly-decaying components of the correlation functions. Given the lack of convergence which has been observed for a number of residues, Figure 3, and the fact that many of the sites are affected by μs motions,[24] we are not in a position to accurately predict $R_{1\rho}$ rates on the basis of the current relatively short MD trajectories. In contrast, R_1 rates are well-suited to draw a comparison between the simulation and experiment. In crystalline samples, ^{15}N R_1 rates are sensitive to the range of motions from about 10 ps to about 100 ns,[72] which is reasonably well-sampled in our MD simulations.

Shown in Figure 6 is the comparison between the experimental and simulated ^{15}N R_1 rates at static magnetic field strength 11.74 T (proton frequency 500 MHz). The experimental dataset is relatively sparse, 35 residues; in particular, it contains no data from residue K11. At the same time, the measurements are fairly precise—the average uncertainty is estimated to be 7%. The erMD simulation has better success in reproducing the experimental data than uMD, as confirmed by the respective rms deviations, 0.023 versus 0.037 s^{-1}. The decrease in rmsd is mainly due to a single residue, G10. In addition, the erMD simulation seems to better reproduce the experimental R_1 profile. Even if G10 is removed from the dataset, the erMD-derived rates show a reasonably strong correlation with the experimental data, $r=0.68$. For uMD simulation, the result is somewhat worse, $r=0.63$.

Similar comparison for data collected at 14.09 T (proton frequency 600 MHz) is illustrated in Figure 7. This data set includes a greater number of residues, 50. However, the measurement error is substantial, on average 13%.[23] The agreement with experiment is not as good as previously found with 11.74 T data. The rms deviation between the simulated and experimental rates is 0.047 s^{-1} for uMD simulation and 0.054 s^{-1} for erMD simulation. The uMD trajectory, therefore, appears to be somewhat more successful. The difference, however, is due to one single residue, K11. Importantly, this residue shows an anomalous dependence of R_1 on static magnetic field strength for which we have no satisfactory explanation (see

[§]Toward this end, we implemented the protocol where all ubiquitin molecules from the MD frames were superimposed onto 3ONS in the least-square sense (via secondary-structure C^{α} atoms). The resulting superposition was then used to calculate *B* factors according to Eq. (3).

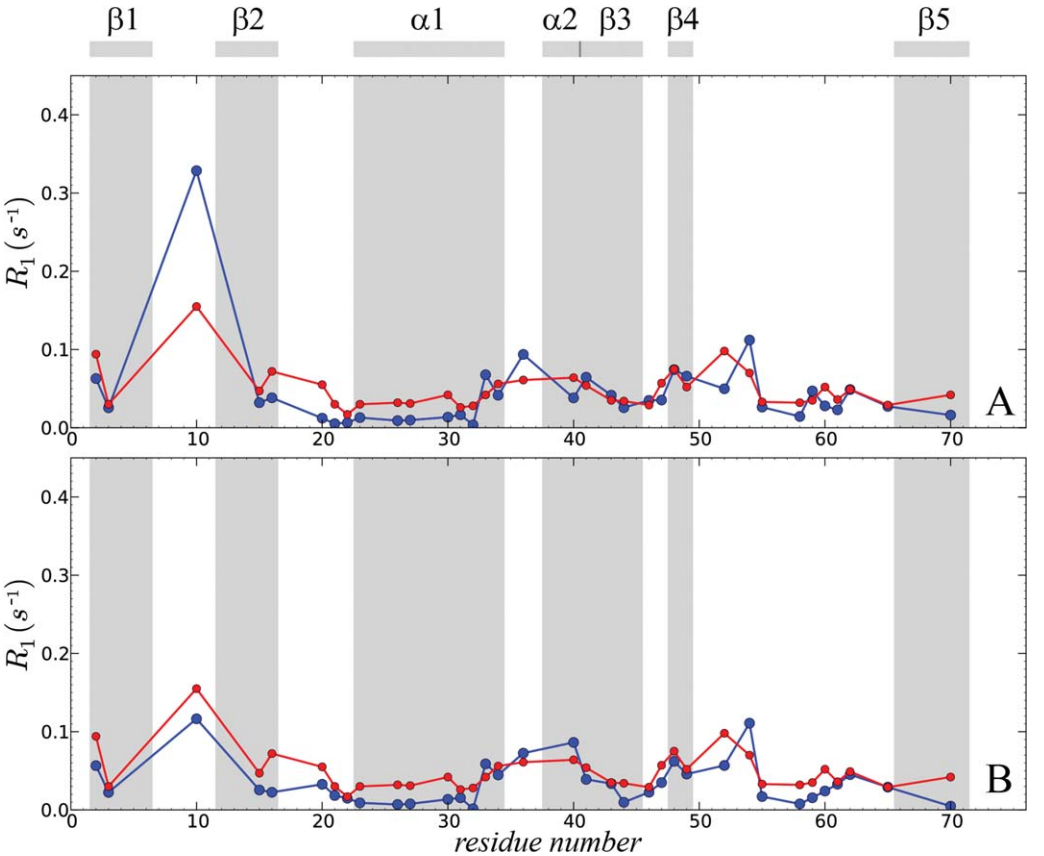

Figure 6. Comparison of the experimental and predicted ^{15}N R_1 relaxation rates in crystalline ubiquitin at static magnetic field strength 11.74 T. Experimental data (red symbols) are as reported by Schanda *et al*.[23] The simulated data (blue symbols) are from: (A) the uMD simulation, k_0=0, and (B) the erMD simulation, k_0=0.1 kcal mol^{-1} Å^{-2}. Each MD trajectory involves a block of four crystal unit cells (4U, 24 ubiquitin molecules) and has a total duration of 400 ns.

below). If this data point is excluded, the results slightly favor erMD simulation over uMD (rmsd of 0.042 and 0.045 s^{-1}, respectively).

Finally, the results at 19.96 T (850-MHz proton frequency) are illustrated in Supporting Information, Figure S4. This data set is comprised of 54 residues; the error is on average 10%. The rms deviations between the simulated and experimental rates are 0.090 and 0.107 s^{-1} for uMD and erMD simulations, respectively. The substantial rmsd values are due to one single residue, K11. Of note, both uMD and erMD trajectories cannot successfully reproduce the experimental R_1 rate for this particular residue (experimental rate 0.90 s^{-1}, uMD rate 0.26 s^{-1}, erMD rate 0.14 s^{-1}). As already pointed out, the experimental data for K11 display an unusual field dependence.[23] Specifically, the ^{15}N R_1 rate for this residue increases from 0.41 ± 0.08 s^{-1} at 600-MHz spectrometer frequency to 0.90 ± 0.10 s^{-1} at 850 MHz. Based on what we know about nitrogen relaxation, there is no good explanation for this result (the CSA relaxation mechanism alone is insufficient to explain 2-fold increase in R_1 rate). It can be suggested that the data involving K11 are contaminated by some sort of experimental error, which may be nontrivial and worthy of further investigation. From our perspective, it is fair to discount or disregard this particular piece of data. With this provision, the performance of erMD model is at least as good, and possibly better than that of the uMD model (cf. Figs. 6 and 7). This result provides a strong validation for the erMD strategy developed in this work.

Concluding Remarks

The applicability of the erMD method is contingent on the assumption that X-ray coordinates faithfully reproduce the average protein structure. Clearly, the very existence of the X-ray coordinates rules out the presence of extensive dynamics such as occurs in disordered proteins. Those elements of the structure that are highly dynamic (e.g., mobile loops or termini) are normally absent from the crystallographic models, so that no restraints are imposed on these mobile fragments (in this sense the erMD approach is "self-regulated"). Likewise, we propose not to impose any restraints on the side chains solved with alternate conformations.

For the major portion of the protein structure, it is safe to assume that the average protein

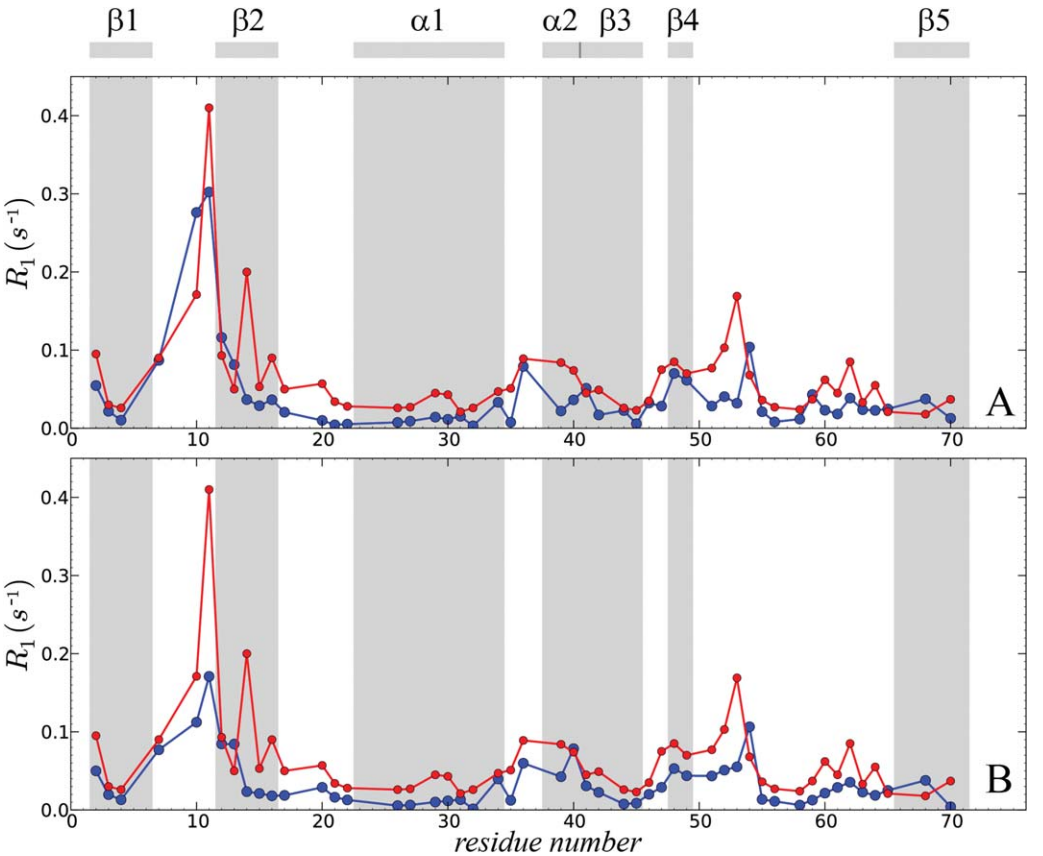

Figure 7. Comparison of the experimental and predicted ^{15}N R_1 relaxation rates in crystalline ubiquitin at static magnetic field strength 14.09 T (same plotting conventions as in Fig. 6).

coordinates fall within 0.2–0.3 Å of the high-resolution crystallographic model, which means that erMD approach is fundamentally sound. One caveat, however, is that X-ray coordinates may correspond to the lowest-energy structure, which is not necessarily the same as the average structure. In other words, X-ray coordinates may reproduce the "ground state" of the protein, while ignoring the "excited states" (i.e., the states with locally different conformations that are populated at the level less than ca. 10%). In the context of our study, we do not see this as a major problem. Given that the restraints used in erMD protocol are weak, we believe that individual protein molecules can sample various excited states without incurring any significant energy penalty [cf. Fig. 2(c)].

In this work, we have tested the restraint coefficients of 0, 0.1, 1, and 10 kcal mol^{-1} Å^{-2} and concluded that the most meaningful results are obtained with k_0=0.1. This is admittedly a rather *ad hoc* and coarse-grained approach. Ideally, we would like to fine-tune the restraint force using a certain measure of quality that is independent of the observables that are used to validate the erMD method. However, any such exercise would require at least ca. 10 different protein systems; the results obtained from ubiquitin alone would be of limited value. Given the scarcity of such systems (i.e., small globular proteins thoroughly characterized by ssNMR), this task would be rather demanding, not to mention computationally expensive. Here, we adopt a more qualitative approach, where we demonstrate the feasibility of the erMD method employing weak restraints. The choice of restraint force is dictated primarily by rmsd to crystallographic target and crystallographic R factors, as well as restraint energies. These metrics point toward k_0=0.1 as the most reasonable option. Other types of data have been used to validate the results. In particular, crystallographic B factors and solid-state ^{15}N R_1 rates have been included *post factum* (when the manuscript was under revision).

The use of the erMD method is contingent on the existence of crystallographic coordinates. This implies that we can only expect to see a limited amount of dynamics in the erMD trajectories. This is in contrast to more general possibilities offered by conventional MD simulations (assuming for a moment that force field is not an issue). Despite such limitations, the new method can provide valuable insights into functionally important forms of protein motion. Relatively recently, Lange *et al.* presented a structural ensemble of ubiquitin which samples a multitude of conformational states

including those observed in 46 different crystal structures.[56] The analysis of this ensemble revealed a dominant motional mode which controls ligand binding via conformational selection mechanism; it also helped to explain the low entropic cost of binding. Later Long and Brüschweiler[73] as well as Fenwick *et al.*[43] used MD simulations to further probe the mechanisms of molecular recognition in ubiquitin, including allosteric effects, cooperative transitions, and formation of an encounter complex. It is anticipated that such studies can benefit from use of the new erMD methodology.

This work draws its inspiration from several sources. A number of ensemble simulations employing solution-NMR restraints have been reported in recent years.[43,56,74–83] A considerable body of work has also been published on "ensemble refinement" of X-ray crystallographic structures.[84–90] Almost all of these simulations, however, consist of short simulated-annealing runs; others are replica-exchange simulations involving high temperatures. In all of these studies, the intention has been to generate structural models with a modicum of conformational diversity; none of them sought to produce a realistic (movie-like) picture of protein motion. This sets our erMD strategy apart from the existing body of work in this area. Of note, our approach is suitable for predicting NMR observables that are dependent on motional correlation times, such as ^{15}N relaxation rates.

The erMD method can be readily generalized for globular proteins in solution, where the crystal structure remains a valid structural template. In principle, protein structure in solution need not necessarily be the same as the X-ray structure obtained from the crystalline sample. Nevertheless, it is generally accepted that crystallographic coordinates provide the best structural models for (single-domain, globular) proteins in solution which are superior to NMR structures.[91–93] This is particularly evident given that X-ray structures lead to better predictions of chemical shifts, residual dipolar couplings, and other independently measured parameters.[94–98] In the case of solution simulations, we envision a modified version of erMD protocol where multiple simulations are run concurrently, with each simulation representing a single protein molecule in a water box. The overarching restraints are imposed to ensure that the average protein structure remains consistent with the X-ray coordinates.

At this time, the best MD force field potentials cannot match the accuracy afforded by the high-resolution crystallographic structures. This shortcoming has a significant adverse impact on fidelity of protein structure in long MD simulations. From this perspective, the crystallography-based restraints used in this study can be thought of as empirical force-field corrections, which remedy small but not-insignificant defects in the force field.[99] Elimination of the "structural drift" is the key achievement of the new erMD methodology. Importantly, the restraints apply only to the ensemble-average coordinates—individual protein molecules in the simulated crystal cell(s) retain their internal dynamics. The restrained MD trajectories recorded in this manner proved to be markedly superior to the conventional unrestrained MD trajectories—they produce better crystallographic R factors, better B factors, better chemical shift predictions, and better predictions for the motional order parameters S^2. They also predict ^{15}N R_1 relaxation rates that are at least as accurate as those obtained from the uMD simulations. The restrained trajectories are characterized by uniquely accurate (average) structure as well as a faithful rendition of internal dynamics; as such, they may be among the most realistic protein MD simulations so far reported.

Acknowledgments

We are thankful to Andrei Fokine and Paul Schanda for valuable discussions. We acknowledge the kind help of Beomsoo Han who prepared for us the customized version of SHIFTX2 software.

References

1. Raval A, Piana S, Eastwood MP, Dror RO, Shaw DE (2012) Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. Proteins 80:2071–2079.
2. Lee MR, Baker D, Kollman PA (2001) 2.1 and 1.8 angstrom average C$^{\alpha}$ RMSD structure predictions on two small proteins, HP-36 and S15. J Am Chem Soc 123: 1040–1046.
3. Fan H, Mark AE (2004) Refinement of homology-based protein structures by molecular dynamics simulation techniques. Protein Sci 13:211–220.
4. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. Science 334:517–520.
5. Lee MR, Tsai J, Baker D, Kollman PA (2001) Molecular dynamics in the endgame of protein structure prediction. J Mol Biol 313:417–430.
6. Chen JH, Brooks CL (2007) Can molecular dynamics simulations provide high-resolution refinement of protein structure? Proteins 67:922–930.
7. Chopra G, Summa CM, Levitt M (2008) Solvent dramatically affects protein structure refinement. Proc Natl Acad Sci USA 105:20239–20244.
8. MacCallum JL, Hua L, Schnieders MJ, Pande VS, Jacobson MP, Dill KA (2009) Assessment of the protein-structure refinement category in CASP8. Proteins 77:66–80.
9. MacCallum JL, Perez A, Schnieders MJ, Hua L, Jacobson MP, Dill KA (2011) Assessment of protein structure refinement in CASP9. Proteins 79:74–90.
10. Ponder JW, Wu CJ, Ren PY, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T (2010) Current status of the AMOEBA polarizable force field. J Phys Chem B 114: 2549–2564.

11. Mackerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25: 1400–1415.
12. Best RB & Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. J Phys Chem B 113:9004–9015.
13. Li DW, Brüschweiler R (2010) NMR-based protein potentials. Angew Chem Int Ed 49:6778–6780.
14. Stocker U, Spiegel K, van Gunsteren WF (2000) On the similarity of properties in solution or in the crystalline state: a molecular dynamics study of hen lysozyme. J Biomol NMR 18:1–12.
15. Meinhold L, Smith JC (2005) Fluctuations and correlations in crystalline protein dynamics: a simulation analysis of Staphylococcal nuclease. Biophys J 88: 2554–2563.
16. Cerutti DS, Freddolino PL, Duke RE, Case DA (2010) Simulations of a protein crystal with a high resolution X-ray structure: evaluation of force fields and water models. J Phys Chem B 114:12811–12824.
17. Chevelkov V, Xue Y, Linser R, Skrynnikov NR, Reif B (2010) Comparison of solid-state dipolar couplings and solution relaxation data provides insight into protein backbone dynamics. J Am Chem Soc 132:5015–5017.
18. Mollica L, Baias M, Lewandowski JR, Wylie BJ, Sperling LJ, Rienstra CM, Emsley JW, Blackledge M (2012) Atomic-resolution structural dynamics in crystalline proteins from NMR and Molecular Simulation. J Phys Chem Lett 3:3657−3662.
19. Martin RW, Zilm KW (2003) Preparation of protein nanocrystals and their characterization by solid state NMR. J Magn Reson 165:162–174.
20. Igumenova TI, Wand AJ, McDermott AE (2004) Assignment of the backbone resonances for microcrystalline ubiquitin. J Am Chem Soc 126:5323–5331.
21. Lorieau JL, McDermott AE (2006) Conformational flexibility of a microcrystalline globular protein: order parameters by solid-state NMR spectroscopy. J Am Chem Soc 128:11505–11512.
22. Manolikas T, Herrmann T, Meier BH (2008) Protein structure determination from ^{13}C spin-diffusion solid-state NMR spectroscopy. J Am Chem Soc 130:3959–3966.
23. Schanda P, Meier BH, Ernst M (2010) Quantitative analysis of protein backbone dynamics in microcrystalline ubiquitin by solid-state NMR spectroscopy. J Am Chem Soc 132:15957–15967.
24. Tollinger M, Sivertsen AC, Meier BH, Ernst M, Schanda P (2012) Site-resolved measurement of microsecond-to-millisecond conformational exchange processes in proteins by solid-state NMR spectroscopy. J Am Chem Soc 134:14800–14807.
25. Huang KY, Amodeo GA, Tong LA, McDermott A (2011) The structure of human ubiquitin in 2-methyl-2,4-pentanediol: a new conformational switch. Protein Sci 20: 630–639.
26. Kohn JE, Afonine PV, Ruscio JZ, Adams PD, Head-Gordon T (2010) Evidence of functional protein dynamics from X-ray crystallographic ensembles. PLoS Comput Biol 6:e1000911.
27. Juers DH & Matthews BW (2001) Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. J Mol Biol 311(4):851–862.
28. Radaelli PG (2011). Symmetry in crystallography: understanding the international tables, Oxford: Oxford University Press.
29. Piana S, Lindorff-Larsen K, Shaw DE (2013) Atomic-level description of ubiquitin folding. Proc Natl Acad Sci USA 110:5915–5920.
30. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725.
31. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78:1950–1958.
32. Bas DC, Rogers DM, Jensen JH (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. Proteins 73:765–783.
33. Sundd M, Iverson N, Ibarra-Molero B, Sanchez-Ruiz JM, Robertson AD (2002) Electrostatic interactions in ubiquitin: stabilization of carboxylates by lysine amino groups. Biochemistry 41:7586–7596.
34. Cerutti DS, Le Trong I, Stenkamp RE, Lybrand TP (2008) Simulations of a protein crystal: explicit treatment of crystallization conditions links theory and experiment in the streptavidin-biotin complex. Biochemistry 47:12065–12077.
35. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D: Biol Crystallogr 66: 213–221.
36. Fokine A, Urzhumtsev A (2002) Flat bulk-solvent model: obtaining optimal parameters. Acta Crystallogr Sect D: Biol Crystallogr 58:1387–1392.
37. Afonine PV, Grosse-Kunstleve RW, Adams PD (2005) A robust bulk-solvent correction and anisotropic scaling procedure. Acta Crystallogr Sect D: Biol Crystallogr 61:850–855.
38. Han B, Liu YF, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50:43–57.
39. Igumenova TI, McDermott AE, Zilm KW, Martin RW, Paulson EK, Wand AJ (2004) Assignments of carbon NMR resonances for microcrystalline ubiquitin. J Am Chem Soc 126:6720–6727.
40. Brüschweiler R, Wright PE (1994) NMR order parameters of biomolecules: a new analytical representation and application to the Gaussian Axial Fluctuation model. J Am Chem Soc 116:8426–8427.
41. Haller JD, Schanda P (2013) Amplitudes and time scales of picosecond-to-microsecond motion in proteins studied by solid-state NMR: a critical evaluation of experimental approaches and application to crystalline ubiquitin. J Biomol NMR 57:263–280.
42. Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 A resolution. J Mol Biol 194: 531–544.
43. Fenwick RB, Esteban-Martin S, Richter B, Lee D, Walter KFA, Milovanovic D, Becker S, Lakomek NA, Griesinger C, Salvatella X (2011) Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. J Am Chem Soc 133:10336–10339.
44. Vitkup D, Ringe D, Karplus M, Petsko GA (2002) Why protein R-factors are so large: a self-consistent analysis. Proteins 46:345–354.

45. DePristo MA, de Bakker PIW, Blundell TL (2004) Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. Structure 12:831–838.
46. Shirts MR, Pitera JW, Swope WC, Pande VS (2003) Extremely precise free energy calculations of amino acid side chain analogs: comparison of common molecular mechanics force fields for proteins. J Chem Phys 119:5740–5761.
47. Seeliger D, de Groot BL (2010) Protein thermostability calculations using alchemical free energy simulations. Biophys J 98:2309–2316.
48. Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M (2003) Refinement of protein structures in explicit solvent. Proteins 50:496–506.
49. Halle B (2002) Flexibility and packing in proteins. Proc Natl Acad Sci USA 99:1274–1279.
50. Robustelli P, Stafford KA, Palmer AG (2012) Interpreting protein structural dynamics from NMR chemical shifts. J Am Chem Soc 134:6365–6374.
51. Li DW, Brüschweiler R (2010) Certification of molecular dynamics trajectories with NMR chemical shifts. J Phys Chem Lett 1:246–248.
52. Lehtivarjo J, Tuppurainen K, Hassinen T, Laatikainen R, Perakyla M (2012) Combining NMR ensembles and molecular dynamics simulations provides more realistic models of protein structures in solution and leads to better chemical shift prediction. J Biomol NMR 52: 257–267.
53. Li DW, Brüschweiler R (2012) PPM: a side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. J Biomol NMR 54: 257–265.
54. Lewandowski JR, Sein J, Blackledge M, Emsley L (2010) Anisotropic collective motion contributes to nuclear spin relaxation in crystalline proteins. J Am Chem Soc 132:1246–1247.
55. Bremi T, Brüschweiler R (1997) Locally anisotropic internal polypeptide backbone dynamics by NMR relaxation. J Am Chem Soc 119:6672–6673.
56. Lange OF, Lakomek NA, Farès C, Schröder GF, Walter KFA, Becker S, Meiler J, Grubmüller H, Griesinger C, de Groot BL (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 320:1471–1475.
57. Xue Y, Ward JM, Yuwen TR, Podkorytov IS, Skrynnikov NR (2012) Microsecond time-scale conformational exchange in proteins: using long Molecular Dynamics trajectory to simulate NMR relaxation dispersion data. J Am Chem Soc 134:2555–2562.
58. Massi F, Grey MJ, Palmer AG (2005) Microsecond timescale backbone conformational dynamics in ubiquitin studied with NMR $R_{1\rho}$ relaxation experiments. Protein Sci 14:735–742.
59. Sidhu A, Surolia A, Robertson AD, Sundd M (2011) A hydrogen bond regulates slow motions in ubiquitin by modulating a β-turn flip. J Mol Biol 411:1037–1048.
60. Garcia AE, Krumhansl JA, Frauenfelder H (1997) Variations on a theme by Debye and Waller: from simple crystals to proteins. Proteins 29:153–160.
61. Kundu S, Melton JS, Sorensen DC, Phillips GN (2002) Dynamics of proteins in crystals: comparison of experiment with simple models. Biophys J 83:723–732.
62. Poon BK, Chen XR, Lu MY, Vyas NK, Quiocho FA, Wang QH, Ma JP (2007) Normal mode refinement of anisotropic thermal parameters for a supramolecular complex at 3.42-A crystallographic resolution. Proc Natl Acad Sci USA 104:7869–7874.
63. Afonine PV, Urzhumtsev A, Grosse-Kunstleve RW, Adams PD (2010) atomic displacement parameters (ADPs), their parameterization and refinement in PHENIX. Comput Crystallogr Newsletter 1:24–31.
64. Merritt EA (2012) To B or not to B: a question of resolution? Acta Crystallogr Sect D: Biol Crystallogr 68: 468–477.
65. Chong SH, Joti Y, Kidera A, Go N, Ostermann A, Gassmann A, Parak F (2001) Dynamical transition of myoglobin in a crystal: comparative studies of X-ray crystallography and Mossbauer spectroscopy. Eur Biophys J 30:319–329.
66. Schmidt M, Achterhold K, Prusakov V, Parak FG (2009) Protein dynamics of a β-sheet protein. Eur Biophys J 38:687–700.
67. Tilton RF, Dewan JC, Petsko GA (1992) Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at 9 different temperatures from 98 K to 320 K. Biochemistry 31:2469–2481.
68. Teeter MM, Yamano A, Stec B, Mohanty U (2001) On the nature of a glassy state of matter in a hydrated protein: relation to protein function. Proc Natl Acad Sci USA 98:11242–11247.
69. Joti Y, Nakasako M, Kidera A, Go N (2002) Nonlinear temperature dependence of the crystal structure of lysozyme: correlation between coordinate shifts and thermal factors. Acta Crystallogr Sect D: Biol Crystallogr 58:1421–1432.
70. Chevelkov V, Faelber K, Diehl A, Heinemann U, Oschkinat H, Reif B (2005) Detection of dynamic water molecules in a microcrystalline sample of the SH3 domain of α-spectrin by MAS solid-state NMR. J Biomol NMR 31:295–310.
71. Brzozowski AM, Dodson EJ, Dodson GG, Murshudov GN, Verma C, Turkenburg JP, de Bree FM, Dauter Z (2002) Structural origins of the functional divergence of human insulin-like growth factor-I and insulin. Biochemistry 41:9389–9397.
72. Chevelkov V, Zhuravleva AV, Xue Y, Reif B, Skrynnikov NR (2007) Combined analysis of ^{15}N relaxation data from solid- and solution-state NMR spectroscopy. J Am Chem Soc 129:12594–12595.
73. Long D, Bruschweiler R (2011) In silico elucidation of the recognition dynamics of ubiquitin. PLoS Comput Biol 7:e1002035.
74. Kim YM, Prestegard JH (1990) Refinement of the NMR structures for Acyl Carrier Protein with scalar coupling data. Proteins 8:377–385.
75. Bonvin AMJJ, Boelens R, Kaptein R (1994) Time and ensemble-averaged direct NOE restraints. J Biomol NMR 4:143–149.
76. Clore GM, Schwieters CD (2004) How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? J Am Chem Soc 126:2923–2938.
77. Tang C, Schwieters CD, Clore GM (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. Nature 449:1078–1082.
78. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M (2005) Simultaneous determination of protein structure and dynamics. Nature 433:128–132.
79. Allison JR, Varnai P, Dobson CM, Vendruscolo M (2009) Determination of the free energy landscape of α-synuclein using spin label nuclear magnetic resonance measurements. J Am Chem Soc 131:18314–18326.
80. Huang JR, Grzesiek S (2010) Ensemble calculations of unstructured proteins constrained by RDC and PRE

data: a case study of urea-denatured ubiquitin. J Am Chem Soc 132:694–705.
81. Robustelli P, Kohlhoff K, Cavalli A, Vendruscolo M (2010) Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. Structure 18:923–933.
82. Esteban-Martin S, Fenwick RB, Salvatella X (2010) Refinement of ensembles describing unstructured proteins using NMR residual dipolar couplings. J Am Chem Soc 132:4626–4632.
83. Im W, Jo S, Kim T (2012) An ensemble dynamics approach to decipher solid-state NMR observables of membrane proteins. BBA Biomembr 1818:252–262.
84. Kuriyan J, Osapay K, Burley SK, Brunger AT, Hendrickson WA, Karplus M (1991) Exploration of disorder in protein structures by X-ray restrained molecular dynamics. Proteins 10:340–358.
85. Burling FT, Brunger AT (1994) Thermal motion and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. Israel J Chem 34:165–175.
86. Gros P, Van Gunsteren WF, Hol WGJ (1990) Inclusion of thermal motion in crystallographic structure by restrained Molecular Dynamics. Science 249:1149–1152.
87. Clarage JB, Phillips GN (1994) Cross-validation tests of time-averaged molecular dynamics refinements for determination of protein structures by X-ray crystallography. Acta Crystallogr Sect D: Biol Crystallogr 50:24–36.
88. Pellegrini M, Gronbech-Jensen N, Kelly JA, Pfluegl GMU, Yeates TO (1997) Highly constrained multiple-copy refinement of protein crystal structures. Proteins 29:426–432.
89. Levin EJ, Kondrashov DA, Wesenberg GE, Phillips GN (2007) Ensemble refinement of protein crystal structures: validation and application. Structure 15:1040–1052.
90. Burnley BT, Afonine PV, Adams PD, Gros P (2012) Modelling dynamics in protein crystal structures by ensemble refinement. eLife Sci 1:e00311.
91. Doreleijers JF, Rullmann JAC, Kaptein R (1998) Quality assessment of NMR structures: a statistical survey. J Mol Biol 281:149–164.
92. Garbuzynskiy SO, Melnik BS, Lobanov MY, Finkelstein AV, Galzitskaya OV (2005) Comparison of X-ray and NMR structures: is there a systematic difference in residue contacts between X-ray and NMR-resolved protein structures? Proteins 60:139–147.
93. Andrec M, Snyder DA, Zhou ZY, Young J, Montellone GT, Levy RM (2007) A large data set comparison of protein structures determined by crystallography and NMR: statistical test for structural differences and the effect of crystal packing. Proteins 69:449–465.
94. Williamson MP, Kikuchi J, Asakura T (1995) Application of ^{1}H NMR chemical shifts to measure the quality of protein structures. J Mol Biol 247:541–546.
95. Neal S, Nip AM, Zhang HY, Wishart DS (2003) Rapid and accurate calculation of protein ^{1}H, ^{13}C, and ^{15}N chemical shifts. J Biomol NMR 26:215–240.
96. Spronk C, Nabuurs SB, Krieger E, Vriend G, Vuister GW (2004) Validation of protein structures derived by NMR spectroscopy. Prog NMR Spectrosc 45:315–337.
97. Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 12:1–16.
98. Simon K, Xu J, Kim C, Skrynnikov NR (2005) Estimating the accuracy of protein structures using residual dipolar couplings. J Biomol NMR 33:83–93.
99. Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G (2004) Making optimal use of empirical energy functions: force-field parameterization in crystal space. Proteins 57:678–683.

Supporting Information

Ensemble MD simulations restrained via crystallographic data: accurate structure leads to accurate dynamics

Yi Xue[1] and Nikolai R. Skrynnikov[1,2*]

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

[2] Laboratory of Biomolecular NMR, St. Petersburg State University, St. Petersburg 199034, Russia

[*] To whom correspondence should be addressed: Nikolai R. Skrynnikov. Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette IN 47907-2084, USA. E-mail: nikolai@purdue.edu. Phone: 1 (765) 494 8519

MD simulation protocol for ubiquitin crystals

Starting coordinates for the crystal MD trajectory were obtained from the high-resolution crystallographic structure 3ONS.[1] This structure misses four flexible C-terminal residues, which give rise to weak and uninterpretable electron density. To address this issue, we prepared 200 structural models based on 3ONS geometry, where the terminal segments were initially generated in a form of random coil[2] and then grafted onto the body of the protein.[*] Each of these models also included the crystallographic water as found in 3ONS. The resulting constructs were packed into a unit cell (space group $P3_221$, six protein molecules per unit cell) using the appropriate tool in Amber 11. The original dimensions of the cell, $a = b = 48.41$ Å and $c = 61.97$ Å, were all multiplied by a factor 1.016 to account for thermal expansion of the protein crystal upon transition from 100 K (temperature at which 3ONS was solved) to 301 K (temperature at which ssNMR data were taken).[3]

As a next step, the protein coordinates were protonated. To determine the protonation status of individual Asp and Glu residues, we performed the PROPKA[4] calculations for ubiquitin in a crystal-lattice environment. The results were generally consistent with the estimations using solution pK_a,[5] except for several residues experiencing the effect of crystal contacts. Since charged side chains are oftentimes involved in crystal contacts, we believe that it is more appropriate to use the computed pK_a values which explicitly take into consideration the effects of crystal packing. The effective pH was assumed to be 4.2, same as in the crystallization buffer.[1] The system was then neutralized by adding eight Cl^- ions per ubiquitin molecule (forty-eight Cl^- ions per unit cell). The number of water molecules to be added to the crystal unit cell was initially estimated based on the simple density considerations.[6] This number was subsequently adjusted such as to ensure that the volume of the crystal cell remains unchanged during the MD production run. Following a series of iterative corrections, we found that it was necessary to add ca. 1650 water molecules (on top of 546 crystallographic waters already contained in the crystal unit cell). Both chlorine ions and water molecules were added using AddToBox facility[7] in Amber 11.[8] We used the SPC/E water model,[9] which has been recommended as the preferred choice for Amber ff99SB force field;[7] this model also showed the best results in our trial simulations. No attempt was made to include 2-methyl-2,4-pentanediol, glycerol, or sodium citrate, which were also a part of the crystallization buffer.[1] None of these compounds appear in the crystallographic structure 3ONS and it is unclear to what degree they are partitioned into the crystal; also force field parameters are not readily available for some of these molecules.

Additional manipulations were performed to optimize the coordinates of the C-terminal residues in each of the 200 starting models. To emulate the crystal lattice environment, periodic boundary conditions have been applied at the faces of the unit cell. Heavy protein atoms, except those in the four C-terminal residues, were restrained to their original coordinates (force constant 500 kcal mol^{-1} $Å^{-2}$). The system was then energy-minimized via 500 steps of steepest descent,

[*] Specifically, C^{α} and C′ atoms in residue R72 were used as the points of attachment.

followed by 500 steps of conjugate gradient minimization. The minimization was conducted in Amber 11 under control of Amber ff99SB force field with Best and ILDN corrections (ff99SB*-ILDN).[10] Subsequently, the system was heated from 0 to 1000 K and then cooled back to 0 K. In doing so, the temperature was incremented (decremented) with the step of 200 K; total duration of the heating and cooling stages was 40 and 120 ps, respectively. During this stage the heavy atoms were restrained with the force constant 10 kcal mol^{-1} $Å^{-2}$.

The 200 structural models processed according to the above scheme were subsequently ranked by energy. Toward this goal, we stripped the system of water and instead applied the implicit solvent (option igb=5 in Amber).[11] Since Amber does not allow for use of periodic boundary conditions in conjunction with implicit solvent, we have modeled the effect of crystal lattice by assembling a block of three identical unit cells. The resulting construct was once again subjected to the energy minimization, where all heavy atoms were fixed while the protons were optimized. Finally, the energy of the obtained system was evaluated using Amber ff99SB*-ILDN potential with igb=5 solvation. The results were used to rank the 200 models by energy and select 10 lowest-energy models.

Next we return to the optimized models containing explicit solvent, focusing on the subset of 10 models identified in the previous step. Recall that these models essentially reproduce the unit crystal cell as seen in the crystallographic coordinate set 3ONS, but with the addition of the ubiquitin C-terminal tail. The inspection of the 10 selected models demonstrates that the C-tails tends to cluster around two preferred conformations (confirmed by the principal component analysis). To test the effect of the tail conformation we recorded a number of MD trajectories beginning from the different initial models. The results of these simulations proved to be similar, indicating that the tail moves sufficiently freely and samples the entire conformational phase space available to it in the time frame of 100 ns. Therefore we have chosen one single model (the one with the lowest energy $E_{implicit}$) as a starting point for all of the following simulations.

The chosen model was subjected to two final rounds of energy minimization prior to the beginning of the production run. At first, water coordinates were optimized while protein atoms were fixed; then all restraints were lifted and the entire model was minimized. After that the temperature of the system was raised from 0 to 301 K by running 20 ps constant-volume simulation with weak restraints applied to all protein atoms (10 kcal mol^{-1} $Å^{-2}$). Finally, the production run was initiated. The first 20 ns of each trajectory were treated as equilibration stage and subsequently discarded. The MD simulation was run at constant pressure (1 atm) and constant temperature (301 K) using the Langevin thermostat. The constant pressure was maintained using the isotropic scaling option, with pressure relaxation time set to 2 ps. The Langevin collision frequency was 3 ps^{-1}. The non-bonded cutoff was 11 Å; we have also conducted erMD simulations using the cutoff of 9 Å and found the results to be identical. The bonds involving hydrogen atoms were constrained using SHAKE algorithm. The integration step was 2 fs and the protein coordinates were stored every 5 ps. The force field, Amber ff99SB*-ILDN, included additionally the crystallography-based pseudopotential, which is discussed in

detail below. A number of comparative studies, in particular those based on the experimental NMR data, favor Amber ff99SB over other force fields. [12-17]

The crystal MD simulations involved either the single unit cell as described above (1U), or the block of two unit cells (2U, dimension a doubled), or the block of four unit cells (4U, dimensions a and b doubled). The starting coordinates for 2U and 4U simulations were obtained by assembling multiple copies of the 1U cell. The resulting system was then equilibrated as reported above (beginning with the solvent energy minimization).

The volume of the system remained remarkably stable during the NPT simulations. For instance, in the case of the unrestrained ubiquitin simulation (1U) the mean volume was only 0.3% above the target value, with rms fluctuations of 0.2%. In the case of erMD trajectory with $k_0 = 0.1$ the corresponding numbers were 0.1% and 0.2%.

The simulations were conducted using two GPU workstations – one equipped with four NVIDIA GeForce GTX480 cards and the other with four GTX580 cards (assembled by Electronics Nexus, Binghamton NY and Colfax International, Sunnyvale CA, respectively). The production rate using CUDA version of pmemd program was 27 ns per day per card for 1U simulation and 9 ns per day per card for 4U simulation.

Additional crystal simulations

We have been concerned about the role of side-chain charges in those Asp and Glu side chains where pK_a happens to fall close to the presumed interstitial crystal pH. In particular, we focused on residue E34, which is capable of forming a salt bridge with K11 and thus may constrain the motion of the β1-β2 loop. This salt bridge is not found in the coordinate set 3ONS, but it occurs in 1UBQ. The PROPKA calculation using 3ONS yields pK_a 4.5 for residue E34, which is identical to the value experimentally measured in solution.[5] According to the protocol described above, at pH 4.2 this residue is deemed to be protonated (uncharged). However, one needs to bear in mind that there is also a substantial fraction of molecules where E34 is deprotonated (charged). It is reasonable to suggest that charged E34 side chain has a propensity to form a salt bridge with K11, thus constraining the motion of β1-β2 loop. Generally speaking, it would be advisable to model both (co-existing) protonated and deprotonated E34 species. It is conceivable that such modification may "rescue" the conventional uMD simulation, i.e. improve the accuracy of $S^2_{i,\,calc}$.

To test this possibility, we have recorded an additional uMD trajectory (1U, 200 ns), where E34 side chain was deprotonated (charged). The results proved to be virtually identical to the reference trajectory where this side chain was protonated (uncharged). In particular, the rmsd between the simulated and experimental order parameters remains unchanged. Thus the problems with uMD simulation are unlikely to be caused by the charge on E34 side chain.

E34 is not the only residue where the protonation state may present a problem. For instance, hydrogen bond formed by the side chain of E24 is likely to influence the conformation of β-turn 52-54.[18] Generally speaking, modeling the variable protonation states presents a challenge for MD simulations. A number of specialized methods have been developed to address this problem,[19-22] but these methods tend to be computationally expensive. In lieu of such specialized tools, standard MD simulations assume fixed protonation states, which is obviously a relatively crude model. The errors associated with this approach can be to a certain degree alleviated by the proposed erMD method.

Solution MD simulations

Unrestrained MD trajectories of ubiquitin in solution have been recorded as a point of comparison. The simulations were conducted using truncated octahedral water box with the thickness of solvation shell of at least 12 Å. The simulation protocol was the same as for the respective crystals, with the exception of crystal lattice periodicity.

Structure-based restraints

The pseudopotential Eq. (1) can be expressed in the expanded form as follows:

$$U_{restraint} = k_0 N_{prot} \sum_{i=1}^{Natom} \left| \frac{\sum_{q=1}^{Nprot} \widehat{\mathbf{R}}^{(q)} (\mathbf{x}_i^{(q)MD} - \mathbf{v}^{(q)MD})}{N_{prot}} - (\mathbf{x}_i^{cryst} - \mathbf{v}^{cryst}) \right|^2 \quad \text{(S1)}$$

Recall that $\mathbf{v}^{(q)MD}$ defines the center of mass of q-th protein molecule in the MD frame (or, strictly speaking, a geometric center because the masses of heavy atoms are taken to be equal); similarly, $\mathbf{v}^{cryst}$ is the center of mass of the crystallographic structure. The force constant k has a form of $k_0 N_{prot}$ where k_0 is an empirically chosen parameter.

Differentiating this expression with respect to the coordinates of the j-th atom in the p-th protein molecule yields the expression for force:

$$\mathbf{F}_j^{(p)} = 2k_0 \widehat{\mathbf{R}}^{(p)T} \left(\frac{\sum_{q=1}^{Nprot} \widehat{\mathbf{R}}^{(q)} (\mathbf{x}_j^{(q)MD} - \mathbf{v}^{(q)MD})}{N_{prot}} - (\mathbf{x}_j^{cryst} - \mathbf{v}^{cryst}) \right) \quad \text{(S2)}$$

In this expression symbol T indicates the transpose of the matrix (equivalent to inverse). The matrices $\tilde{\mathbf{R}}^{(p)T}$ are the same as the crystallographic symmetry transformation matrices (rotation part, 3×3) listed in the headers of the PDB files. Note that the term $\mathbf{v}^{(q)MD}$ is also dependent on coordinates $x_j^{(p)}$; however, the respective contribution to force is zero. Finally note that the forces applied to individual atoms are proportional to k_0 and do not depend on the size of the simulated system.

Diffraction-based restraints

In addition to the erMD protocol detailed above, we have also implemented an alternative protocol where the restraints are derived directly from the crystallographic structure factors. For this purpose we introduced the pseudopotential:

$$U_{diffraction} = k_0 N_{prot} \frac{\sum_{(h,k,l)} (q\,|\overline{F_{calc}^{(q)}(h,k,l)}| - |F_{obs}(h,k,l)|)^2}{\sum_{(h,k,l)} |F_{obs}(h,k,l)|^2} \qquad \text{(S3).}$$

Here q is the overall scaling factor and other notations are the same as used in the text. Using the "direct summation" formula for $F_{calc}^{(q)}(h,k,l)$,[23] we differentiated this expression with respect to atomic coordinates and thus defined forces (in analogy to standard crystallographic refinement programs). Each ubiquitin molecule in the periodic boundary box was treated as an independent entity, with no assumptions regarding crystal symmetry. The calculation of forces based on Eq. (S3) was implemented in GPU CUDA code and integrated with the Amber 11 simulation engine. The production rate achieved for 1U simulation of crystalline ubiquitin was 14 ns/day.

Conceptually, the idea of erMD simulation based on Eq. (S3) is appealing. Indeed, raw diffraction data contain the information which is both more accurate and more complete than the information that can be found in the derivative crystallographic model. In particular, diffraction data encode more information about the conformational diversity of the system, i.e. internal protein dynamics. Nevertheless, the simulations using this algorithm proved to be unsuccessful. The energy landscape of $U_{diffraction}$ is highly non-local* and therefore extremely rugged. Consequently, the forces associated with $U_{diffraction}$ do not point toward the global minimum (i.e. the true structure), but rather toward a nearby local minimum. In the context of MD simulations, where the protein coordinates constantly change, these forces acquire a quasi-random character: they rapidly fluctuate while pointing in seemingly random directions. This makes them useless or even harmful, since they destabilize the simulation.

One possible *ad hoc* solution in this situation is to calculate time-averaged forces, thus reducing the element of randomness. This strategy has been originally proposed two decades ago[24] and very recently successfully implemented by Gros *et al*. in the context of single-molecule refinement.[25] While such restrained trajectories lead to improved crystallographic

* In other words, the movement of any single atom generates force on all other atoms.

models (conformational ensembles), they cannot be viewed as a realistic representation of protein dynamics. In summary, the potential Eq. (S3) is well suited for minimization algorithms as used in crystallographic refinement, but cannot be easily integrated in *bona fide* MD simulations.

Figures

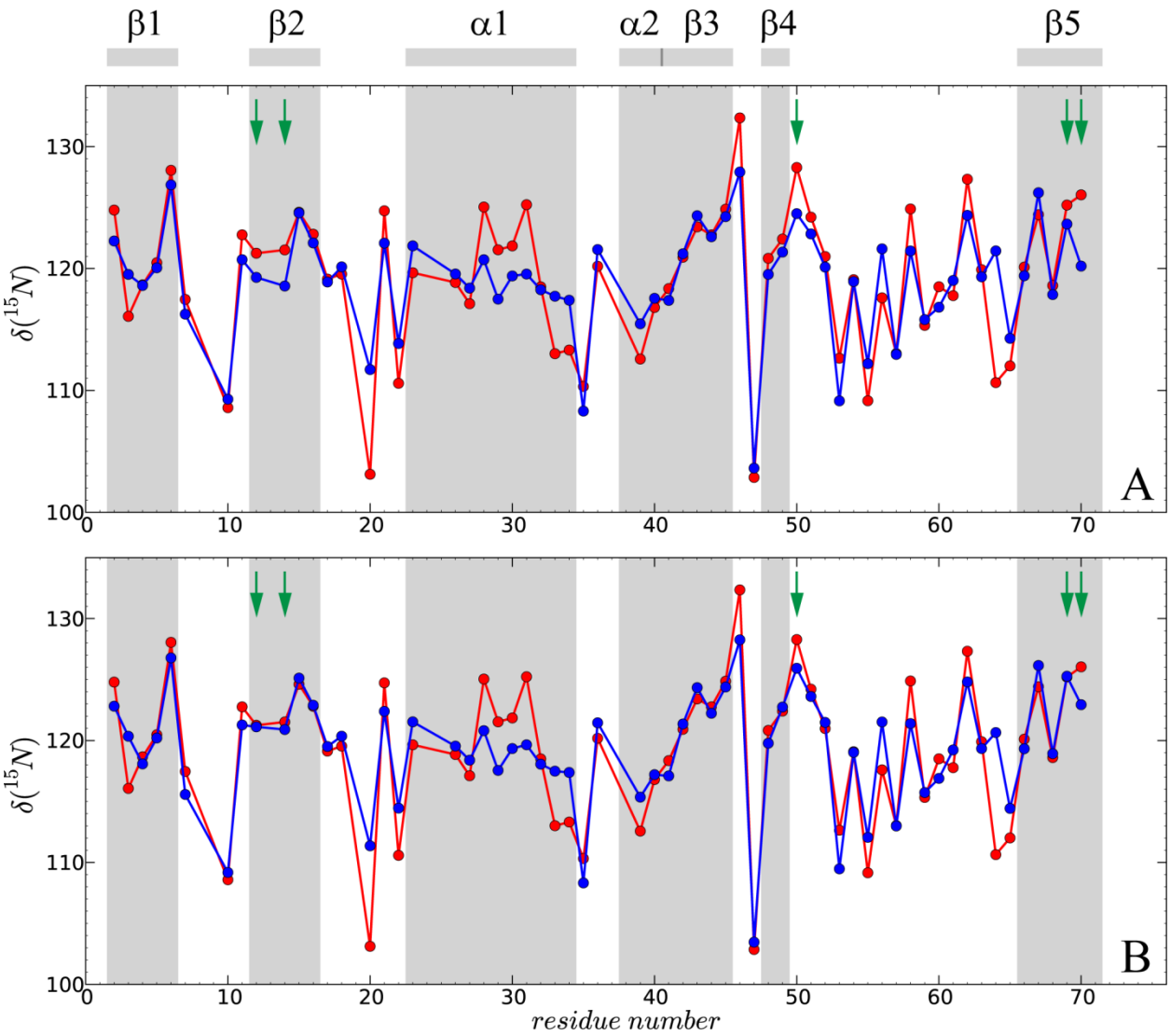

Fig. S1. Comparison of the experimental and predicted ^{15}N chemical shifts in crystalline ubiquitin. Experimental data (red symbols) are from Schanda *et al.*[26] The simulated data (blue symbols) are from application of the program SHIFTX+ [27] to (**A**) the uMD simulation, $k_0 = 0$, and (**B**) the erMD simulation, $k_0 = 0.1$ kcal mol^{-1} Å^{-2}. Each MD trajectory involves a single crystal unit cell (1U, 6 ubiquitin molecules) and has a total duration of 1 μs. The program SHIFTX+ has been customized as described in the text. The sites where erMD-based predictions display the most significant improvement over uMD-based predictions are marked with green arrows.

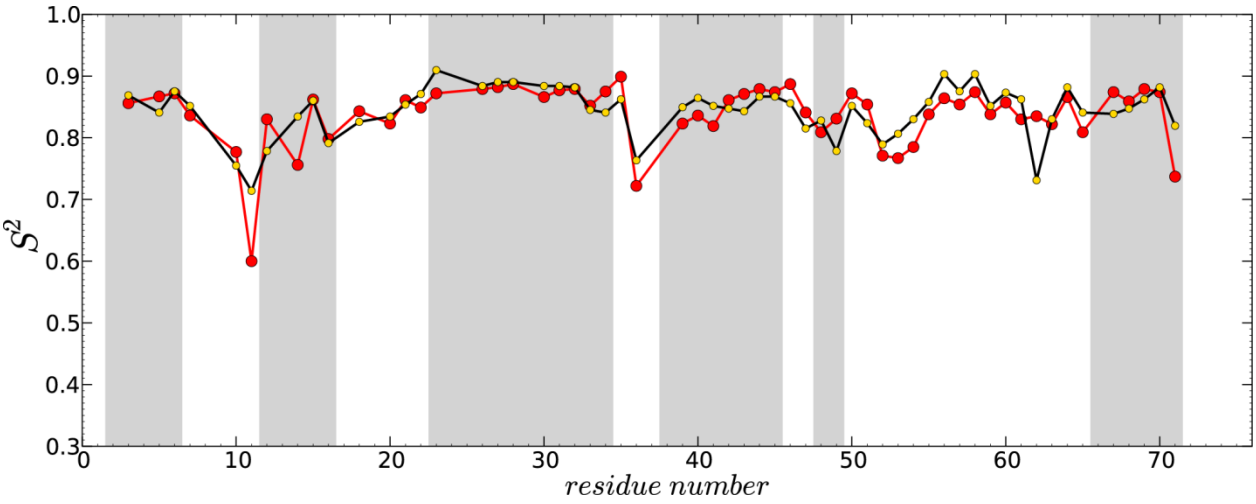

Fig. S2. Comparison of the experimental ^{15}N-^{1}H^{N} dipolar order parameters from crystalline ubiquitin with the experimental ^{15}N-relaxation-based order parameters from ubiquitin in solution. Solid-state data (red symbols) are from Schanda *et al*.[28] Solution data (black & gold) are from Showalter and Brüschweiler,[29] who reinterpreted the original results by Lienin *et al*.[30] The rms deviation between the solution- and solid-state S^2_{exptl} as presented in this plot is 0.035; the Pearson correlation coefficient is 0.73. The conspicuous difference at the site Q62 is likely due to the effect of the crystal contact.

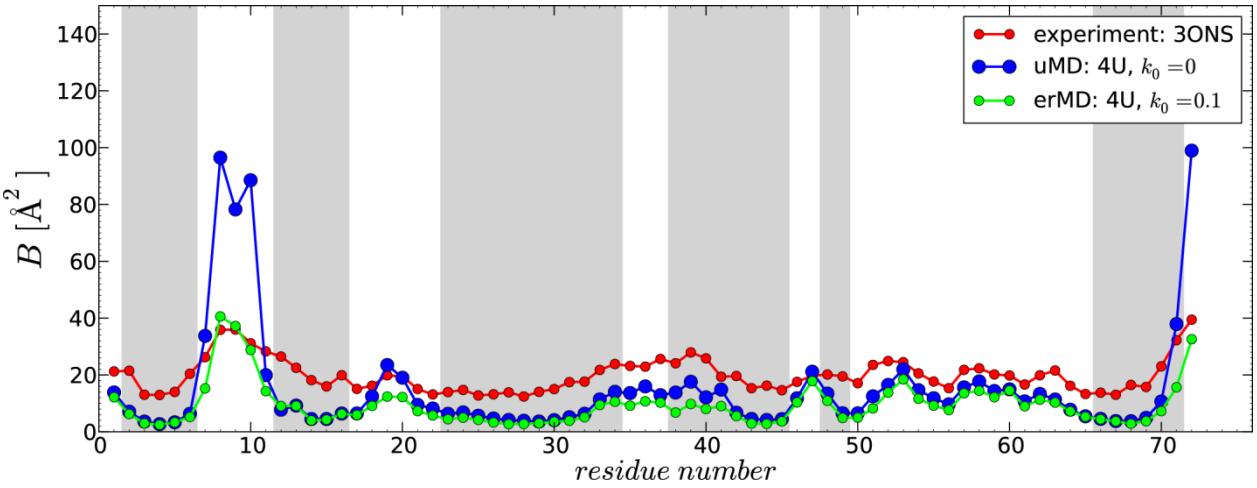

Fig. S3. Comparison of the experimental and predicted B factors in crystalline ubiquitin. The computational protocol has been modified compared to the one used in generating Fig. 5. Specifically, all ubiquitin molecules from the MD frames were superimposed onto 3ONS in the least-square sense (via secondary-structure C$^{\alpha}$ atoms). The resulting superposition was then used to calculate B factors according to Eq. (3). From this calculation we have also obtained the amplitudes of rotational fluctuations experienced by ubiquitin molecules: on average, 4.1 and 4.5° for uMD and erMD $k_0 = 0.1$ trajectories, respectively. If mean MD coordinates are used as a superposition template, the corresponding numbers become 3.5 and 4.4°.

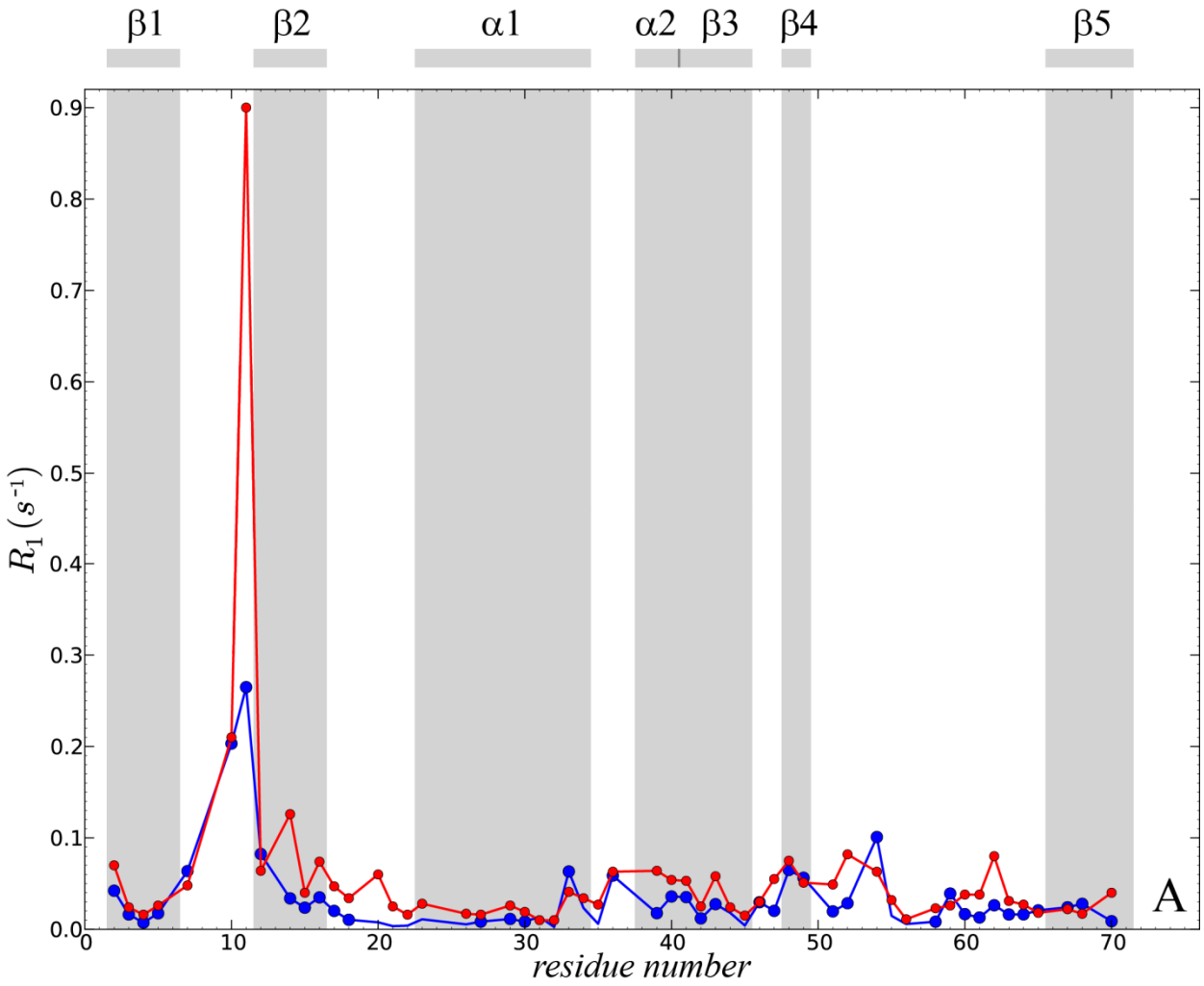

Fig. S4 (**A**). Comparison of the experimental and predicted ^{15}N R_1 relaxation rates in crystalline ubiquitin at static magnetic field strength 19.96 T (proton frequency 850 MHz). Experimental data (red symbols) are as reported by Schanda *et al*.[26] The simulated data (blue symbols) are from the uMD simulation (4U, 400 ns). This plot has been generated with the same aspect ratio as Figs. 6 and 7.

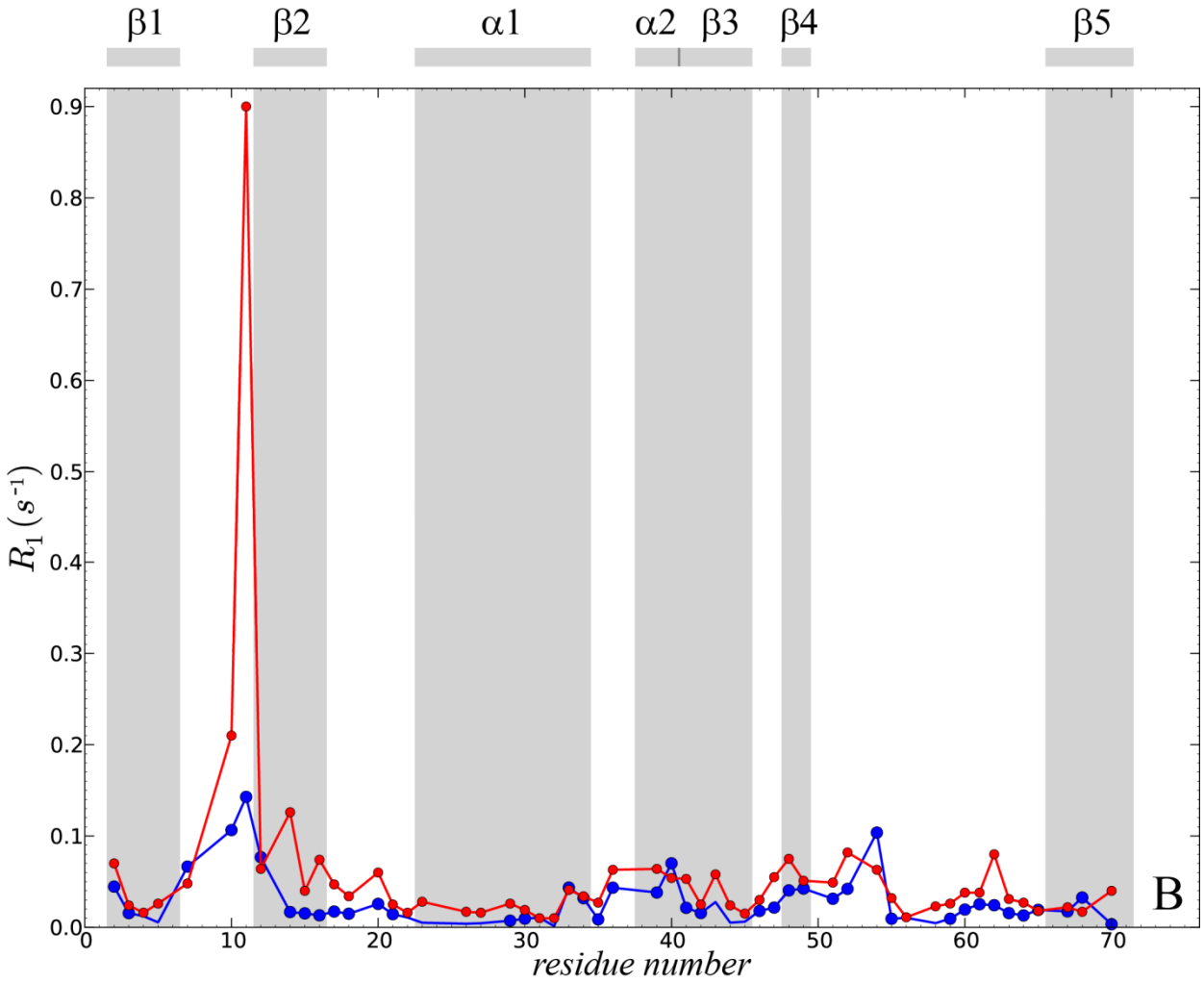

Fig. S4 (**B**). Comparison of the experimental and predicted ^{15}N R_1 relaxation rates in crystalline ubiquitin at static magnetic field strength 19.96 T (proton frequency 850 MHz). Experimental data (red symbols) are as reported by Schanda *et al.*[26] The simulated data (blue symbols) are from the erMD simulation ($k_0 = 0.1$, 4U, 400 ns). This plot has been generated with the same aspect ratio as Figs. 6 and 7.

References

(1) Huang, K. Y.; Amodeo, G. A.; Tong, L. A.; McDermott, A. *Protein Sci.* **2011**, *20*, 630.

(2) Feldman, H. J.; Hogue, C. W. V. *Proteins: Struct. Funct. Genet.* **2000**, *39*, 112.

(3) Juers, D. H.; Matthews, B. W. *J. Mol. Biol.* **2001**, *311*, 851.

(4) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins* **2008**, *73*, 765.

(5) Sundd, M.; Iverson, N.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M.; Robertson, A. D. *Biochemistry* **2002**, *41*, 7586.

(6) Harpaz, Y.; Gerstein, M.; Chothia, C. *Structure* **1994**, *2*, 641.

(7) Cerutti, D. S.; Le Trong, I.; Stenkamp, R. E.; Lybrand, T. P. *Biochemistry* **2008**, *47*, 12065.

(8) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

(9) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

(10) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *PLoS One* **2012**, *7*.

(11) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383.

(12) Showalter, S. A.; Johnson, E.; Rance, M.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 14146.

(13) Penev, E.; Ireta, J.; Shea, J. E. *J. Phys. Chem. B* **2008**, *112*, 6872.

(14) Aliev, A. E.; Courtier-Murias, D. *J. Phys. Chem. B* **2010**, *114*, 12358.

(15) Cerutti, D. S.; Freddolino, P. L.; Duke, R. E.; Case, D. A. *J. Phys. Chem. B* **2010**, *114*, 12811.

(16) Li, D. W.; Brüschweiler, R. *J. Phys. Chem. Lett.* **2010**, *1*, 246.

(17) Beauchamp, K. A.; Lin, Y. S.; Das, R.; Pande, V. S. *J. Chem. Theory Comput.* **2012**, *8*, 1409.

(18) Sidhu, A.; Surolia, A.; Robertson, A. D.; Sundd, M. *J. Mol. Biol.* **2011**, *411*, 1037.

(19) Borjesson, U.; Hunenberger, P. H. *J. Chem. Phys.* **2001**, *114*, 9706.

(20) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *Proteins* **2004**, *56*, 738.

(21) Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157.

(22) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmuller, H. *J. Chem. Theory Comput.* **2011**, *7*, 1962.

(23) Afonine, P. V.; Urzhumtsev, A. *Acta Crystallogr A* **2004**, *60*, 19.

(24) Gros, P.; Van Gunsteren, W. F.; Hol, W. G. J. *Science* **1990**, *249*, 1149.

(25) Burnley, B. T.; Afonine, P. V.; Adams, P. D.; Gros, P. *eLife Sciences* **2012**, *1*, e00311.

(26) Schanda, P.; Meier, B. H.; Ernst, M. *J. Am. Chem. Soc.* **2010**, *132*, 15957.

(27) Han, B.; Liu, Y. F.; Ginzinger, S. W.; Wishart, D. S. *J. Biomol. NMR* **2011**, *50*, 43.

(28) Haller, J. D.; Schanda, P. *J. Biomol. NMR* **2013**, *57*, 263.

(29) Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961.

(30) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870.