Supporting Information

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

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

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, 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 $\widehat{\mathbf{R}}^{(p)T}$ 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

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

(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 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

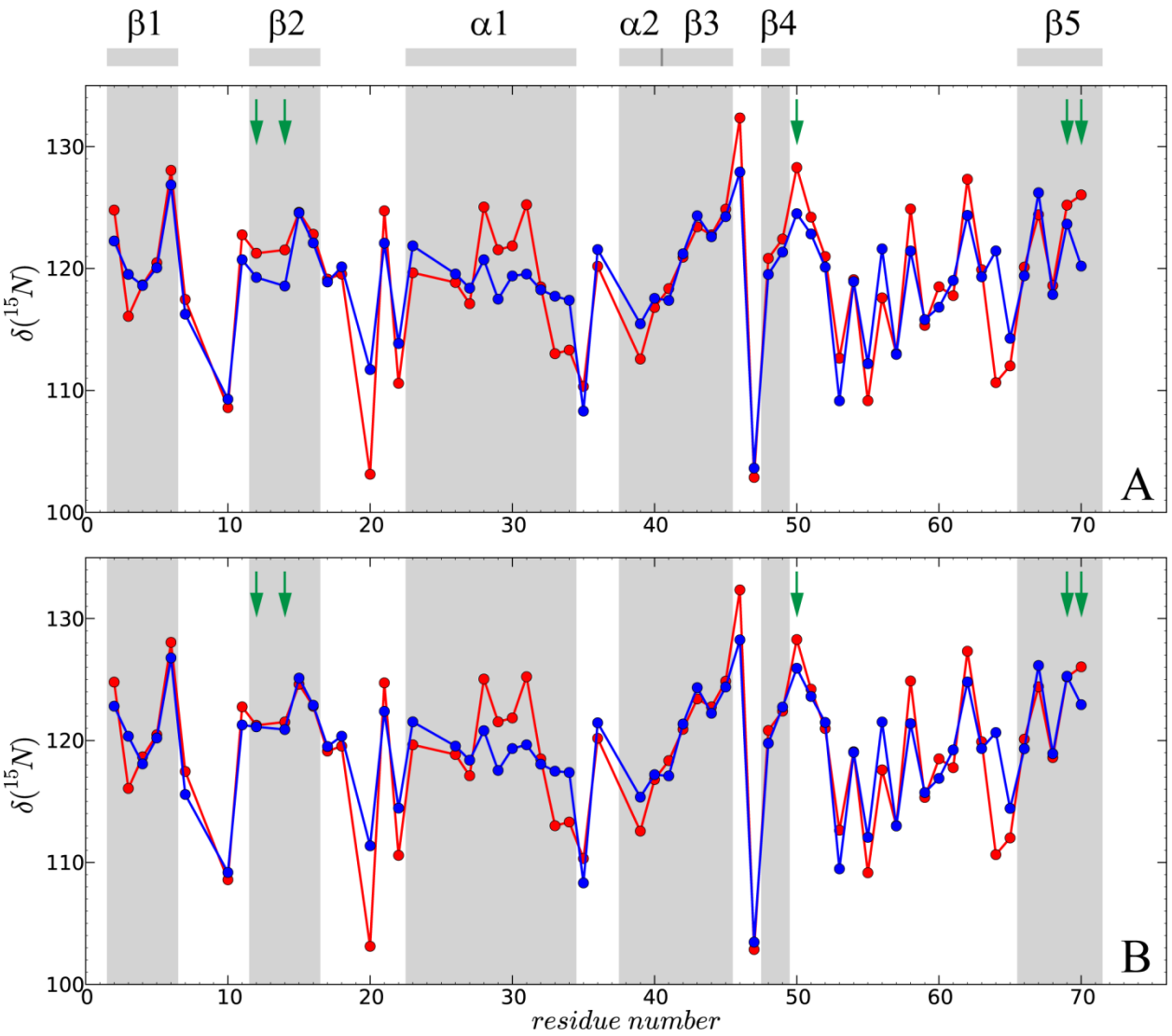

Fig. S1. Comparison of the experimental and predicted ^{15}N chemical shifts in crystalline ubiquitin. Experimental data (red symbols) are from Schanda *et al*. (26). The simulated data (blue symbols) are from application of the program SHIFTX+ (27) to (**A**) the uMD simulation, $k_0 = 0$, and (**B**) 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.

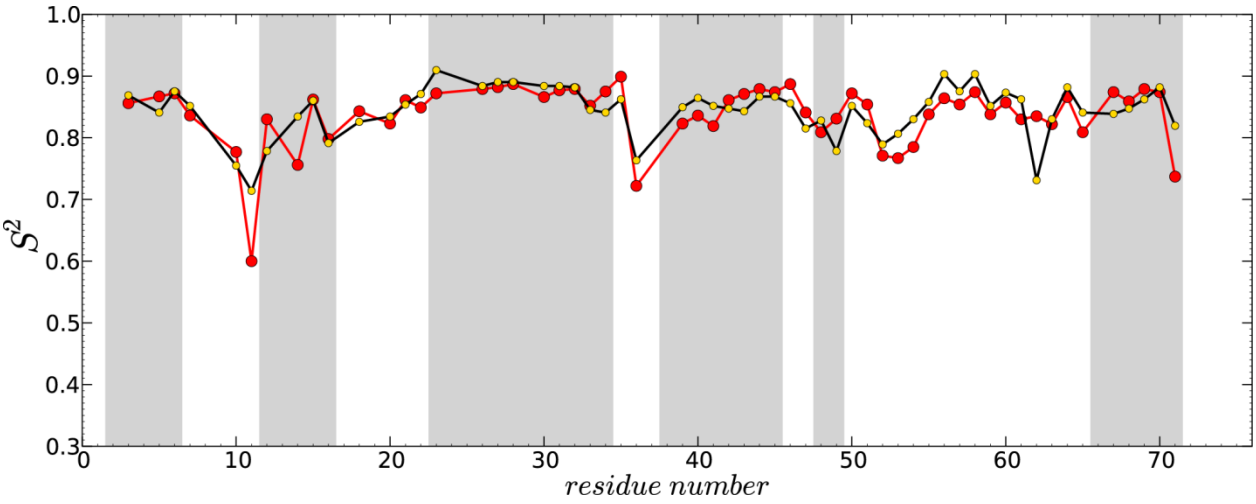

Fig. S2. Comparison of the experimental ^{15}N-^{1}H^{N} dipolar order parameters from crystalline ubiquitin with the experimental ^{15}N-relaxation-based order parameters from ubiquitin in solution. Solid-state data (red symbols) are from Schanda *et al*. (28). Solution data (black & gold) are from 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.

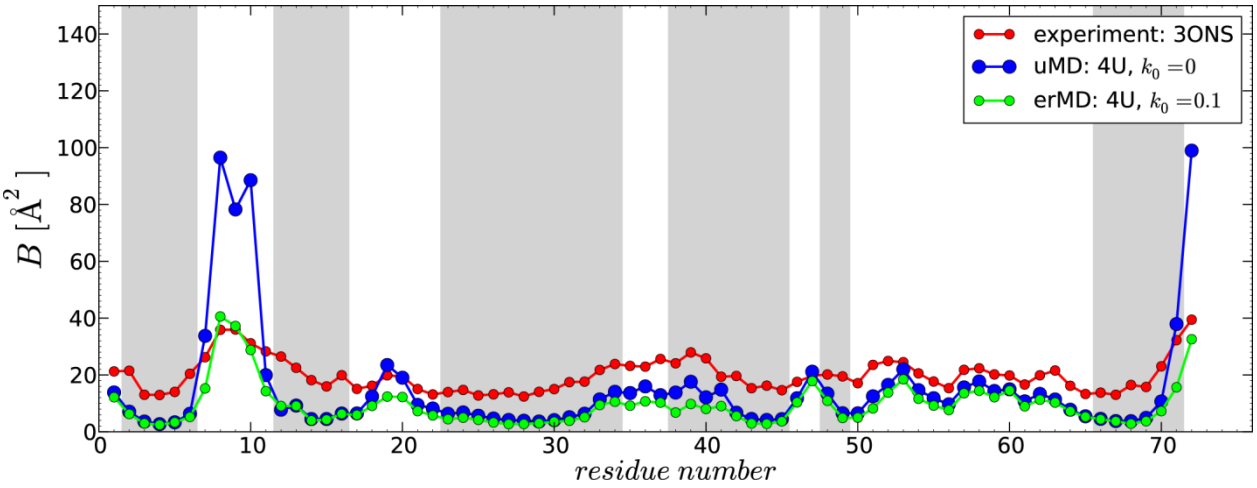

Fig. S3. Comparison of the experimental and predicted B-factors in crystalline ubiquitin. The computational protocol has been modified compared to the one used in generating Fig. 5. Specifically, all ubiquitin molecules from the MD frames were superimposed onto 3ONS in the least-square sense (via secondary-structure 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°.

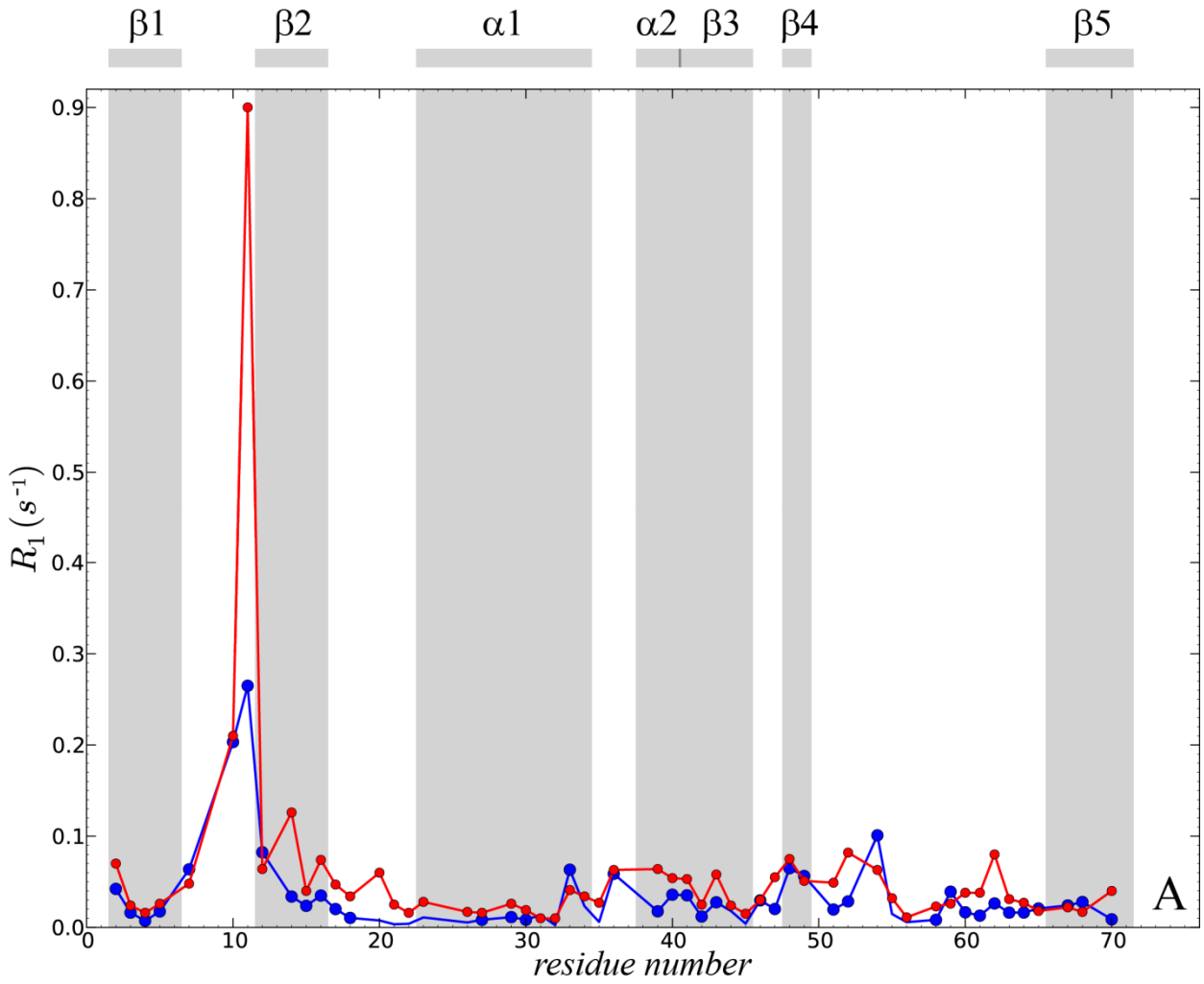

Fig. S4(A). Comparison of the experimental and predicted ^{15}N R_1 relaxation rates in crystalline ubiquitin at static magnetic field strength 19.96 T (proton frequency 850 MHz). Experimental data (red symbols) are as reported by Schanda *et al*. (26). The simulated data (blue symbols) are from the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

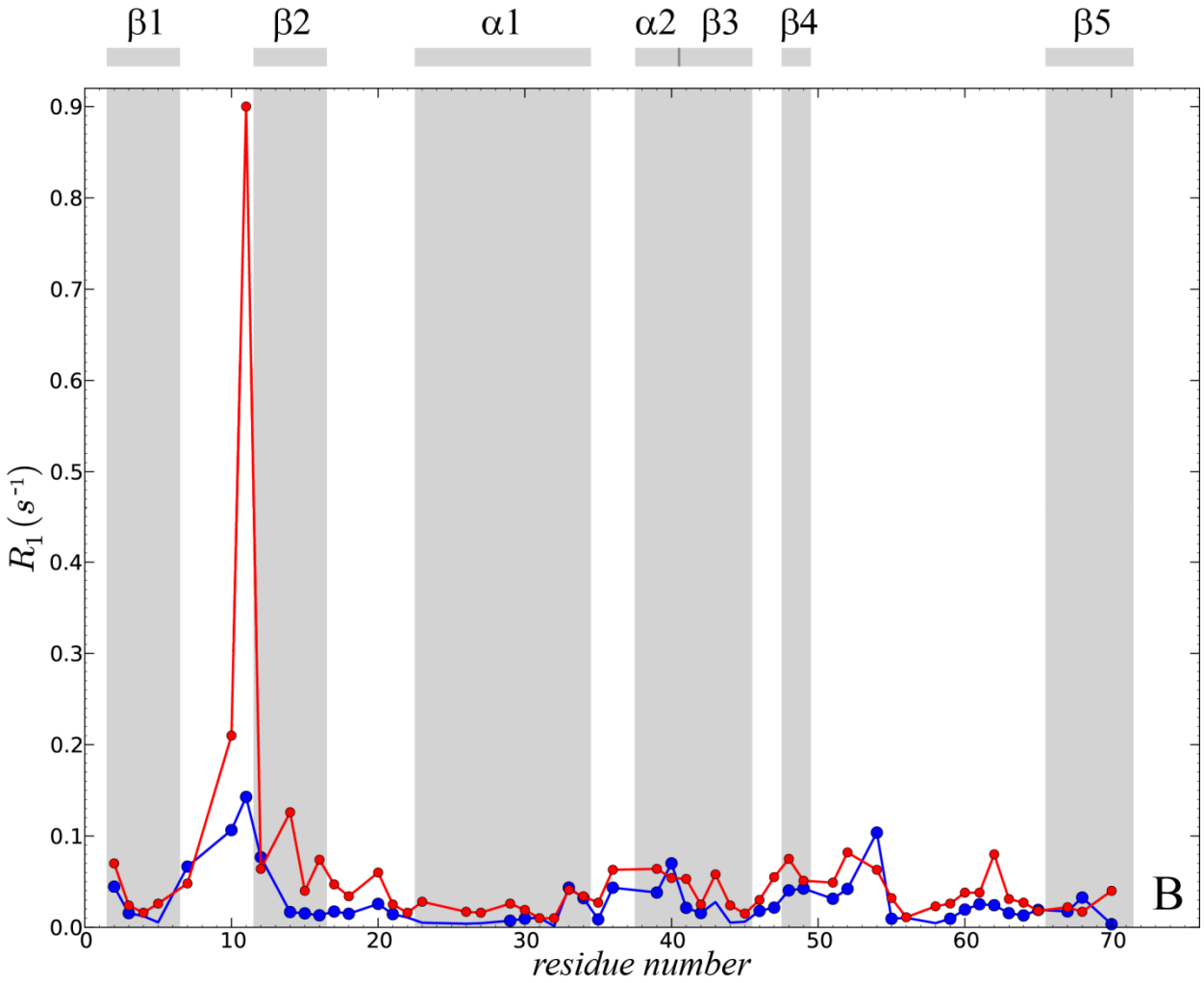

Fig. S4 (B). Comparison of the experimental and predicted ^{15}N R_1 relaxation rates in crystalline ubiquitin at static magnetic field strength 19.96 T (proton frequency 850 MHz). Experimental data (red symbols) are as reported by Schanda *et al*. (26). The simulated data (blue symbols) are from the 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 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(3):630-639.
2. Feldman HJ & Hogue CWV (2000) A fast method to sample real protein conformational space. *Proteins: Struct Funct Genet* 39(2):112-131.
3. Juers DH & Matthews BW (2001) Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. *J Mol Biol* 311(4):851-862.
4. Bas DC, Rogers DM, & Jensen JH (2008) Very fast prediction and rationalization of pK_a values for protein-ligand complexes. *Proteins* 73(3):765-783.
5. 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(24):7586-7596.
6. Harpaz Y, Gerstein M, & Chothia C (1994) Volume changes on protein folding. *Structure* 2(7):641-649.
7. 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(46):12065-12077.
8. Case DA*, et al.* (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26(16):1668-1688.
9. Berendsen HJC, Grigera JR, & Straatsma TP (1987) The missing term in effective pair potentials. *J Phys Chem* 91(24):6269-6271.
10. Lindorff-Larsen K*, et al.* (2012) Systematic validation of protein force fields against experimental data. *PLoS One* 7(2).
11. Onufriev A, Bashford D, & Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55(2):383-394.
12. Showalter SA, Johnson E, Rance M, & Brüschweiler R (2007) Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. *J Am Chem Soc* 129(46):14146-14147.
13. Penev E, Ireta J, & Shea JE (2008) Energetics of infinite homopolypeptide chains: a new look at commonly used force fields. *J Phys Chem B* 112(22):6872-6877.
14. Aliev AE & Courtier-Murias D (2010) Experimental verification of force fields for Molecular Dynamics simulations using Gly-Pro-Gly-Gly. *J Phys Chem B* 114(38):12358-12375.
15. 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(40):12811-12824.
16. Li DW & Bruschweiler R (2010) Certification of molecular dynamics trajectories with NMR chemical shifts. *J Phys Chem Lett* 1(1):246-248.
17. Beauchamp KA, Lin YS, Das R, & Pande VS (2012) Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J Chem Theory Comput* 8(4):1409-1414.
18. Sidhu A, Surolia A, Robertson AD, & Sundd M (2011) A hydrogen bond regulates slow motions in ubiquitin by modulating a ß-turn flip. *J Mol Biol* 411(5):1037-1048.
19. Borjesson U & Hunenberger PH (2001) Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. *J Chem Phys* 114(22):9706-9719.

20. Lee MS, Salsbury FR, & Brooks CL (2004) Constant-pH molecular dynamics using continuous titration coordinates. *Proteins* 56(4):738-752.
21. Mongan J & Case DA (2005) Biomolecular simulations at constant pH. *Curr Opin Struct Biol* 15(2):157-163.
22. Donnini S, Tegeler F, Groenhof G, & Grubmuller H (2011) Constant pH molecular dynamics in explicit solvent with λ-dynamics. *J Chem Theory Comput* 7(6):1962-1978.
23. Afonine PV & Urzhumtsev A (2004) On a fast calculation of structure factors at a subatomic resolution. *Acta Crystallogr A* 60:19-32.
24. Gros P, Van Gunsteren WF, & Hol WGJ (1990) Inclusion of thermal motion in crystallographic structure by restrained Molecular Dynamics. *Science* 249(4973):1149-1152.
25. Burnley BT, Afonine PV, Adams PD, & Gros P (2012) Modelling dynamics in protein crystal structures by ensemble refinement. *eLife Sciences* 1:e00311.
26. 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(45):15957-15967.
27. Han B, Liu YF, Ginzinger SW, & Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* 50(1):43-57.
28. 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*:ePub ahead of print.
29. Showalter SA & Brüschweiler R (2007) Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB force field. *J Chem Theory Comput* 3(3):961-975.
30. Lienin SF, Bremi T, Brutscher B, Bruschweiler R, & Ernst RR (1998) Anisotropic intramolecular backbone dynamics of ubiquitin characterized by NMR relaxation and MD computer simulation. *J Am Chem Soc* 120(38):9870-9879.