

Relative Stability of Protein Structures Determined by X-ray Crystallography or NMR Spectroscopy: A Molecular Dynamics Simulation Study

Hao Fan and Alan E. Mark*

Groningen Biomolecular Sciences and Biotechnology Institute, Department of Biophysical Chemistry, University of Groningen, Groningen, The Netherlands

ABSTRACT The relative stability of protein structures determined by either X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy has been investigated by using molecular dynamics simulation techniques. Published structures of 34 proteins containing between 50 and 100 residues have been evaluated. The proteins selected represent a mixture of secondary structure types including all α , all β , and α/β . The proteins selected do not contain cysteine–cysteine bridges. In addition, any crystallographic waters, metal ions, cofactors, or bound ligands were removed before the systems were simulated. The stability of the structures was evaluated by simulating, under identical conditions, each of the proteins for at least 5 ns in explicit solvent. It is found that not only do NMR-derived structures have, on average, higher internal strain than structures determined by X-ray crystallography but that a significant proportion of the structures are unstable and rapidly diverge in simulations. *Proteins* 2003;53:111–120.

© 2003 Wiley-Liss, Inc.

Key words: X-ray crystallography; NMR spectroscopy; molecular dynamic simulation; protein structure; structure refinement

INTRODUCTION

Currently, the primary means to determine the tertiary structures of proteins at atomic resolution is either using solution nuclear magnetic resonance (NMR) spectroscopy techniques or by X-ray diffraction (X-ray) from crystals.¹ NMR spectroscopy has the advantage that structures can be obtained in solution without the constraints imposed by crystal packing but is limited to molecules below a given size that do not aggregate at high concentration.² X-ray crystallography is not in principle limited to molecules of a given size but does require crystallization. Both approaches are very powerful. However, neither can be used to solve the structure of a protein directly, except in rare cases. Instead, molecular modeling and simulation techniques are used to generate a structure or a collection of structures consistent with the experimental data. In X-ray crystallography, this involves fitting a model to electron density data.³ In NMR-derived structures, it involves fitting a collection of models to a series of distance and angle constraints.⁴ Thus, not only is the environment in

which proteins are studied when using NMR or X-ray crystallographic techniques different but, so too are the nature of the structural data obtained and the distribution of structural information within the molecule. For example, in NMR-derived structures, redundant distance information is normally only available in tightly packed regions of the structure. Resonance assignments can also be ambiguous. Furthermore, in both cases the experimental data itself corresponds to a time and ensemble average rather than reflecting the properties of a single molecule.⁴

As a consequence of the above, structures of proteins determined by NMR or X-ray crystallography will never be identical. They are ultimately models that have been fitted to different experimental data. Fortunately, the structures of proteins determined by both NMR and X-ray crystallography have in general been shown to have very similar backbone folds.^{5–9} Nevertheless, they will differ in detail such as how groups of atoms and side-chains are packed. The question is, how significant are these differences? In particular, are structures obtained from NMR and X-ray crystallography equally suitable for use as a starting point for further studies such as modeling protein dynamics, protein–protein interactions, or protein–ligand interactions?

In an attempt to address this question, molecular dynamics simulation techniques have been used to investigate the relative stability of protein structures solved by using either NMR or X-ray crystallography or both. A total of 34 proteins ranging in size from 50 to 100 amino acid residues were selected for the study. None of the proteins contain cysteine–cysteine bridges. All bound metal ions, ligands, and crystallographic waters were removed. Each protein structure was simulated in explicit solvent under identical conditions. Key structural and dynamic properties of the systems were then analyzed. The concept behind this study was that by keeping all other parameters constant, variations in the properties analyzed would depend primarily on the quality of initial structural data.

*Correspondence to: Alan E. Mark, Groningen Biomolecular Sciences and Biotechnology Institute, Department of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. E-mail: mark@chem.rug.nl

Received 12 October 2002; Accepted 8 April 2003

TABLE I. A Summary of the Properties of the Structures Determined by X-ray Crystallography. Units kJmol^{-1} and nm

No.	PDBID	Description	Reference	Resolution	N_{res}	N_{α}	N_{β}	Charge %	Net Charge	Energy protein	Energy per res	Rg	RMSD	RMSD (Bfac1)	RMSD (Bfac2)
1	1vif	Dihydrofolate reductase	Narayana et al., 1995 ¹⁰	0.18	60	0	26	17	0	-2688	-44.8	1.08	0.29	0.19 (4)	0.29 (0)
2	1tuc	A-spectrin	Viguera et al., 1995 ¹¹	0.18	61	3	25	30	0	-2752	-45.1	1.08	0.21	0.21 (1)	0.16 (2)
3	1vcc	DNA topoisomerase I	Sharma et al., 1994 ¹²	0.16	77	11	25	25	-1	-4221	-54.8	1.22	0.23	0.23 (0)	0.23 (0)
4	1ail	Nonstructural protein NSI	Liu et al., 1997 ¹³	0.19	70	60	0	29	2	-3996	-57.1	1.27	0.27	0.27 (0)	0.19 (7)
5	1cei	Colicin E7 immunity protein	Chak et al., 1996 ¹⁴	0.18	85	49	0	34	-9	-3413	-40.2	1.23	1.18	1.18 (0)	1.13 (1)
6	1rpo	ROP mutant (dimer)	Vlassi et al., 1994 ¹⁵	0.14	61	55	0	30	-6 -12	-2381 -39.0	1.43	0.60	0.60 (0)	0.24 (38)	0.09 (2)
7	1a32	Ribosomal protein S15 (residues 23-85)	Clemons et al., 1998 ¹⁶	0.21	85	65	0	36	7	-3951	-46.5	1.84	1.34	1.33 (4)	1.35 (2)
8	1ae3	Gene V protein (dimer)	Su et al., 1997 ¹⁷	0.20	86	6	41	17	1 2	-2220 -25.8	1.42	0.40	0.40 (1)	0.39 (4)	0.30 (2)
9	1ctf	50 S ribosomal protein	Leijonmarck et al., 1987 ¹⁸	0.17	68	38	18	35	-2	-3074	-45.2	1.12	0.17	0.17 (0)	0.17 (2)
10	1pgx	Protein G B2 domain (residues 7-62)	Achari et al., 1992 ¹⁹	0.17	70	14	28	26	-4	-1858	-26.5	1.42	0.50	0.50 (0)	0.15 (11)
11	1tif	Transition initiation factor 3	Biou et al., 1995 ²⁰	0.18	76	27	22	38	5	-3439	-45.2	1.38	0.24	0.24 (0)	0.21 (3)
12	2acy	Acylphosphate	Thunnissen et al., 1997 ²¹	0.18	98	24	41	26	1	-4590	-46.8	1.25	0.21	0.21 (2)	0.21 (1)
13	2fxb	Ferredoxin	Fukuyama et al., 2002 ²²	0.09	81	16	14	31	-17	-2545	-31.4	1.13	0.23	0.23 (0)	0.23 (1)
14	1r69	Gene regulation protein 434 repressor (R1-69)	Mondragon et al., 1989 ²³	0.20	63	40	0	22	4	-3322	-52.7	1.05	0.18	0.18 (0)	0.18 (0)
15	1bm8	Transcription factor MBP1	Xu et al., 1997 ²⁴	0.17	99	32	28	28	6	-5139	-51.9	1.24	0.29	0.28 (4)	0.29 (1)
16	2ci2	Chymotrypsin inhibitor 2	McPhalen et al., 1987 ²⁵	0.20	63	13	22	33	-1	-3340	-53.0	1.11	0.10	0.10 (0)	0.10 (0)
17	1pgb	Protein G B1 domain	Gallagher et al., 1994 ²⁶	0.19	56	14	30	28	-4	-3089	-55.2	1.05	0.13	0.13 (0)	0.13 (0)
18	1shg	α -spectrin	Musacchio et al., 1992 ²⁷	0.18	57	3	28	33	1	-3259	-57.2	1.03	0.09	0.08 (2)	0.08 (1)
19	1ubi	Ubiquitin	Ramage et al., 1994 ²⁸	0.18	76	12	23	29	0	-3653	-48.1	1.16	0.11	0.11 (0)	0.10 (4)
20	1a19	Barstar	Ratnaparkhi et al., 1998 ²⁹	0.28	89	39	15	24	-6	-3634	-40.8	1.22	0.15	0.15 (17)	0.15 (0)

PDBID, PDB identifier; N_{res} , number of residues; N_{α} , the number of helical residues; N_{β} , the number of strand residues; Charge %, the percentage of charged residues; Energy_{protein}, the initial internal energy; Energy_{perres}, internal energy per residue; Rg, the initial radius of gyration; RMSD refers to backbone root mean square positional deviation of the average structure during the last 1 ns with respect to the experimental structure. Bfac1 excludes residues (number in brackets) for which the B-factor of the C_{α} in the experimental structure is high ($\geq 60.0 \text{ \AA}^2$). Bfac2 excludes residues (number in brackets) for which the apparent B-factor calculated from the simulation is $\geq 100.0 \text{ \AA}^2$ (see explanation in text). Two C-terminal residues of 2ci2 were removed for compatibility with 3ci2.

MATERIALS AND METHODS

Structure Selection

The 39 structures used in this study,⁹⁻⁴⁷ corresponding to 34 different proteins were taken from the Protein Data Bank (PDB).⁴⁸ Of these proteins, 26 were identical to those used by Baker and coworkers to test the ROSETTA method for ab initio protein structure prediction.⁴⁹ The PDB identification code, the method used for solving the structures, and a range of structural properties of the 39 structures are summarized in Tables I and II. Of these proteins, 15 were solved with X-ray diffraction only, 14 were solved with NMR spectroscopy only, and 5 with both techniques. All of the proteins in this study contained between 50 and 100 residues. None of the structures contained cysteine-cysteine bridges. All but

two of the proteins are believed to be monomeric in solution. The 39 structures are divided into two lists. The first contains the 20 structures determined by X-ray crystallography (Table I). The second contains the 19 structures determined with use of NMR techniques (Table II). Of these NMR structures, nine correspond to energy minimized average structures where only a single structure was given in the PDB. In the remaining cases, where multiple structures have been deposited in the PDB, the first structure in each set was chosen to represent the molecule.

Molecular Dynamics Simulations

All simulations were performed in explicit water using the GROMACS (Groningen Machine for Chemical Simula-

TABLE II. A Summary of Refinement Details and Structural Properties for the Structures Determined by NMR. Units kJmol^{-1} and nm

No.	PDBID	Description	Reference	NOE's	N_{res}	N_{α}	N_{β}	Charge %	Net Charge	Energy protein	Energy per res	Rg	RMSD	RMSD (Bfac1)	RMSD (Bfac2)
1	1aoy	Arginine repressor	Sunnerhagen et al., 1997 ³⁰	2288	78	32	8	27	3	-2281	-29.2	1.32	0.42	0.11 (15)	0.42 (0)
2	1stu	Mat. Effect protein staufen	Bycroft et al., 1995 ³¹	621	68	26	20	28	5	988	14.5	1.23	0.32	0.20 (14)	0.32 (1)
3	1sro	PNPase	Bycroft et al., 1997 ³²	723	76	4	27	30	1	-485	-6.4	1.24	0.29	0.24 (12)	0.27 (5)
4	1sap	Sac7d	Edmondson et al., 1995 ³³	775	66	20	27	45	6	-3267	-49.5	1.16	0.33	0.16 (17)	0.28 (1)
5	1afi	MerP	Steele et al., 1997 ³⁴	918	72	21	21	24	3	-540	-7.5	1.12	0.15	—(64)	0.15 (0)
6	1bb8	Integrase	Connolly et al., 1998 ³⁵	1172	71	11	14	41	5	-2120	-29.8	1.28	0.73	0.71 (6)	0.52 (14)
7	2bby	Rap30	Groft et al., 1998 ³⁶	1196	69	35	4	30	3	-1285	-18.6	1.19	0.21	0.20 (2)	0.20 (2)
8	2fmr	FMRI protein	Musco et al., 1997 ³⁷	907	65	18	18	28	-4	-824	-12.7	1.19	0.54	0.40 (9)	0.53 (1)
9	1alz	FADD protein	Eberstadt et al., 1998 ³⁸	1036	83	62	0	32	-3	-1169	-14.1	1.23	0.26	0.26 (0)	0.24 (1)
10	1bw6	Centromere protein B	Iwahara et al., 1998 ³⁹	626	56	32	0	36	6	-254	-4.5	1.12	0.71	0.54 (11)	0.68 (5)
11	1coo	RNA polymerase alpha subunit	Jeon et al., 1995 ⁴⁰	822	81	35	0	28	-3	-590	-7.3	1.22	0.26	0.20 (7)	0.20 (3)
12	1lea	LexA rep. DNA binding domain	Fogh et al., 1994 ⁴¹	613	72	39	0	28	2	-4060	-56.4	1.12	0.22	—	0.19 (1)
13	2af8	Actinorhodin polyketide synthase acyl carrier	Crump et al., 1997 ⁴²	699	86	43	0	30	-12	-455	-5.3	1.31	0.66	—(45)	0.59 (5)
14	2ezh	Tranposase	Clubb et al., 1997 ⁴³	969	65	45	0	34	-2	-1288	-19.8	1.20	0.30	0.27 (6)	0.29 (1)
15	3ci2	Chymotrypsin inhibitor 2	Ludvigsen et al., 1991 ⁴⁴	961	63	11	10	32	0	-2817	-44.7	1.10	0.16	0.14 (6)	0.16 (0)
16	2gbl	Protein G B1 domain	Gronenborn et al., 1991 ⁴⁵	854	56	13	22	28	-4	-1834	-32.8	1.07	0.16	0.16 (0)	0.16 (0)
17	1aey	α -spectrin	Blanco et al., 1997 ⁹	687	57	3	28	33	1	-3995	-70.1	1.04	0.14	0.14 (3)	0.14 (1)
18	1d3z	Ubiquitin	Cornilescu et al., 1998 ⁴⁶	2727	76	12	23	29	0	-1834	-24.1	1.17	0.22	—(76)	0.15 (3)
19	1bta	Barstar	Lubienski et al., 1994 ⁴⁷	1613	89	37	16	24	-6	-2050	-23.0	1.21	0.18	0.18 (3)	0.18 (0)

NOE's, number of distance restraints used in the refinement. The C-terminal residue of 1aey was removed for compatibility 1shg. Only a single minimized average structure was given for 1sap, 1alz, 1bw6, 1coo, 1lea, 2af8, 2ezh, 2gbl and 1bta in the PDB. RMSD refers to the backbone RMSD as described in Table 1. Bfac1 excludes residues (number in brackets) for which the apparent B-factor for the C_{α} atoms calculated from the variation in the experimentally derived structures is $\geq 100.0 \text{ \AA}^2$. Where RMSD values per atom were given in the PDB file these values were used to estimate the apparent B-factor. Where no RMSD values were given the RMSD was estimated from the ensemble of structures deposited in the PDB. No values are given for 1lea as only a single average structure was deposited which did not include RMSD values per atom. In addition no estimate for the overall RMSD (Bfac1) is given where $> 50\%$ of the residues were excluded. Bfac2 excludes residues (number in brackets) for which the apparent B-factor calculated from the simulation is $\geq 100.0 \text{ \AA}^2$.

tion) package^{50–52} in conjunction with the GROMOS96 43a1 force field for condensed-phase simulations.^{53,54} The Simple Point Charge (SPC) model was used to represent water.⁵⁵ The protonation state of ionizable groups in each of the proteins was chosen appropriate for pH 7.0. No counterions were added to neutralize the system. Bound metal ions, crystallographic waters, and ligands were removed from the experimental structures before simulation. The molecular dynamics simulations were performed at constant temperature and pressure in a periodic truncated octahedral box. The minimum distance between any atom of the protein and the box wall was 1.0 nm. Depending on the size and shape of the protein, this resulted in between 3,000 and 18,000 water molecules in the simulation box with most systems containing 4000–6000 water molecules. Nonbonded interactions were evaluated by using a twin-range method. Coulomb and van der Waals

interactions within a shorter-range cutoff of 0.9 nm were evaluated at every timestep. Longer-range Coulomb and van der Waals interactions between 0.9 and 1.4 nm were updated every five steps and held constant for the intervening steps. To minimize the effects of truncating the electrostatic interactions beyond the 1.4-nm long-range cutoff, a reaction field correction⁵⁶ was applied using a relative dielectric constant of 78. To remove high-frequency degrees of freedom, explicit hydrogen atoms in the force field were replaced by dummy atoms, the positions of which were constructed each step from the coordinates of the heavy atoms to which they are attached. This allows a timestep of 4 fs to be used without affecting significantly the thermodynamic properties of the system.⁵⁷ Covalent bonds in the protein were constrained by using the LINCS algorithm.⁵⁸ The SETTLE algorithm⁵⁹ was used to constrain the geometry of the water molecules. To generate

the starting configuration for each system, the following protocol was used. After energy minimization (EM) using a steepest descent algorithm, 10 ps of molecular dynamics with position restraints on the protein (PRMD) were performed at 250 K to gently relax the system. Unrestrained molecular dynamics (MD) simulations were then performed at 300 K for 5-ns to assess the stability of the structures. During the simulations, the temperature and the pressure were maintained at 300 K and 1 bar by coupling to an external heat and an isotropic pressure bath.⁶⁰ The relaxation times were 0.1 ps and 0.5 ps, respectively.

RESULTS AND DISCUSSION

Intermolecular Energy

Ten steps of energy minimization using a steepest descent algorithm were performed on each of the 39 structures to release possible strain in the initial structures before initiating the dynamics simulations. The internal energy of each of the structures after minimization is listed in Tables I and II. The internal energy of the X-ray derived structures was in the range of -5139 kJmol⁻¹ to -1858 kJmol⁻¹. The mean internal energy is -3328 kJmol⁻¹. In contrast, the internal energies of the NMR-derived structures were in the range of -4060 kJmol⁻¹ to 988 kJmol⁻¹ with a mean internal energy of -1587 kJmol⁻¹. For the X-ray-derived structures, there is, as expected, a weak linear correlation between the number of residues in the structure and the internal energy with a negative slope (slope = -33.5 ; correlation coefficient = -0.55). The larger the molecule, the more negative is the energy. In NMR-derived structures, no such relation is evident (slope = 24.8 ; correlation coefficient = 0.18).

The average energy per residue in each structure is also listed in Tables I and II. Again, it can be seen that the average energy per residue is significantly lower for the X-ray-derived structures with a smaller standard deviation (-45 kJmol⁻¹ \pm 9) than the NMR-derived structures (-23 kJmol⁻¹ \pm 21). However, we note that those NMR structures for which both an NMR structure and an X-ray structure are available, the energy per residue is comparable to those of the X-ray structures. The relatively high energies suggest that most, although not all, of the NMR structures examined are under internal strain.

Ten steps of minimization is, of course, very limited. The structures are not fully relaxed. The minimization was used primarily to remove possible strain due to van der Waals overlap arising from minor differences between the force field used to refine the structures and the GROMOS96 force field without allowing any significant structural changes. Thus, the energies listed in these two tables reflect the structures as deposited in the PDB. However, we note that comparisons made after 100 steps of minimization were qualitatively identical.

Structure Deviations in MD Simulations

The root-mean-square positional deviation (RMSD) of all backbone atoms from the respective experimental structures (NMR or X-ray) after a performing a least-

square best fit was used to monitor structural changes after minimization and during the simulations. In all cases, the RMSD after minimization was as expected small, <0.02 nm. The RMSD after a 0.01-ns equilibration was similar for all structures and in the order of 0.1 nm. Simulations of at least a nanosecond were required to reliably resolve significant structural changes. The last three columns in Tables I and II list the RMSD for the average structure over the last 1 ns of the 5-ns simulations. The third last column lists the RMSD calculated for all backbone atoms. The second last column lists the RMSD calculated after excluding highly mobile residues as determined during the refinement of the experimental structure. The last column lists the RMSD calculated after excluding residues found to be highly mobile in the simulations (normally residues at the N- and C-termini). Consider first only the RMSD calculated for all backbone atoms, which we think is the most objective approach. Of the 20 X-ray-derived structures, 2 deviate ≤ 0.1 nm, 7 ≤ 0.2 nm, 15 ≤ 0.3 nm, and 5 deviate > 0.3 nm. Of the 19 NMR-derived structures, 0 deviate ≤ 0.1 nm, 5 ≤ 0.2 nm, 12 ≤ 0.3 nm shown, and 7 deviate > 0.3 nm. However, this comparison does not accurately reflect the stability of the structures. Five X-ray-derived structures deviate by >0.3 nm within 5 ns in the simulations. In four of these cases, the deviations from the starting crystal structures are clearly due to the release of crystal-packing forces. The ROP mutant protein (1rpo), which was a member of the test set proposed by Baker and coworkers, was in fact crystallized as a dimer.¹⁵ Simulations of the dimer show a backbone RMSD of only 0.13 nm. The mutant R82C of GENE V protein (1ae3) also crystallizes and is active as a dimer. The backbone RMSD, which is 0.40 nm when simulated as monomer, decreases to 0.32 nm when simulated as a dimer. The ribosomal protein S15 (1a32) crystallizes in a highly extended form, essentially as two domains. The high RMSD in this case reflects the relative motion and the mutual collapse of the two domains in solution [Fig. 1(a)]. The RMSD calculated for the primary domain alone (residues 23–85) is 0.56 nm. In the immunoglobulin-binding protein G B2 domain (1pgx) shown in Figure 1(b), the RMSD is dominated by the motion of the highly flexible N- and C-terminal residues. For comparison, the NMR and X-ray structures of the structurally related G B1 domain, which contains only 56 residues corresponding to the central core of G B2, has an RMSD of < 0.2 nm. Considering the equivalent core region in G B2 (residues 7–62), the RMSD is only 0.11 nm. Excluding the case of ribosomal protein S15 (1a32), for which the RMSD is clearly an artifact, we find that of the 19 remaining X-ray-derived structures, 2 deviate ≤ 0.1 nm, 9 ≤ 0.2 nm, 17 ≤ 0.3 nm, and only 2 deviate > 0.3 nm. The large deviations in the case of the X-ray-derived structure of Colicin E7 immunity protein (1cei) illustrated in Figure 1(c) could not be obviously explained by the loss of specific crystal contacts. This predominantly helical structure deviates considerably from the initial structure (1.18 nm at 5 ns). The protein does carry a high net charge of -9 ,

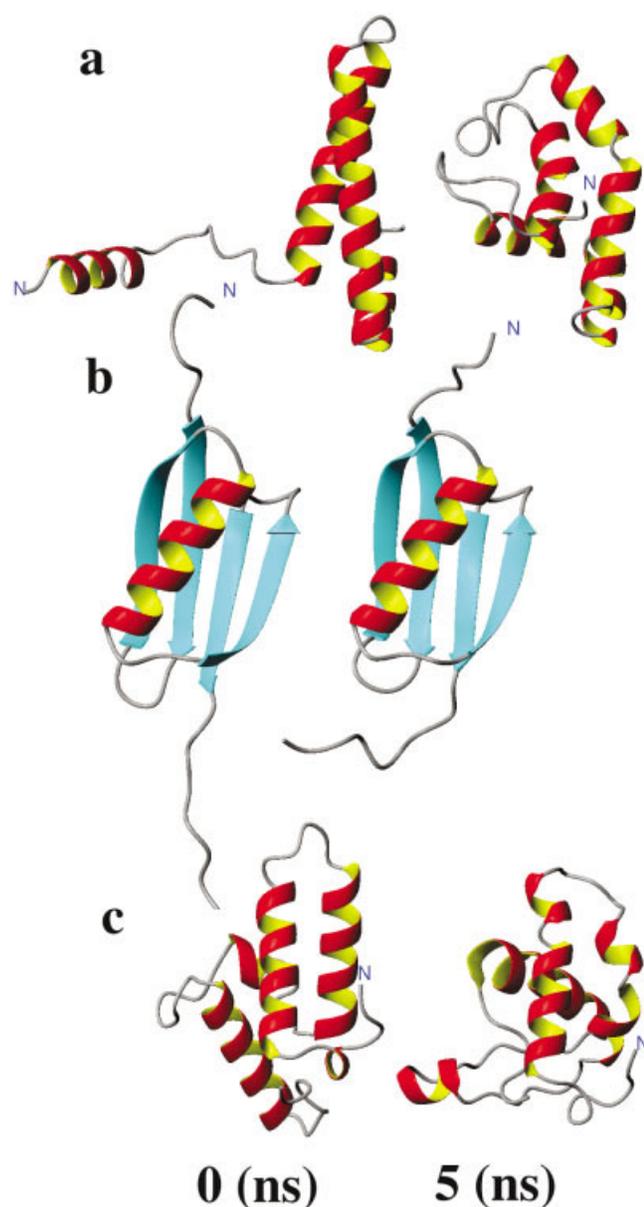


Fig. 1. Conformations of three proteins, solved by X-ray diffraction, which show large structural deviations during molecular dynamics (MD) simulations illustrating the type of deviations observed and the effects of crystal packing. On the left are shown the initial conformations (0 ns); on the right conformations after 5 ns of simulation (5 ns): ribosomal protein S15 (1a32) (a), protein G B2 domain (1pgx) (b) and colicin E7 immunity protein (1cei) (c).

and it is possible that this resulted in the structural changes observed.

We note that no counter ions were included in any of the simulations. Counter ions were not included in the simulations for three reasons. 1) They diffuse slowly on the MD timescale and the position at which ions are placed can significantly affect how specific configurations evolve in time. 2) The effects of surface charges are effectively screened by the dielectric response of water. 3) Because the box volumes are small and vary considerably between the simulations, the inclusion of counter ions would result

in large differences in the effective ionic strength. It is possible that the deviations in the case of 1cei are due to the fact that this molecule is highly charged. However, there is no obvious evidence of internal strain (energy per residue after minimization was -40 kJmol^{-1}). In addition, considering the set of X-ray-derived structures as a whole, there was no significant correlation between the net charge on the molecules and the RMSD after 5 ns (slope = 0.20; correlation coefficient = 0.27). Furthermore, the ROP mutant dimer that deviated only 0.13 nm carried a net charge of -12 (see Table I).

For the NMR structures, which represent proteins in solution, there is no factor equivalent to crystal-packing effects that could explain cases with large RMSD deviations. As illustrated in Figure 2(a), some structures (e.g., native form of mercury detoxification protein, 1afi) remain very close to their initial structures. However, more than a third of the structures examined deviated by $>0.3 \text{ nm}$ backbone RMSD. Two examples, corresponding to Centromere protein B (1bw6) and FMR1 protein (2fmr), are illustrated in Figure 2(b and c). In both of these cases, the initial structures are compact but show considerable structural rearrangements during the simulations. There is no obvious reason why these structures should display such large deviations. No significant correlation was found between the overall RMSD and other properties such as proportion of α -helix and/or β -sheet, the percentage of charged residues, and the initial shape of the molecule as reflected in the radius of gyration. There was also no correlation between the overall RMSD and the apparent quality of the experimental structures as determined by the AVE measure from the PROCHECK series of tests.^{61,62} The highest correlation coefficient between the behavior observed in the simulation and a physical property was 0.62 for the NMR-derived structures. This corresponded to a linear relationship between the backbone RMSD and the overall net charge. However, considering all the structures involved in current work, the correlation between net charge and RMSD was only 0.35, casting doubt on whether the net charge is in fact significant.

To estimate the statistical significance of the difference between the distributions obtained for the X-ray and NMR structures, a two-sample t -test was applied.⁶³ The structures were scored according to their RMSD: 0 = $\text{RMSD} \leq 0.1 \text{ nm}$; 1 = $0.1 \text{ nm} < \text{RMSD} \leq 0.2 \text{ nm}$; 2 = $0.2 \text{ nm} < \text{RMSD} \leq 0.3 \text{ nm}$; 3 = $\text{RMSD} > 0.3 \text{ nm}$. If no allowance is made for the effect of packing forces or the fact that 1rpo and 1ae3 are dimers in solution, the confidence level that the NMR structures show a larger deviation is 0.7–0.8. Statistically, this is not very significant. However, simulating 1rpo and 1ae3 as dimers, the confidence level increases to 0.8–0.9. Furthermore, when 1a32 is removed from the set, as its extended structure is clearly determined by packing forces, and only the central portion of 1pgx is considered, the confidence level increases to 0.95–0.98, which clearly is statistically significant.

Figure 3 shows the time evolution of the RMSD from the respective experimental structures for the five proteins solved by both techniques. Because no coordinates were

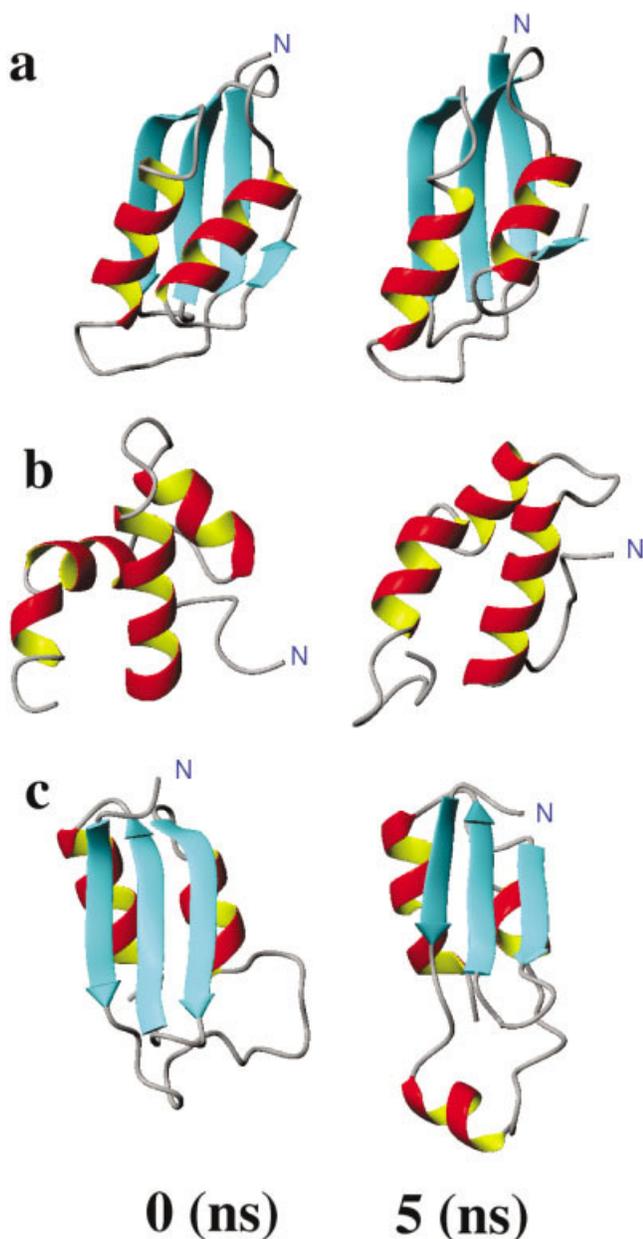


Fig. 2. Conformations of three proteins, the structures of which were solved with NMR techniques. On the left are shown the initial conformations (0 ns); on the right conformations after 5 ns of simulation (5 ns). Bacterial mercury detoxification protein (1afi) (a), centromere protein B (1bw6) (b), and FMR1 protein (2fmr) (c).

given for the two C-terminal residues in the NMR structure of chymotrypsin inhibitor 2 (3ci2), these two residues were deleted from the corresponding X-ray structure (2ci2) before the system was simulated. In addition, there is a difference in the sequence with residue 58GLN in the NMR structure being substituted by 58GLU in the X-ray structure. In the X-ray structure of the ribonuclease inhibitor (1a19), there was a single substitution at residue 82, ALA instead of CYS as in the corresponding NMR structure (1bta). In α -spectrin (SH3 domain), the C-terminal residue was deleted from the NMR structure

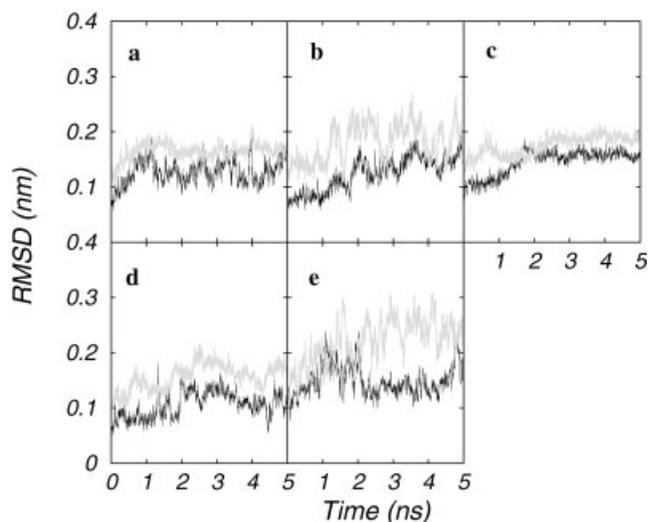


Fig. 3. Time evolution of the backbone RMSD from the starting structure during 5 ns of simulation for each of the five proteins, the structures of which were solved by both X-ray diffraction (black) and NMR (gray). Chymotrypsin inhibitor 2 (2ci2, 3ci2) (a), protein G B1 domain (1pgb, 2gb1) (b), ribonuclease inhibitor, Barstar (1a19, 1bta) (c), α -spectrin (1shg, 1aey) (d), and Ubiquitin (1ubi, 1d3z) (e).

(1aey) before the system was simulated to make it consistent with the corresponding X-ray structure (1shg). As can be readily seen in Figure 3, the RMSD curves for the five X-ray structures (black lines) consistently show lower deviations than the corresponding curves for the NMR structures (gray lines) during the simulations. The differences are not large but are systematic. Applying a Student *t*-test and assuming that the X-ray and NMR structures have an equal probability of deviating more from the experimental structure, we conclude that the NMR structures deviate more than the X-ray-derived structures with a confidence level of > 0.98 .⁶³

As noted by one referee of the manuscript, the results could be biased by the inclusion of residues that experimentally show high relative mobility. The second last column in Tables I and II lists the RMSD calculated after excluding residues, which are highly mobile or for which the positions are uncertain. In the X-ray-derived structures, experimental B-factors were used as an indication of mobility. In the NMR-derived structures, apparent B-factors were used. These apparent B-factors were derived from the RMSD of the atomic positions obtained during the refinement or from the spread of the ensemble of structures deposited in the PDB where RMSD values per atom were not given. We note that inferring absolute mobility of residues from B-factors for comparison with simulation data is not straightforward.⁶⁴ In addition to thermal motion, uncertainty caused by crystal defects and measurement errors can also contribute to crystallographic B-factors. Upper constraints are also often applied to B-factors during the refinement of X-ray structures.³ In structures solved by NMR, the atomic RMSD values included in most files instead of B-factors reflect the spread of structures obtained during the refinement procedure as opposed to true motion. For these reasons, it is not

possible to make a one-to-one comparison between the X-ray-derived B-factors and the apparent B-factors estimated from the spread of the NMR structures. In particular, the latter are on average much larger.

For the X-ray-derived structures, residues were considered highly mobile if the B-factor of the C_{α} was $> 60 \text{ \AA}^2$. For the NMR-derived structures, a larger cutoff of $> 100 \text{ \AA}^2$ was used. These atoms were excluded from the superposition of the simulated structure onto experimental structure and the determination of the overall RMSD. The number of residues excluded in each case is listed in Tables I and II, together with the RMSD. In the X-ray-derived structures given in Table I, 8 of the 20 structures have residues for which the B-factor of the C_{α} is $> 60 \text{ \AA}^2$. However, only in the case of 1vif does excluding the more mobile residues make a significant difference to the RMSD. Clearly, there is no correlation between the overall RMSD found in the simulation and the number of residues with high B-factors.

In NMR-derived structures, the apparent B-factors are highly dependent on the refinement protocol. Two structures (1alz and 2gb1) contain no residues for which the apparent B-factor $\geq 100 \text{ \AA}^2$. In contrast, in three structures (1afi, 2af8, and 1d3z), $> 50\%$ of the residues have apparent B-factors $\geq 100 \text{ \AA}^2$. In three cases (1aoy, 1stu, and 1asp), in which 15, 14, and 17 residues are excluded, respectively, it is apparent that excluding the more mobile residues (or residues for which the positions are uncertain) has a marked effect on the RMSD. Overall, however, there is still little correlation between the number of apparently mobile residues and the calculated RMSD. For example, 1afi is stable in the simulations with a very low overall RMSD despite 64 of 72 residues presenting relatively high apparent B-factors. Clearly, it is not possible to account for the difference in the behavior of the X-ray- and NMR-derived structures in the simulations by excluding residues with high apparent mobility as determined during refinement in a rigorous and consistent manner.

The last column in Tables I and II lists the RMSD calculated after excluding residues found to be highly mobile in the simulation, which could indicate locations dominated by crystal packing forces or regions poorly defined in NMR structures. Highly mobile residues were determined by calculating an apparent B-factor from the root-mean-square fluctuation of atoms obtained during the last nanosecond of the 5-ns simulations. Residues for which any backbone atom had an apparent B-factor of $> 100 \text{ \AA}^2$ were excluded in a similar manner as for the experimentally derived B-factors. Again, the number of residues excluded in each case is listed in Tables I and II, together with the RMSD. In most cases, the number of residues excluded is small, between 0 and 10% of the total number of residues. In almost all cases, the highly mobile residues corresponded to residues at the N- and/or C-termini. Elimination of the most highly mobile residues in general results in a reduction in the RMSD. Note that because the RMSD is calculated from an average structure, the elimination of highly mobile residues does not necessarily result in a decrease in the overall RMSD if the

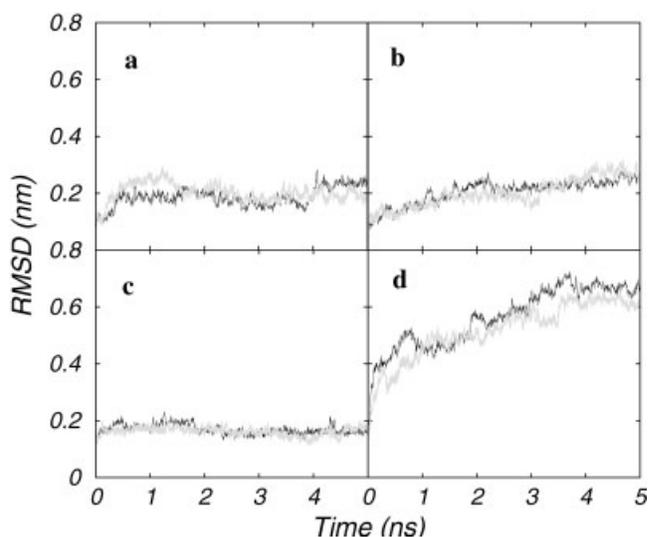


Fig. 4. The RMSD from the starting structure as a function of time illustrating the sensitivity of the simulations to changes in the initial velocities. α -Spectrin (1tuc) (a), DNA topoisomerase I (1vcc) (b), bacterial mercury detoxification protein (1afi) (c), actinorhodin polyketide synthase acyl carrier (2af8) (d).

fluctuations were with respect to a well-defined average position. For only one structure is there a clear anomaly. When 1rpo was simulated as a monomer, $> 60\%$ of the residues had high apparent B-factors as a result of the molecule deforming. However, simulated as a dimer, only two residues were considered highly mobile. Considering 1rpo and 1ae3 as dimers but making no other corrections, we find of the 20 X-ray-derived structures, 4 deviate $\leq 0.1 \text{ nm}$, 11 $\leq 0.2 \text{ nm}$, 18 $\leq 0.3 \text{ nm}$, and 2 deviate $> 0.3 \text{ nm}$. Of the 19 NMR-derived structures, 0 deviate $\leq 0.1 \text{ nm}$, 9 $\leq 0.2 \text{ nm}$, 13 $\leq 0.3 \text{ nm}$ shown, and 6 deviate $> 0.3 \text{ nm}$. Thus, the overall trend remains the same after the exclusion of highly mobile residues. If anything, the differences between the X-ray- and NMR-derived structures is more evident. We also note that no allowance is required in 1pgr because the extended N- and C-terminal regions observed in the crystal structure are highly mobile in the simulation and thus excluded automatically from the fit.

Consistency Checks

To test the consistency of the simulations and to determine the sensitivity of the simulations to the choice of starting velocities, four structures were selected at random and the simulations were repeated with different starting velocities. A comparison of RMSD from the starting structure over 5 ns for each of the proteins is shown in Figure 4. For each of the structures, two derived from X-ray diffraction and two using NMR, similar behavior was observed irrespective of the initial velocities. The choice of initial atom velocities did not result in large differences in the evolution of the structures in 5-ns MD simulation. Therefore, it is unlikely that the difference observed in the behavior of structures derived from X-ray or NMR studies is simply due to statistical fluctuations.

To examine the convergence of the systems as a function of time, the simulations of two of the proteins (1tuc and

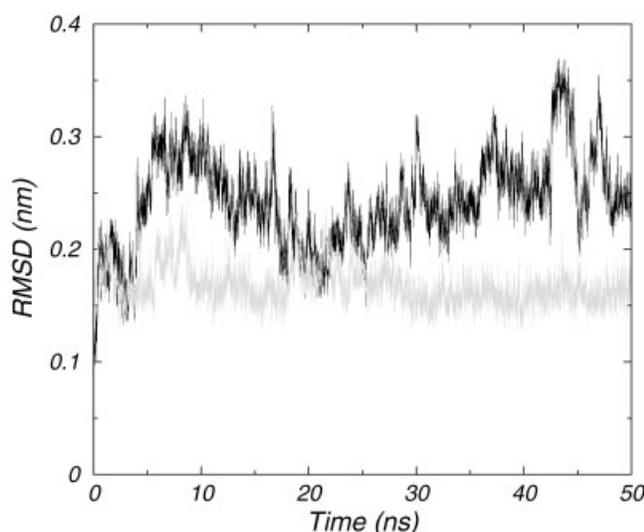


Fig. 5. The RMSD from the starting structure as a function of time illustrating the effect of extending the simulation time for α -spectrin 1tuc (black) and bacterial mercury detoxification protein (gray).

1afi) were extended to 50 ns. The corresponding RMSD plots are presented in Figure 5. In these two cases, no large structural fluctuations were observed after the first 5 ns. This finding suggests that structural changes induced by stress within the initial structures can manifest rapidly and that although longer simulation times are clearly desirable, 5 ns is sufficient to assess the relative stability of the initial structures given the 39 structures incorporated in this study.

CONCLUSIONS

Using molecular dynamics simulation techniques, we have analyzed the relative stability of 34 proteins, the structures of which were determined by X-ray crystallography, by NMR spectroscopic techniques, or by both. Comparing the RMSD from the experimentally derived structure after 5 ns of simulation we find that, on average, X-ray-derived structures are significantly more stable (show smaller deviations than NMR-derived structures). No correlation between the magnitude of the RMSD from the starting structure after 5 ns and any property of the system (such as size, secondary structure content, or charge distribution) other than the method used to determine the structures could be identified. In addition, we found that, on average, the NMR-derived structures had substantially higher internal energies than the X-ray-derived structures of comparable size.

It could be argued that because the structures of proteins determined by NMR reflect that in solution, without the distortions imposed by crystal-packing forces, NMR structures should distort less than X-ray-derived structures when simulated as isolated molecules. Except in rare cases in which the structure in the crystal is clearly dominated by crystal-packing forces, this appears not to be the case. The work suggests that the current approaches used to determine NMR structures frequently yield struc-

tures with considerable internal strain.⁶⁵ What we have not done is attempt to determine which of the various NMR refinement protocols most often lead to unstable structures. For this, the sample size is too small, and there is already much debate in the literature about the relative merits of different refinement approaches.^{66–72} Overall, we must conclude that currently, structures determined with use of X-ray techniques, when available, are in general more suited for use as a starting point for modeling studies to investigate protein motions or protein-ligand interactions.

REFERENCES

1. Branden C, Tooze J. Introduction to protein structure. New York: Garland Publishing; 1991.
2. Billeter M. Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals. *Q Rev Biophys* 1992;25:325–377.
3. Drenth J. Principles of protein x-ray crystallography. New York: Springer-Verlag; 1999.
4. Wuthrich K. NMR of proteins and nucleic acids. New York: Wiley-Interscience; 1986.
5. Robillard GT, Tarr CE, Vosman F, Berendsen HJC. Similarity of the crystal and solution structure of yeast tRNAPhe. *Nature* 1976;262:363–369.
6. Williams G, Clayden NJ, Moore GR, Williams RJ. Comparison of the solution and crystal structures of mitochondrial cytochrome c. Analysis of paramagnetic shifts in the nuclear magnetic resonance spectrum of ferricytochrome c. *J Mol Biol* 1985;183:447–460.
7. Billeter M, Kline AD, Braun W, Huber R, Wuthrich K. Comparison of the high-resolution structures of the alpha-amylase inhibitor tandemistat determined by nuclear magnetic resonance in solution and by x-ray diffraction in single crystals. *J Mol Biol* 1989;206:677–687.
8. Billeter M, Vendrell J, Wider G, Aviles FX, Coll M, Guasch A, Huber R, Wuthrich K. Comparison of the NMR solution structure with the x-ray crystal structure of the activation domain from procarboxypeptidase B. *J Biomol NMR* 1992;2:1–10.
9. Blanco FJ, Ortiz AR, Serrano L. 1H and 15N NMR assignment and solution structure of the SH3 domain of spectrin: comparison of unrefined and refined structure sets with the crystal structure. *J Biomol NMR* 1997;9:347–357.
10. Narayana N, Matthews DA, Howell EE, Nguyen-Huu X. A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D2-symmetric active site. *Nat Struct Biol* 1995;2:1018–1025.
11. Viguera AR, Blanco FJ, Serrano L. The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. *J Mol Biol* 1995;247:670–681.
12. Sharma A, Hanai R, Mondragon A. Crystal structure of the amino-terminal fragment of vaccinia virus DNA topoisomerase I at 1.6 Å resolution. *Structure* 1994;2:767–777.
13. Liu J, Lynch PA, Chien CY, Montelione GT, Krug RM, Berman HM. Crystal structure of the unique RNA-binding domain of the influenza virus NS1 protein. *Nat Struct Biol* 1997;4:896–899.
14. Chak KF, Safo MK, Ku WY, Hsieh SY, Yuan HS. The crystal structure of the immunity protein of colicin E7 suggests a possible colicin-interacting surface. *Proc Natl Acad Sci USA* 1996;93:6437–6442.
15. Vlasi M, Steif C, Weber P, Tsernoglou D, Wilson KS, Hinz HJ, Kokkinidis M. Restored heptad pattern continuity does not alter the folding of a four-alpha-helix bundle. *Nat Struct Biol* 1994;1:706–716.
16. Clemons WM Jr, Davies C, White SW, Ramakrishnan V. Conformational variability of the N-terminal helix in the structure of ribosomal protein S15. *Structure* 1998;6:429–438.
17. Su S, Gao YG, Zhang H, Terwilliger TC, Wang AH. Analyses of the stability and function of three surface mutants (R82C, K69H, and L32R) of the gene V protein from Ff phage by X-ray crystallography. *Protein Sci* 1997;6:771–780.
18. Leijonmarck M, Liljas A. Structure of the C-terminal domain of the ribosomal protein L7/L12 from *Escherichia coli* at 1.7 Å. *J Mol Biol* 1987;195:555–579.

19. Achari A, Hale SP, Howard AJ, Clore GM, Gronenborn AM, Hardman KD, Whitlow M. 1.67-Å X-ray structure of the B2 immunoglobulin-binding domain of streptococcal protein G and comparison to the NMR structure of the B1 domain. *Biochemistry* 1992;31:10449–10457.
20. Biou V, Shu F, Ramakrishnan V. X-ray crystallography shows that translational initiation factor IF3 consists of two compact alpha/beta domains linked by an alpha-helix. *EMBO J* 1995;14:4056–4064.
21. Thunnissen MM, Taddei N, Liguri G, Ramponi G, Nordlund P. Crystal structure of common type acylphosphatase from bovine testis. *Structure* 1997;5:69–79.
22. Fukuyama K, Okada T, Kakuta Y, Takahashi Y. Atomic resolution structures of oxidized [4Fe-4S] ferredoxin from *Bacillus Thermoproteolyticus* in two crystal forms: systematic distortion of [4Fe-4S] cluster in the protein. *J Mol Biol* 2002;315:1155–1166.
23. Mondragon A, Subbiah S, Almo SC, Drottner M, Harrison SC. Structure of the amino-terminal domain of phage 434 repressor at 2.0 Å resolution. *J Mol Biol* 1989;205:189–200.
24. Xu RM, Koch C, Liu Y, Horton JR, Knapp D, Nasmyth K, Cheng X. Crystal structure of the DNA-binding domain of Mbp1, a transcription factor important in cell-cycle control of DNA synthesis. *Structure* 1997;5:349–358.
25. McPhalen CA, James MN. Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry* 1987;26:261–269.
26. Gallagher T, Alexander P, Bryan P, Gilliland GL. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* 1994;33:4721–4729.
27. Musacchio A, Noble M, Pauptit R, Wierenga R, Saraste M. Crystal structure of a Src-homology 3 (SH3) domain. *Nature* 1992;359:851–855.
28. Ramage R, Green J, Muir TW, Ogunjobi OM, Love S, Shaw K. Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem J* 1994;299:151–158.
29. Ratnaparkhi GS, Ramachandran S, Udgaonkar JB, Varadarajan R. Discrepancies between the NMR and X-ray structures of uncomplexed barstar: analysis suggests that packing densities of protein structures determined by NMR are unreliable. *Biochemistry* 1998;37:6958–6966.
30. Sunnerhagen M, Nilges M, Otting G, Carey J. Solution structure of the DNA-binding domain and model for the complex of multifunctional hexameric arginine repressor with DNA. *Nat Struct Biol* 1997;4:819–825.
31. Bycroft M, Grunert S, Murzin AG, Proctor M, St Johnston D. NMR solution structure of a dsRNA binding domain from *Drosophila* staufer protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J* 1995;14:3563–3571.
32. Bycroft M, Hubbard TJ, Proctor M, Freund SM, Murzin AG. The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell* 1997;88:235–242.
33. Edmondson SP, Qiu L, Shriver JW. Solution structure of the DNA-binding protein Sac7d from the hyperthermophile *Sulfolobus acidocaldarius*. *Biochemistry* 1995;34:13289–13304.
34. Steele RA, Opella SJ. Structures of the reduced and mercury-bound forms of MerP, the periplasmic protein from the bacterial mercury detoxification system. *Biochemistry* 1997;36:6885–6895.
35. Connolly KM, Wojciak JM, Clubb RT. Site-specific DNA binding using a variation of the double stranded RNA binding motif. *Nat Struct Biol* 1998;5:546–550.
36. Groft CM, Uljon SN, Wang R, Werner MH. Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. *Proc Natl Acad Sci USA* 1998;95:9117–9122.
37. Musco G, Kharrat A, Stier G, Fraternali F, Gibson TJ, Nilges M, Pastore A. The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nat Struct Biol* 1997;4:712–716.
38. Eberstadt M, Huang B, Chen Z, Meadows RP, Ng SC, Zheng L, Lenardo MJ, Fesik SW. NMR structure and mutagenesis of the FADD (Mort1) death-effector domain. *Nature* 1998;392:941–945.
39. Iwahara J, Kigawa T, Kitagawa K, Masumoto H, Okazaki T, Yokoyama S. A helix-turn-helix structure unit in human centromere protein B (CENP-B). *EMBO J* 1998;17:827–837.
40. Jeon YH, Negishi T, Shirakawa M, Yamazaki T, Fujita N, Ishihama A, Kyogoku Y. Solution structure of the activator contact domain of the RNA polymerase alpha subunit. *Science* 1995;270:1495–1497.
41. Fogh RH, Otleben G, Ruterjans H, Schnarr M, Boelens R, Kaptein R. Solution structure of the LexA repressor DNA binding domain determined by 1H NMR spectroscopy. *EMBO J* 1994;13:3936–3944.
42. Crump MP, Crosby J, Dempsey CE, Parkinson JA, Murray M, Hopwood DA, Simpson TJ. Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2). *Biochemistry* 1997;36:6000–6008.
43. Clubb RT, Schumacher S, Mizuuchi K, Gronenborn AM, Clore GM. Solution structure of the I gamma subdomain of the Mu end DNA-binding domain of phage Mu transposase. *J Mol Biol* 1997;273:19–25.
44. Ludvigsen S, Shen HY, Kjaer M, Madsen JC, Poulsen FM. Refinement of the three-dimensional solution structure of barley serine proteinase inhibitor 2 and comparison with the structures in crystals. *J Mol Biol* 1991;222:621–635.
45. Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* 1991;253:657–661.
46. Cornilescu G, Marquardt JL, Ottiger M, Bax A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J Am Chem Soc* 1998;120:6836–6837.
47. Lubienski MJ, Bycroft M, Freund SM, Fersht AR. Three-dimensional solution structure and 13C assignments of barstar using nuclear magnetic resonance spectroscopy. *Biochemistry* 1994;33:8866–8877.
48. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–242.
49. Simons KT, Strauss C, & Baker D. Prospects for ab initio protein structural genomics. *J Mol Biol* 2001;306:1191–1199.
50. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comp Phys Comm* 1995;91:43–56.
51. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 2001;7:306–317.
52. van der Spoel D, van Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, Hess B, Feenstra KA, Lindahl E, van Drunen R, Berendsen HJC. Gromacs user manual version 3.0. Nijenborgh 4, 9747 AG Groningen, The Netherlands, 2001; Internet: <http://www.gromacs.org>.
53. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger PKHC, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: the GROMOS96 manual and user guide. vdf Hochschulverlag AG, Zürich, 1996.
54. van Gunsteren WF, Daura X, Mark AE. The GROMOS force field. In: von Rague Schleyer P, editor. *Encyclopedia of computational chemistry*, vol. 2. New York: John Wiley & Sons; 1998. p 1211–1216.
55. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. *Intermolecular forces*. Dordrecht: Reidel; 1981. p 331–342.
56. Tironi IG, Sperb R, Smith PE, van Gunsteren WF. A generalized reaction field method for molecular dynamics simulations. *J Chem Phys* 1995;102:5451–5459.
57. Feenstra KA, Hess B, Berendsen HJC. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comp Chem* 1999;20:786–798.
58. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comp Chem* 1997;18:1463–1472.
59. Miyamoto S, Kollman PA. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comp Chem* 1992;13:952–962.
60. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81:3684–3690.
61. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PRO-

- CHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26:283–291.
62. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins* 1992;12:345–364.
 63. Grafen A, Hails R. *Modern statistics for the life sciences*. New York: Oxford University; 2002.
 64. Hünenberger PH, Mark AE, van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol* 1995;252:492–503.
 65. van Gunsteren WF. On testing theoretical models by comparison of calculated with experimental data. *Stud Phys Theor Chem* 1990;71:463–478.
 66. Brunger AT, Nilges M. Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy. *Q Rev Biophys* 1993;26:49–125.
 67. Nilges M. Structure calculation from NMR data. *Curr Opin Struct Biol* 1996;6:617–623.
 68. Nilges M. Ambiguous distance data in the calculation of NMR structures. *Fold Des* 1997;2:S53–57.
 69. Kuszewski J, Gronenborn AM, Clore GM. Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J Magn Reson* 1997;125:171–177.
 70. Doreleijers JF, Rullmann JA, Kaptein R. Quality assessment of NMR structures: a statistical survey. *J Mol Biol* 1998;281:149–164.
 71. Doreleijers JF, Raves ML, Rullmann T, Kaptein R. Completeness of NOEs in protein structure: a statistical analysis of NMR. *J Biomol NMR* 1999;14:123–132.
 72. Spronk CA, Linge JP, Hilbers CW, Vuister GW. Improving the quality of protein structures derived by NMR spectroscopy. *J Biomol NMR* 2002;22:281–289.