Supporting Information

69451 Weinheim, Germany

SUPPORTING INFORMATION

NMR Sample preparation

^{15}N-labeled ubiquitin was expressed and purified following standard procedures.[1] The sample was exchanged to a buffer containing 8 M urea, 10 mM Glycine-HCl pH 2.5 and 7% (v/v) D_2O using an Amicon 3K centrifugal filter device (Millipore) and concentrated to ~300 µM in a final volume of ~450 µL. To obtain the paramagnetic state, Gd-DTPA-BMA (gadodiamide, commercially known as Omniscan^{TM}) from a 0.5 M stock solution was added to the sample at 1 mM, 2 mM and 4 mM final concentrations.

NMR spectroscopy and data analysis

NMR experiments were recorded at 298 K on a Bruker Avance III 800 MHz spectrometer equipped with a cryogenic probe. NMR spectra were processed using NMRPipe[2] and analyzed using Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Backbone chemical shift assignments were transferred from previously published work obtained from a sample under identical buffer conditions to the one used in this study.[3] The assignments were further verified using 3D HNCACB and CBCA(CO)NH experiments. Transverse amide proton relaxation rates ($^1H^N$-R_2) were obtained using a HSQC-based pulse sequence by Kay et al.[4] A 2.5 ms REBURP pulse[5] centered at 8.15 ppm was used for selective refocusing of amide protons. Relaxation delays were set to 6, 30, 60, 90, 120 and 180 ms for the diamagnetic (reference) and in the range of 6-90 ms for the paramagnetic states. Relaxation rates were obtained by fitting the data to a mono-exponential decay using the Sparky relaxation module. Relaxation rate errors were estimated by Monte Carlo estimation of the spectral noise. Transverse proton sPRE rates ($^1H^N$-Γ_2) were calculated as the $^1H^N$-R_2 difference between the diamagnetic and paramagnetic states ($^1H^N$-Γ_2 = $^1H^N$-R_2^{para} - $^1H^N$-R_2^{dia}) as described previously.[6] Errors in sPRE were obtained from errors of diamagnetic and paramagnetic $^1H^N$-R_2 rates by standard error propagation. The average error in sPRE is on the order of 5%. One-bond H-N, Cα-Hα and Cα-C' RDCs measured in stretched polyacrylamide gels were taken from previously published work.[7]

A new module for calculation of sPRE

Despite their importance and practicality, sPRE data have not been regularly used as restraints in structure calculations, due to computational challenges associated with back-calculating sPRE from a structure. A grid-based approach was previously proposed by Otting[8] and LeMaster[9] for sPRE back-calculations. In this approach, the coordinates of the protein are placed in a 3D grid where each grid point represents a paramagnetic probe and the sPRE for an atom is calculated by a $1/r^6$-weighted integration of the atom-grid distances within a given radius. However, as mentioned earlier,[10] two problems with this approach prevent its use for structure calculations. First, due to 3D integration and presence of explicit probes this approach is computationally very expensive. The problem becomes even more severe in the case of disordered proteins, where the sPRE for each conformer in a very large ensemble needs to be calculated at every time-step during the calculation. Second, computing the gradient of this function with respect to atomic coordinates is not straightforward. Here we overcome these challenges by using an alternative approach based on the molecular surface. In brief, the surface of the protein is tessellated with triangle patches as described previously[11] using the algorithm originally developed for the program SURF.[12] This creates the probe-excluded surface of the protein, which is essentially the surface of the protein as sensed by a paramagnetic probe of a given radius. The sPRE is then evaluated by numerically integrating over the triangular patches, in an approach which also supplies an approximate gradient.[13] We have implemented this function as a new energy potential (PSolPot) for sPRE in the structure calculation program Xplor-NIH[14] and tested it on several experimental sPRE datasets. As an example, the experimental[15] and theoretical sPREs calculated using this approach from the native ubiquitin structure (PDB: 1D3Z[16]) show an excellent agreement (correlation coefficient r = 0.95 and sPRE Q-factor = 0.17), demonstrating the validity of our approach (Figure S2). Note that compared to the general solvent accessibility term (NBTarget) proposed earlier,[17] the back-calculated sPRE values using PsolPot show significantly better correlation to experimental data *(Kooshapur et al., unpublished)*. The PSolPot module enables direct refinement against sPRE data in ensemble simulated annealing calculations in a computationally-efficient manner. On a computer with a 2.3 GHz Intel Core i7 processor one can compute a single 5-member ensemble in approximately 50 min and an arbitrary number of ensembles can be computed simultaneously using multiple processors.

Ensemble structure calculations

Structures of urea-denatured ubiquitin were calculated using Xplor-NIH version 2.47.[14] In addition to standard potentials for covalent geometry (bond lengths, bond angles and improper torsion angles) and nonbonded van der Waals repulsions, the following energy terms: PsolPot (sPRE), SardcPot (steric alignment RDC)[18] and SolnXRayPot (SAXS)[19] were used in the calculations. The SAXS data[20] were kindly provided by Frank Gabel (IBS, Grenoble). As with other observables, the ensemble average sPRE for nucleus i is calculated as:

$$\Gamma_i = \sum_j w_j \Gamma_{ij}$$

where w_j is the population of ensemble member *j* and Γ_{ij} is the sPRE calculated for nucleus *i* of ensemble member *j*. An overview of ensemble refinement within Xplor-NIH is given in.[21] Starting from an extended structure, the energy minimization protocol comprised a high temperature molecular dynamics step, simulated annealing, gradient minimization in torsion angle space and a final gradient minimization in Cartesian space. We calculated 10,000 conformers (2,000 structures with ensemble size of 5) and the 50% lowest-energy structures were selected for analysis. Contact maps were generated using Xplor-NIH's contactMap helper program with residues considered in contact if their respective Cα atoms are within 8 Å of each other.

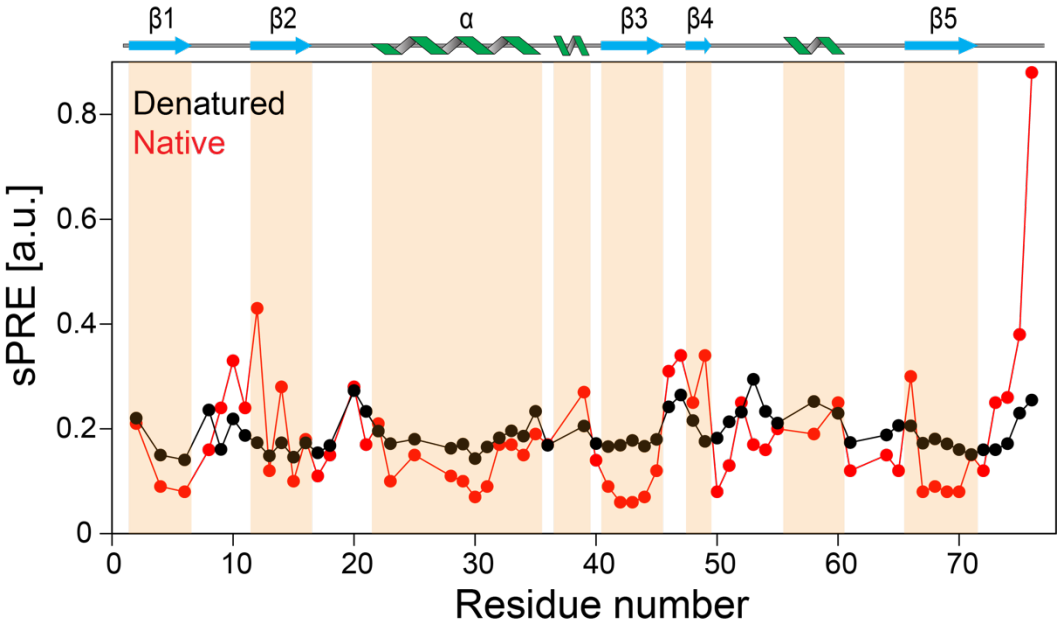

Figure S1. Comparison of sPRE data of urea-denatured (black) and native (red) ubiquitin. The native ubiquitin sPRE data are taken from Madl et al.[15] To account for different experimental conditions in which the two datasets were measured, the denatured ubiquitin sPRE data were scaled with a constant factor such that they best fit the native ubiquitin sPRE data. Secondary structure elements of the native state are shown above the plot and the corresponding regions are shaded. The patterns in the two datasets, particularly for residues 1-35, are very similar: the correlation coefficient between the two datasets for β1 (residues 2-6), β2 (residues 12-16) and the α-helix (residues 22-35) is 0.99, 0.76 and 0.84, respectively.

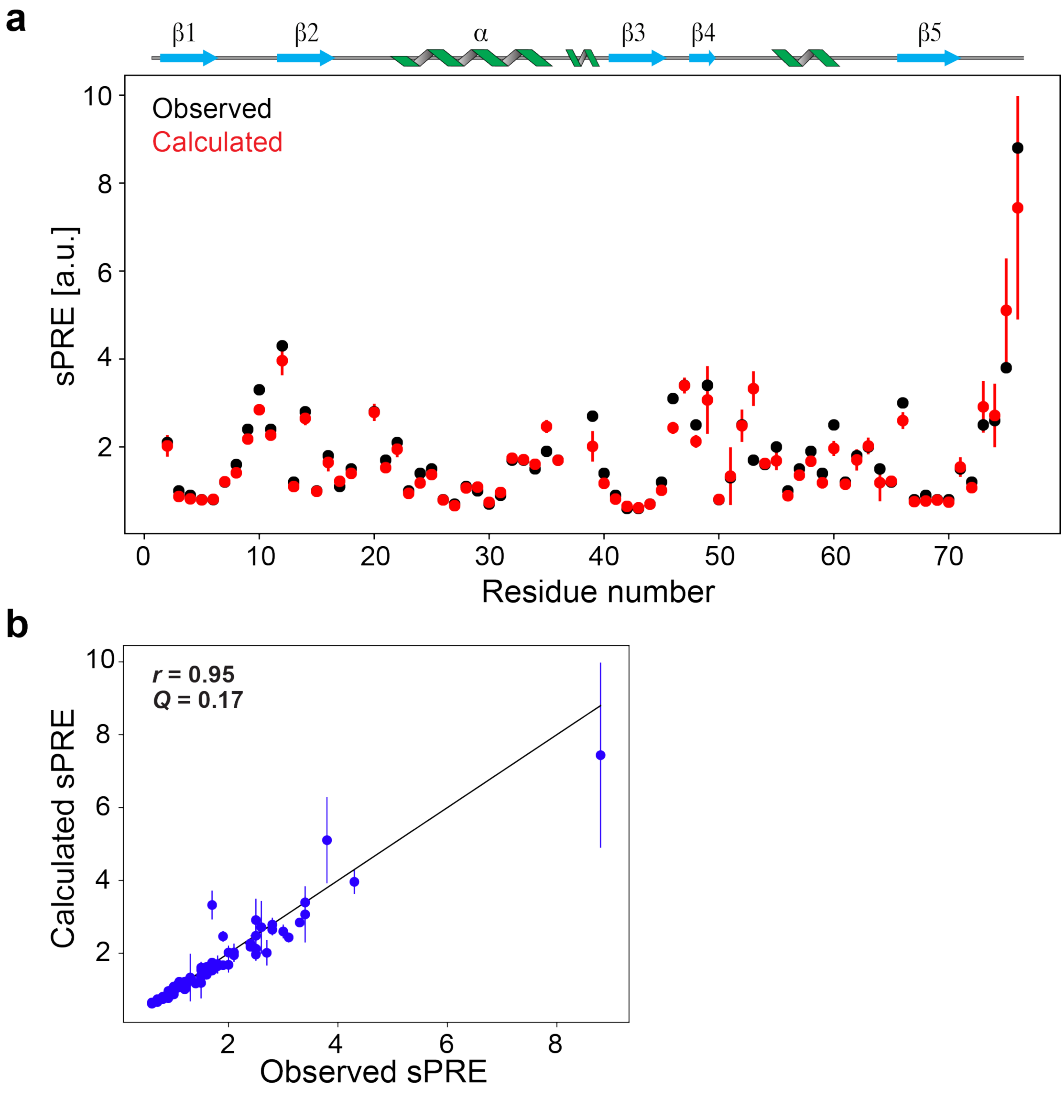

Figure S2. (a) sPREs observed (black) and back-calculated from the NMR structure of ubiquitin (PDB:1D3Z, red) are plotted against the residue number. The experimental sPRE data are taken from Madl et al.[15] Secondary structure elements are shown above the plot. (b) Correlation between observed and back-calculated sPRE of native ubiquitin. The correlation coefficient (*r*) and sPRE *Q*-factor are indicated. The error bars in panels a and b reflect the variation in sPREs calculated from the ensemble of 10 lowest-energy structures in the pdb file. Back-calculation from the single lowest energy structure yields an equally good fit to the experiment with $r = 0.96$.

References

[1] G. A. Lazar, J. R. Desjarlais, T. M. Handel, *Protein Sci* **1997**, *6*, 1167-1178.
[2] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J Biomol NMR* **1995**, *6*, 277-293.
[3] S. Meier, M. Strohmeier, M. Blackledge, S. Grzesiek, *J Am Chem Soc* **2007**, *129*, 754-755.
[4] L. W. Donaldson, N. R. Skrynnikov, W. Y. Choy, D. R. Muhandiram, B. Sarkar, J. D. Forman-Kay, L. E. Kay, *J Am Chem Soc* **2001**, *123*, 9843-9847.
[5] H. Geen, R. Freeman, *Journal of Magnetic Resonance (1969)* **1991**, *93*, 93-141.
[6] J. Iwahara, C. Tang, G. Marius Clore, *Journal of magnetic resonance* **2007**, *184*, 185-195.
[7] S. Meier, S. Grzesiek, M. Blackledge, *J Am Chem Soc* **2007**, *129*, 9799-9807.
[8] G. Pintacuda, G. Otting, *J Am Chem Soc* **2002**, *124*, 372-373.
[9] G. Hernandez, C. L. Teng, R. G. Bryant, D. M. LeMaster, *J Am Chem Soc* **2002**, *124*, 4463-4472.
[10] G. M. Clore, J. Iwahara, *Chemical Reviews* **2009**, *109*, 4108-4139.
[11] Y. E. Ryabov, C. Geraghty, A. Varshney, D. Fushman, *J Am Chem Soc* **2006**, *128*, 15432-15444.
[12] A. Varshney, F. P. Brooks, W. V. Wright, *IEEE Computer Graphics and Applications* **1994**, *14*, 19-25.
[13] Z. Gong, C. D. Schwieters, C. Tang, *Methods* **2018**.
[14] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, *Journal of magnetic resonance* **2003**, *160*, 65-73.
[15] T. Madl, W. Bermel, K. Zangger, *Angewandte Chemie* **2009**, *48*, 8259-8262.
[16] G. Cornilescu, J. L. Marquardt, M. Ottiger, A. Bax, *Journal of the American Chemical Society* **1998**, *120*, 6836-6837.
[17] Y. Wang, C. D. Schwieters, N. Tjandra, *Journal of magnetic resonance* **2012**, *221*, 76-84.
[18] J. R. Huang, S. Grzesiek, *J Am Chem Soc* **2010**, *132*, 694-705.
[19] C. D. Schwieters, G. M. Clore, *Prog Nucl Magn Reson Spectrosc* **2014**, *80*, 1-11.
[20] F. Gabel, M. R. Jensen, G. Zaccai, M. Blackledge, *J Am Chem Soc* **2009**, *131*, 8769-8771.
[21] C. D. Schwieters, G. A. Bermejo, G. M. Clore, *Protein Sci* **2018**, *27*, 26-40.