Simultaneous determination of protein structure and dynamics

Kresten Lindorff-Larsen[1]*†, Robert B. Best[1]*†, Mark A. DePristo[2]*, Christopher M. Dobson[1] & Michele Vendruscolo[1]

[1]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK*
[2]*Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK*

* Present addresses: Department of Protein Chemistry, Institute of Molecular Biology, University of Copenhagen, Øster Farimagsgade 2A, 4, DK-1353 Copenhagen K, Denmark (K.L.-L.); Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-0520, USA (R.B.B.); Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA (M.A.D.)
† These authors contributed equally to this work

We present a protocol for the experimental determination of ensembles of protein conformations that represent simultaneously the native structure and its associated dynamics. The procedure combines the strengths of nuclear magnetic resonance spectroscopy—for obtaining experimental information at the atomic level about the structural and dynamical features of proteins—with the ability of molecular dynamics simulations to explore a wide range of protein conformations. We illustrate the method for human ubiquitin in solution and find that there is considerable conformational heterogeneity throughout the protein structure. The interior atoms of the protein are tightly packed in each individual conformation that contributes to the ensemble but their overall behaviour can be described as having a significant degree of liquid-like character. The protocol is completely general and should lead to significant advances in our ability to understand and utilize the structures of native proteins.

Traditional procedures for the experimental determination of protein structures have relied on X-ray crystallographic and nuclear magnetic resonance (NMR) methods that have been designed to determine the average structure of a protein with high accuracy. It has long been recognized, however, that the dynamical properties associated with backbone and side-chain mobilities are also crucial determinants of many aspects of protein behaviour, including stability, folding and function[1–6]. Information about such dynamic processes can be obtained from a variety of different experimental techniques, most notably those detecting NMR relaxation phenomena[7]. In addition, molecular dynamics simulations can provide insight into a wide variety of phenomena associated with molecular motion[3]. The dynamical properties of proteins are, however, generally treated in isolation from the structure determination process, giving rise to considerable uncertainty as to the distribution of conformations that is sampled by a protein under a given set of conditions.

We describe a method of determining protein structures in which information about the average structure is combined with explicit information obtained from NMR relaxation experiments[7,8] about the structural variability in solution. We call the method dynamic-ensemble refinement (DER) and illustrate it on human ubiquitin, for which a comprehensive range of high-quality NMR data are available[9–16]. We used experimentally determined order parameters (S^2) for the native state of ubiquitin[9,13] as well as distance information from nuclear Overhauser effect (NOE) data[11] as restraints in molecular dynamics simulations[8,17]. The order parameters contain atomic-detailed information about the amplitude of molecular motion on a picosecond to nanosecond timescale, and thus about the variability of the native state ensemble on this timescale[7]. By requiring that a set of ubiquitin conformations is simultaneously consistent with both the NOE data and the S^2 restraints we obtain an experimental ensemble that represents both the structure and the dynamical variability in the native state of ubiquitin.

Ensemble averaging in structure determination

To apply the DER approach to ubiquitin we used molecular dynamics simulations in combination with a simulated annealing protocol to determine 128 conformations of ubiquitin that, as an ensemble, are compatible with both the experimental NOE and S^2 data (see Supplementary Methods for details of all methods). Because the NOE-derived distances are averages over the large ensemble of molecules that is present in an experimental sample, we enforce them on an average calculated over a computer-generated ensemble of structures rather than on a single conformer[18–20]. To validate the conformations we have back-calculated backbone residual dipolar couplings (RDCs)[11] and side-chain scalar couplings[15] and compared them with experimental values of these parameters. As these data were not used in the structure determination procedure, the high correlations obtained between the predicted and experimental values (Fig. 1) show that, despite the heterogeneity enforced through the S^2 restraints, the structures are in remarkable agreement with independent data describing backbone and side-chain conformations in ubiquitin, a result that strongly validates the DER approach.

The ensemble averaged q-factor, defined as $q = \sqrt{\Sigma(\mathrm{RDC_{calc}} - \mathrm{RDC_{exp}})^2}/\sqrt{\Sigma(\mathrm{RDC_{calc}})^2}$, calculated from the RDCs is 26%, which is comparable to that obtained by calculating these values from the crystal structure of ubiquitin[21] (24%). The q-factor from the published NMR ensemble[11] is lower (14%) but in this case the RDCs were included as restraints in the structure determination protocol. The correlation between experimental and calculated scalar couplings from our ensemble has $r^2 = 0.96$ compared with 0.84 and 0.89 from the crystal structure and NMR ensemble, respectively, showing that the inclusion of conformational averaging around the side-chain dihedrals markedly improves the agreement with the experimental data[22]. In support of this conclusion we find that our ensemble as a whole provides a notably better prediction of the scalar coupling data than any of the 128 individual conformations within the ensemble (mean r^2 0.81, maximum 0.91). A similar result is obtained for the q-factor for the RDCs, where the ensemble average is 26% compared with a mean of 41% and a minimum value of 33% for the individual conformations. Results of this type have been found generally in the comparison of the DER structures with independent experimental data. This is in contrast to structures determined using single-molecule refinement as indicated by the calculation of a set of independently determined side-chain RDC values[12]. The ensemble

averaged q-factor for the DER ensemble (48%) from this set of RDCs is comparable to that derived from the NMR structure[11] (47%). However, whereas q-factors for individual conformations within the DER ensemble cover a broad range from 52–89% (mean 72%) and only show good agreement to the experimental data as an ensemble, all the individual conformations within the NMR ensemble have low q-factors (47–52%, mean 49%). For this set of RDCs the crystal structure[21] has $q = 77\%$.

To quantify the heterogeneity introduced by the use of S^2 values as restraints (Fig. 2a) we calculated the root mean square deviations (r.m.s.d.) of the positions of all heavy atoms in the ubiquitin molecule (Fig. 2b, c). In addition to the carboxy-terminal region (residues 72–76), which is known to be highly flexible[9], we find that a large number of residues show marked variability within the calculated ensemble. These residues include not only those that are surface exposed, but also a considerable proportion of the atoms in the core of the protein. We used the generalized Lindemann criterion[23,24] to compare the amplitude of the atomic fluctuations (r.m.s.d.) with the average distances between atoms, quantified as the ratio of these two values (the Lindemann Δ value). A critical value of $\Delta \approx 0.15$ indicates a transition between solid-like (low Δ) and liquid-like (high Δ) behaviour for a wide range of compounds[23,24].

If we exclude residues 72–76 from the analysis we find that the Lindemann Δ value is 0.14 for the heavy atoms of the backbone and 0.29 for those of the side chains, showing that on average the molecule can be characterized as having solid-like rigidity in the backbone with liquid-like side chains attached. Surprisingly, this conclusion is still valid if one examines only the 90 heavy atoms that form the internal core of the protein, defined as those within 6 Å of the centre of mass[24] ($\Delta = 0.25$ for side-chain atoms and 0.12 for backbone atoms); by comparison, the remaining 473 heavy atoms have $\Delta = 0.30$ for the side-chain atoms and 0.15 for the backbone atoms. Finally, the 94 most exterior backbone heavy atoms (>12 Å from the centre of mass) have mobilities that on average are liquid-like ($\Delta = 0.17$). Thus, our results show that ubiquitin in solution is not only 'surface molten'[24], but also 'side-chain molten' even in the core of the protein.

Comparison with conventional structure determination methods

We then compared the heterogeneity of the structures obtained by DER with that of several ensembles determined by other methods (Fig. 2b, c). We first examined the published NMR ensemble[11] and an ensemble (the 'rapper ensemble') consisting of 50 conformations that individually are all compatible with the published X-ray structure factors[21], and which we determined using a recently described procedure[25]. These two ensembles both show a lower heterogeneity than the one we determined using S^2 values as restraints—a result that was anticipated given that they were determined with the objective of defining an accurate estimate of the average atomic positions, rather than of the associated dynamics. Finally, we performed a 6-ns molecular dynamics simulation in explicit solvent to sample the native-state dynamics of ubiquitin. For both main-chain and side-chain atoms the molecular dynamics simulation gives rise to r.m.s.d. values whose magnitudes are comparable to those in our ensemble; although, as described below, the local structures in the two ensembles can differ substantially. Nonetheless, this result demonstrates that molecular dynamics simulations provide an overall description of the magnitude of molecular fluctuations that is consistent with the experimental NMR data.

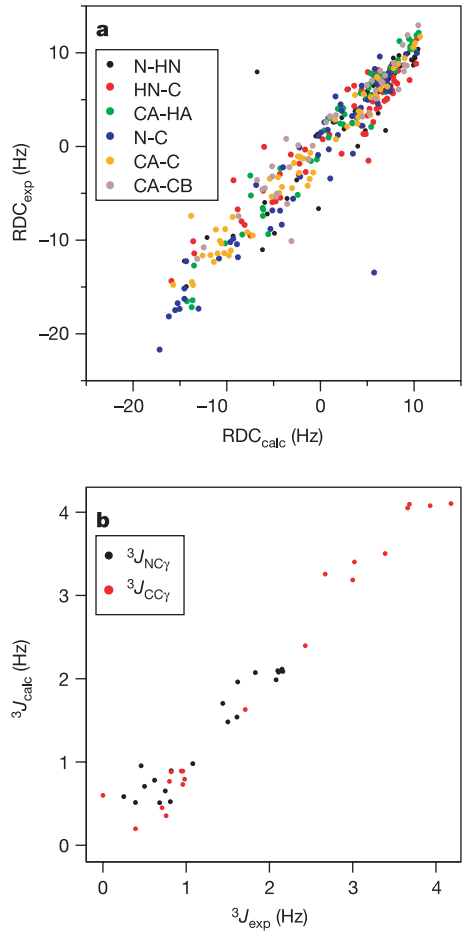

Figure 1 Cross-validation of ubiquitin structures by comparison with independently determined NMR data. Calculation of residual dipolar couplings (RDCs)[11] (**a**) and side-chain scalar couplings[15] (**b**). The NMR data were back-calculated from an ensemble of conformations that was determined using DER as described in the text and in the Supplementary Information. For clarity, the magnitudes of the residual dipolar couplings were normalized to those for an amide NH in the same orientation by scaling according to bond lengths and gyromagnetic ratios[10]. The data point labels in **a** describe the atoms between which the RDCs were measured. In **b**, $^3J_{NC\gamma}$ and $^3J_{CC\gamma}$ are scalar couplings between the side-chain γ carbon and the backbone amide nitrogen and carbonyl carbon, respectively.

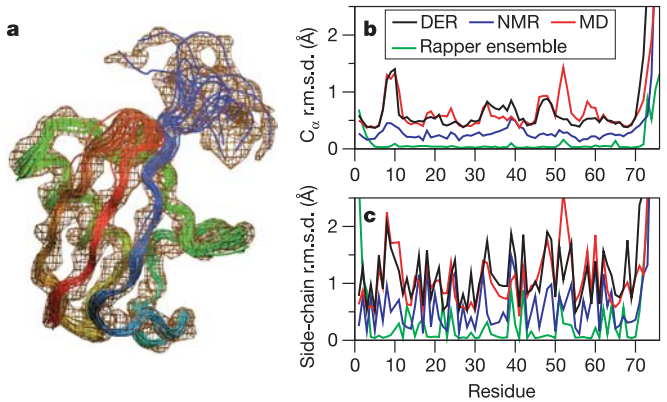

Figure 2 Variability in structural ensembles of ubiquitin. **a**, Backbone trace of 15 representative conformations obtained from a clustering procedure. The structures are coloured from the N terminus (red) to the C terminus (blue) and are traced within an atomic density map[46] representing the 20% amplitude isosurface of the density of atoms in the polypeptide main chain. The r.m.s.d. values of backbone (Cα) atoms (**b**) and side-chain atoms (**c**) in ubiquitin ensembles were determined by dynamic-ensemble refinement, NMR[11], X-ray diffraction[21,25] and molecular dynamics (MD) simulations.

The liquid-like mobility of the side-chain atoms in our ensemble stems primarily from their ability to occupy multiple rotameric states, a general characteristic already observed both in experiments and molecular dynamics simulations[15,26,27]. To illustrate this fact we calculated the distributions of selected side-chain dihedral angles for a set of aliphatic residues in our ensemble and compared them with those in the X-ray and NMR structures (Fig. 3). In the ensemble determined using DER, many side chains occupy multiple rotamers and show considerable variability within each rotamer. For example, 38 out of the 68 non-Ala/Gly residues populate more than one χ_1 rotamer[28] to an extent of at least 10%. For comparison, only 15 of the residues in the NMR ensemble and six residues in the rapper ensemble show this level of heterogeneity, and all of these are located at the solvent-exposed surface of the protein.

A more quantitative analysis of the heterogeneity in the different ensembles was performed by back-calculating backbone and side-chain order parameters (Fig. 4). The ensemble that we have determined in the present study using both NOE and S^2 restraints is in good agreement with experiment ($r^2 = 0.96$), due to the inclusion of these values as restraints. For comparison, the lower variability in the X-ray rapper and NMR ensembles manifests itself as a large number of calculated order parameters that are close to unity and in general are larger than the experimentally observed values, giving rise to lower coefficients of correlation ($r^2 = 0.37$ and 0.62 for the X-ray and NMR ensembles, respectively).

We also determined ensembles of conformations using only NOE restraints enforced onto either a single conformation or a set of molecules. From the ensemble determined by enforcing the NOEs on a single conformation we obtain a correlation between predicted and experimental S^2 values ($r^2 = 0.61$) that is similar to that for the published NMR ensemble given above. However, if ensemble averaging is introduced for NOEs, and thus more variability is allowed, the correlation coefficient increases to $r^2 = 0.76$, again highlighting the importance of an ensemble description of the protein structure. Finally, although the ensemble obtained by unrestrained molecular dynamics simulations shows an overall level of variability that is similar to that derived using S^2 restraints (Fig. 2), the lower correlation with the experimental order parameters ($r^2 = 0.62$) indicates that a detailed description of the dynamics of many individual residues is substantially enhanced by the incorporation of S^2 restraints, thereby overcoming the present limitations in conventional molecular dynamics simulations due to limited sampling and force-field imperfections[3].

The consequences of protein dynamics

The liquid-like mobility of side-chain atoms could in principle affect the tight packing of amino acid residues known to be a hallmark of the native structures of proteins[29,30]. To examine whether the introduction of heterogeneity through the DER method presented here can be realized without disrupting the high packing density, we calculated[30] the volume taken up by 14 residues that dominate the hydrophobic core of ubiquitin in our ensemble (2,268 ± 41 Å^3; mean ± s.d.) as well as in the NMR (2,303 ± 14 Å^3) and crystal (2,304 Å^3) structures. These results

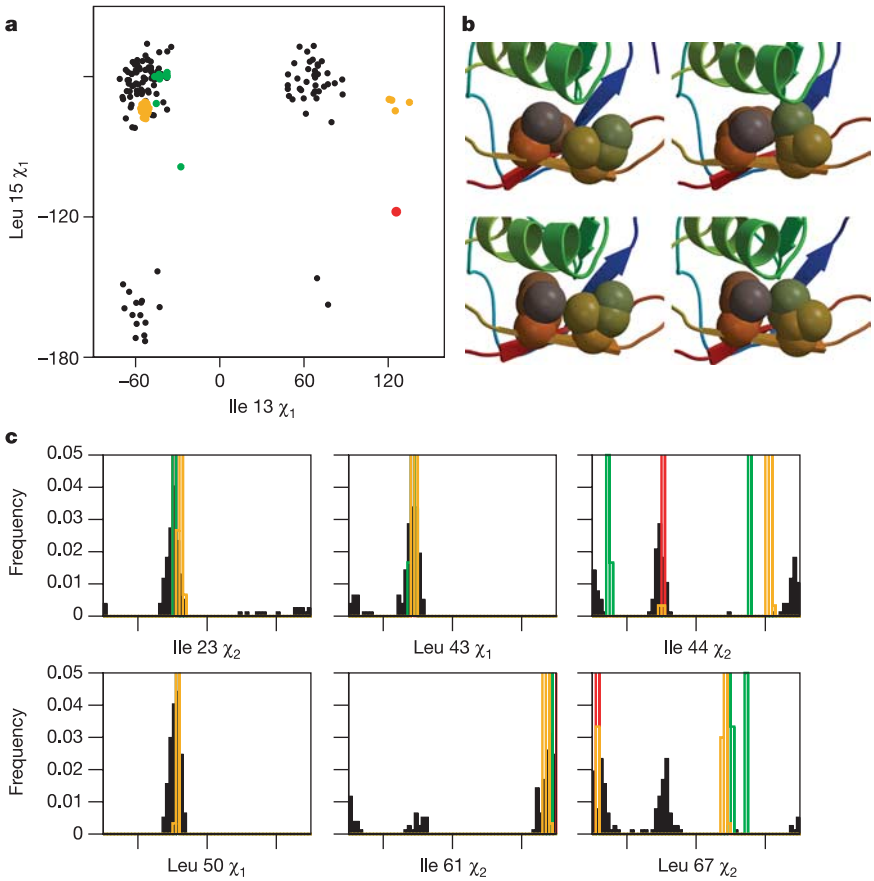

Figure 3 Examples of the liquid-like mobility of side chains in the DER ensemble. **a**, Joint distribution of the χ_1 dihedral angles[28] in Ile 13 and Leu 15 in our 128 conformer ensemble (black), the crystal structure (red), the X-ray rapper ensemble (orange) and the published NMR ensemble[11] (green). **b**, Four structures chosen from our DER ensemble to represent the four groupings of dihedrals evident in **a**; the four structures are arranged to match the four regions. Heavy atoms in the side chains of Ile 13 and Leu 15 are shown as van der Waals spheres (Ile 13 is located to the right of Leu 15). **c**, Distribution of side-chain χ_1 and χ_2 dihedral angles[28] of selected hydrophobic residues. Colouring scheme as in **a**. In some of the plots the histograms of dihedral angles for the X-ray, NMR and rapper structures overlap.

show that the core is at least as tightly packed in the ensemble that we have determined as in the crystal and NMR structures, despite the fact that the structures determined here by DER show considerable variability.

Protein structures are widely used as an important tool in the design and analysis of experimental studies. A typical example is the rationalization and prediction of the effects of mutations on protein stability. To illustrate the utility of an accurate description of the native state heterogeneity in an application of this type, we used FOLD-X[31] to predict the stability changes accompanying 28 mutations distributed throughout the ubiquitin sequence for which experimental data are available (S. E. Jackson, personal communication). The correlation (r^2) and r.m.s.d. between the experimental values and values predicted from the ensemble that we have determined using DER are 0.69 and 0.7 kcal mol^{-1}, respectively. The corresponding numbers obtained using the published X-ray structure and NMR ensemble are 0.58 (0.9 kcal mol^{-1}) and 0.47 (1.0 kcal mol^{-1}), respectively, showing the importance of the DER method for predicting such thermodynamic properties. As in the case of RDCs and scalar couplings, the predictions from our ensemble as a whole are better than those from the individual conformations of which it is composed (mean values of r^2 and r.m.s.d. are 0.61 and 0.8 kcal mol^{-1}, respectively).

The fact that many order parameters calculated from the NMR and X-ray ensembles are higher than those measured experimentally could at least in part be caused by the use of energy minimization and simulated annealing routines in the structure determination protocols. To explore this effect we energy minimized each of the 128 conformations determined in this work and estimated the order parameters from the resulting ensemble. Amide and methyl group order parameters increased on average by 0.056 and 0.034, respectively, a result that is in accord with the observation that backbone librations play an important role in the dynamics probed by amide order parameters[32]. These librations are significantly reduced by the minimization procedure, causing the order parameters to increase. Nonetheless, the variation over the minimized ensemble is still much greater than over the NMR and X-ray structures as a result of anharmonic effects such as multiple rotamer populations.

We suggest that although energy minimization and simulated annealing to very low temperatures in a structure determination protocol can be used to decrease the enthalpy of a structure, these procedures could in addition decrease the entropy; that is, reduce the magnitudes of thermal vibrations. Thus, energy minimization procedures move each individual conformation towards the 'average' conformation but reduce the variability of the ensemble as a whole. As an illustration of this effect, energy minimization causes a decrease in the backbone RDC q-factor of the individual conformations (the average value of the individual q-factors is 40.1% before and 35.8% after minimization) without changing the ensemble average (the ensemble averaged q-factor is 25.7% both before and after minimization).

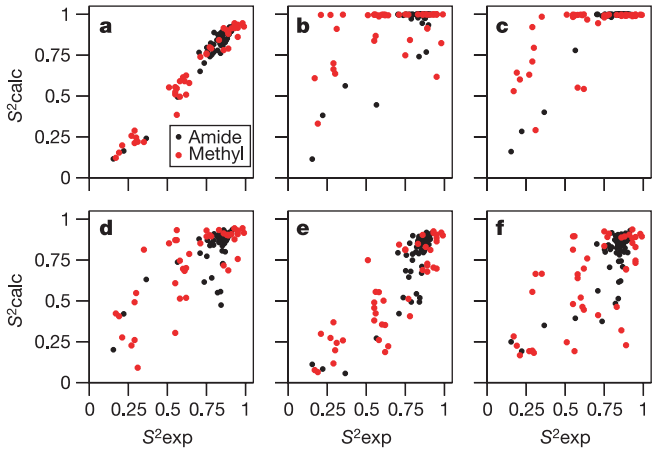

Figure 4 Quantifying backbone and side-chain variability in ubiquitin ensembles by comparison of experimental and back-calculated order parameters. In addition, the q-factors for a set of backbone RDCs[11] are given for the individual ensembles. **a**, 128 conformer DER ensemble from this work using both NOEs and S^2 values as restraints ($q = 26\%$). **b**, X-ray rapper ensemble ($q = 24\%$). **c**, The published NMR ensemble ($q = 14\%$). **d**, 64 conformer ensemble from this work using NOEs enforced as restraints on a single conformer ($q = 23\%$). **e**, 128 conformer ensemble from this work using NOEs enforced as restraints on an ensemble of molecules ($q = 24\%$). **f**, Ensemble obtained from molecular dynamics simulations ($q = 52\%$). Structural ensembles were obtained as described in the text and in more detail in the Supplementary Information.

Significance for understanding protein behaviour

Well-established structure determination protocols based on either NMR or X-ray diffraction data can be used to determine very precise molecular representations of the dominant conformations of proteins. The dramatic achievements of structural biology show that these methods provide a finely detailed view of protein structures that, in conjunction with molecular dynamics simulations, can be used to rationalize and predict biological function. Complementary to the structural information available from these techniques, NMR relaxation experiments probe directly the dynamics of a protein. We have shown here that a combination of experimental parameters that reflect the structural and dynamical properties of proteins with molecular dynamics simulations extends conventional structure determination methods to include an accurate atomistic description of the conformational variability of the native states of proteins.

Application of the DER approach to ubiquitin reveals clearly that the native state must be considered as a heterogeneous ensemble of conformations that interconvert on the picosecond to nanosecond timescale, as well as populating more expanded conformations arising from rare but large fluctuations on much longer timescales, such as those revealed by hydrogen exchange experiments[33]. We find that many side chains, even in the core of the protein, occupy multiple rotameric states and can be considered to have liquid-like characteristics. On the basis of the available NMR relaxation data[27,34] this observation is expected to be general and may have important biological and functional consequences. We believe therefore that the DER approach will enable a more detailed understanding to be gained about such fundamental properties as the contribution of entropic terms to native state stability[3,35], the effects of dynamics on enzyme function[6] and the ability of proteins to retain their structure even when subjected to marked mutational changes such as the complete redesign of the hydrophobic core[36,37].

By using a range of different types of experimental data as restraints in computer simulations it is becoming possible to obtain detailed structural models of proteins in denatured[38], intermediate[39] and transition[40,41] states. Although it is clear that it is crucial to take ensemble averaging into account in structural studies of denatured proteins[38], the results described here demonstrate the importance of an ensemble interpretation of structural data for native states of proteins. By using the same principles, the DER approach may readily be extended to incorporate structural information from other types of experimental measurements including NMR chemical shifts and RDCs, and perhaps even data from X-ray diffraction. Finally, dynamic-ensemble refinement should substantially enhance our ability to design drug molecules on a rational basis[5], to provide a detailed description of protein–protein interfaces, and to perform such procedures as homology modelling, structure prediction and other methods that require a high level of definition of the structural details of proteins. □

Methods

Structure determination was performed by solvating ubiquitin in a shell of 595 TIP3P water molecules[43] and carrying out eight cycles of simulated annealing using an energy function of the form: $E_{tot} = E_{CHARMM} + E_{NOE} + E_{S^2}$. In this expression, E_{CHARMM} is the CHARMM22 (ref. 44) force-field energy and E_{NOE} and E_{S^2} are the energies due to the NOE and S^2 restraints, respectively. NOE-derived distance restraints[11] and experimental S^2 values[9,13] were obtained from the literature. As the experimental NOEs and S^2 values are averages over a large ensemble of molecules, we enforced these restraints on an average calculated over a computer-generated ensemble, as described in detail in the Supplementary Methods section. The restraints were applied using biased molecular dynamics in combination with ensemble simulations as described[8,45]. The X-ray rapper ensemble was determined as described previously[25]. A 6-ns unrestrained molecular dynamics simulation was performed essentially as described previously[46]. Details of the structural analysis of the simulations can be found in the Supplementary Methods section.

Received 16 September; accepted 28 October 2004; doi:10.1038/nature03199.

1. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions on proteins. *Science* **254,** 1598–1603 (1991).
2. Rasmussen, B. F., Stock, A. M., Ringe, D. & Petsko, G. A. Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature* **357,** 423–424 (1992).
3. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Struct. Biol.* **9,** 646–652 (2002).
4. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science* **295,** 1520–1523 (2002).
5. Wong, C. F. & McCammon, J. A. Protein flexibility and computer-aided drug design. *Annu. Rev. Pharmacol. Toxicol.* **43,** 31–45 (2003).
6. Benkovic, S. J. & Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **301,** 1196–1202 (2003).
7. Kay, L. E. Protein dynamics from NMR. *Nature Struct. Biol.* **5,** 513–517 (1998).
8. Best, R. B. & Vendruscolo, M. Determination of ensembles of structures consistent with NMR order parameters. *J. Am. Chem. Soc.* **126,** 8090–8091 (2004).
9. Tjandra, N., Feller, S. E., Pastor, R. W. & Bax, A. Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J. Am. Chem. Soc.* **117,** 12562–12566 (1995).
10. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278,** 1111–1114 (1997).
11. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120,** 6836–6837 (1998).
12. Ottiger, M. & Bax, A. How tetrahedral are methyl groups in proteins? A liquid crystal NMR study. *J. Am. Chem. Soc.* **121,** 4690–4695 (1999).
13. Lee, A. L., Flynn, P. F. & Wand, A. J. Comparison of ^{2}H and ^{13}C NMR relaxation techniques for the study of protein methyl group dynamics in solution. *J. Am. Chem. Soc.* **121,** 2891–2902 (1999).
14. Peti, W., Meiler, J., Brüschweiler, R. & Griesinger, C. Model-free analysis of protein backbone motion from residual dipolar couplings. *J. Am. Chem. Soc.* **124,** 5822–5833 (2002).
15. Chou, J. J., Case, D. A. & Bax, A. Insights into the mobility of methyl-bearing side chains in proteins from $^3J_{CC}$ and $^3J_{CN}$ couplings. *J. Am. Chem. Soc.* **125,** 8959–8966 (2003).
16. Clore, G. M. & Schwieters, C. D. 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 (2004).
17. Kitao, A. & Wagner, G. A space-time structure determination of human CD2 reveals the CD58-binding mode. *Proc. Natl Acad. Sci. USA* **97,** 2064–2068 (2000).
18. Scheek, R. M., Torda, A. E., Kemmink, J. & van Gunsteren, W. F. in *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy* (eds Hoch, J. C., Redfield, C. & Poulsen, F. M.) 209–217 (Plenum, New York, 1991).
19. Bonvin, A. M. J. J., Rullmann, J. A. C., Lamerichs, R. M. J. N., Boelens, R. & Kaptein, R. 'Ensemble' iterative relaxation matrix approach: A new NMR refinement protocol applied to the solution structure of crambin. *Proteins* **15,** 385–400 (1993).
20. Choy, W. Y. & Forman-Kay, J. D. Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol.* **308,** 1011–1032 (2001).
21. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194,** 531–544 (1987).
22. Mierke, D. F., Scheek, R. M. & Kessler, H. Coupling constants as restraints in ensemble driven dynamics. *Biopolymers* **34,** 559–563 (1994).
23. Stillinger, F. H. & Stillinger, D. K. Computational study of transition dynamics in 55-atom clusters. *J. Chem. Phys.* **93,** 6013–6024 (1990).
24. Zhou, Y., Vitkup, D. & Karplus, M. Native proteins are surface-molten solids: Application of the Lindemann criterion for the solid versus liquid state. *J. Mol. Biol.* **285,** 1371–1375 (1999).
25. DePristo, M. A., de Bakker, P. I. & Blundell, T. L. Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* **12,** 831–838 (2004).
26. Hoch, J. C., Dobson, C. M. & Karplus, M. Vicinal coupling constants and protein dynamics. *Biochemistry* **24,** 3831–3841 (1984).
27. Best, R. B., Clarke, J. & Karplus, M. The origin of protein sidechain order parameter distributions. *J. Am. Chem. Soc.* **126,** 7734–7735 (2004).
28. Markley, J. L. *et al.* Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J. Mol. Biol.* **280,** 933–952 (1998).
29. Richards, F. M. The interpretation of protein structures: Total volume, group volume distribution and packing density. *J. Mol. Biol.* **82,** 1–14 (1974).
30. Pontius, J., Richelle, J. & Wodak, S. J. Deviation from standard atomic volumes as a quality measure for protein crystal structure. *J. Mol. Biol.* **264,** 121–136 (1996).
31. Guerois, R., Nielsen, J. E. & Serrano, L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J. Mol. Biol.* **320,** 369–387 (2002).
32. Buck, M. & Karplus, M. Internal and overall peptide group motion in proteins: Molecular dynamics simulations for lysozyme compared with results from X-ray and NMR spectroscopy. *J. Am. Chem. Soc.* **121,** 9645–9658 (1999).
33. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. *J. Am. Chem. Soc.* **125,** 15686–15687 (2003).
34. Lee, A. L. & Wand, A. J. Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature* **411,** 501–504 (2001).
35. Yang, D. & Kay, L. E. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: Application to protein folding. *J. Mol. Biol.* **263,** 369–382 (1996).
36. Johnson, E. C., Lazar, G. A., Desjarlais, J. R. & Handel, T. M. Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Struct. Fold. Des.* **7,** 967–976 (1999).
37. Benítez-Cardoza, C. G. *et al.* Exploring sequence/folding space: Folding studies on multiple hydrophobic core mutants of ubiquitin. *Biochemistry* **43,** 5195–5203 (2004).
38. Lindorff-Larsen, K. *et al.* Determination of a broad structural ensemble representing the denatured state of the bovine acyl-coenzyme A binding protein. *J. Am. Chem. Soc.* **126,** 3291–3299 (2004).
39. Korzhnev, D. M. *et al.* Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430,** 586–590 (2004).
40. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **409,** 641–645 (2001).
41. Lindorff-Larsen, K., Paci, E., Vendruscolo, M. & Dobson, C. M. Transition states for protein folding have native topologies despite high structural variability. *Nature Struct. Mol. Biol.* **11,** 443–449 (2004).
42. Schwieters, C. D. & Clore, G. M. Reweighted atomic densities to represent ensembles of NMR structures. *J. Biomol. NMR* **23,** 221–225 (2002).
43. Jorgensen, W. J., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).
44. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **4,** 187–217 (1983).
45. Paci, E. & Karplus, M. Forced unfolding of fibronectin type 3 modules: An analysis by biased molecular dynamics simulations. *J. Mol. Biol.* **288,** 441–459 (1999).
46. Fox, T. & Kollman, P. A. The application of different solvation and electrostatic models in molecular dynamics simulations of ubiquitin: How well is the X-ray structure 'maintained'? *Proteins* **25,** 315–334 (1995).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We are very grateful to P. I. de Bakker and T. L. Blundell for assistance in determining the X-ray rapper ensemble of ubiquitin. We thank S. E. Jackson for sharing the experimental data on ubiquitin stability changes before publication. K.L.L. is supported by the Danish Research Agency. M.A.D. was funded by the Marshall Aid Commemoration Commission, US National Science Foundation, and Cambridge Overseas Trust. M.V. is a Royal Society University Research Fellow. The research of M.V. and C.M.D. is supported in part by Programme Grants from the Wellcome and Leverhulme Trusts.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.V. (mv245@cam.ac.uk) or C.M.D. (cmd44@cam.ac.uk).