Refinement of Protein Structures in Explicit Solvent

Jens P. Linge,[1] Mark A. Williams,[2] Christian A.E.M. Spronk,[3] Alexandre M. J. J. Bonvin,[4]* and Michael Nilges[1]*

[1]*Unité de Bio-Informatique Structurale, Institut Pasteur, Paris, France*
[2]*Department of Biochemistry and Molecular Biology, University College London, London, United Kingdom*
[3]*Centre for Molecular and Biomolecular Informatics, University of Nijmegen, Nijmegen, The Netherlands*
[4]*Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*

ABSTRACT We present a CPU efficient protocol for refinement of protein structures in a thin layer of explicit solvent and energy parameters with completely revised dihedral angle terms. Our approach is suitable for protein structures determined by theoretical (e.g., homology modeling or threading) or experimental methods (e.g., NMR). In contrast to other recently proposed refinement protocols, we put a strong emphasis on consistency with widely accepted covalent parameters and computational efficiency. We illustrate the method for NMR structure calculations of three proteins: interleukin-4, ubiquitin, and crambin. We show a comparison of their structure ensembles before and after refinement in water with and without a force field energy term for the dihedral angles; crambin was also refined in DMSO. Our results demonstrate the significant improvement of structure quality by a short refinement in a thin layer of solvent. Further, they show that a dihedral angle energy term in the force field is beneficial for structure calculation and refinement. We discuss the optimal weight for the energy constant for the backbone angle omega and include an extensive discussion of meaning and relevance of the calculated validation criteria, in particular root mean square Z scores for covalent parameters such as bond lengths. Proteins 2003;50: 496–506. © 2003 Wiley-Liss, Inc.

© 2003 Wiley-Liss, Inc.

INTRODUCTION

Rapid calculation of protein structures necessitates severe simplifications of the nonbonded interactions. The unrealistic treatment of electrostatic and van der Waals (vdW) interactions can lead to artefacts, such as nonoptimal packing and unsatisfied hydrogen bond donors or acceptors. As a consequence, validation programs often yield unsatisfactory quality indices such as the Ramachandran plot appearance or the side-chain packing quality for these structures. Refinement in a full molecular dynamics force field including electrostatic and Lennard–Jones nonbonded potentials and interaction with solvent is one way to improve the quality of the structures. The aim is to bring the structures closer to physical reality without moving

them away from the data. However, care has to be taken to avoid new artefacts introduced by the force field.

Energy parameters for structure calculation and refinement have to meet several requirements. The energy parameters should maintain the structure appropriately close to ideal geometry, yet ensure that the final structures accurately reflect all of the experimental data. The relative weights on experimental and geometric terms in the force field must be correctly balanced. For this purpose, standard molecular dynamics force fields may not put sufficient weight on some energy terms, for example, planarity, because they are developed for dynamics calculations rather than refinement. Computational efficiency is a major issue, especially for NMR, where large ensembles of structures are iteratively calculated. Consequently, the physical reality of the nonbonded attractive and repulsive interactions is often only crudely approximated by a simple repulsive volume exclusion term for the structure calculation.

Several recent articles have shown that refinement in explicit solvent [1–3] significantly improves NMR structures. In our previous study,[1] the force field used in the refinement was a hybrid force field based on the covalent parameters defined by Engh and Huber for X-ray crystallography refinement[4] and slightly modified nonbonded parameters from the OPLS force field.[5] The main reason for developing this hybrid force field was to provide a more realistic description of nonbonded interactions in the refinement without modifying the covalent parameters. This is important because all common structure validation programs (e.g., WHATCHECK,[6] PROCHECK,[7] and SQUID[8]) are based on the Engh and Huber parameters. In contrast, common molecular dynamics force fields show systematic deviations from these parameters, and may use different relative weighting of energy terms. A recent article[2] presented a significant improvement over our previous results. The authors used CHARMM22[9] parameters, immersed the protein in a large solvent box, and used longer simulation times. It was therefore important to investigate if their results were due to the force field or the

*Correspondence to: Alexandre M. J. J. Bonvin, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. E-mail: abonvin@nmr.chem.uu.nl, or Michael Nilges, Unité de Bio-Informative Structurale, Institut Pasteur, 25–28 rue du Dr Roux, 75015 Paris, France

refinement procedure. In this article, we present calculations with an improved refinement protocol and revised energy parameters. The validation results are as good as those obtained in Spronk et al.,[2] indicating that a minimal approach with just a few solvent layers is sufficient to obtain significant improvements in validation parameters over unrefined structures. The refinement is CPU efficient, in particular when compared to other methods, most notably refinement with a generalized Born energy function.[3]

MATERIALS AND METHODS

We derived the water refinement protocol and the energy parameters from those used for the calculations in Linge and Nilges.[1] We used three data sets: interleukin-4 (IL-4), crambin, and ubiquitin.

NMR Data

We calculated IL-4 structures from the NMR data of Protein Data Bank (PDB) entry 1BBN,[10] which comprises 918 unambiguously assigned NOEs (329 long, 305 medium, and 284 short range) and 174 torsion angles restraints (101 ϕ and 73 ψ restraints). The amino acid sequence contains 133 residues and has 4 additional residues (Glu Ala Glu Ala) at the N-terminus and two Asn to Asp mutations for residues 38 and 105 compared to the wild type. We defined the three disulphide bridges (between cysteines 3–127, 24–65, and 46–99) in the topology and not separately as distance restraints. Essentially, we repeated calculations 5 and 8 from our previous article (see Table 2 in Linge and Nilges[1]) with the revised protocols in ARIA 1.2[11] and the updated PARALLHDG 5.3 force field. The X-ray structures 2INT (2.25 Å resolution)[12] and 1RCB (2.35 Å resolution)[13] served as references for root mean square derivative (RMSD) calculations.

For the 76-residue protein ubiquitin, we took the NMR data of PDB entry 1D3Z,[14] comprising 2727 distance restraints, 27 hydrogen bond restraints, and 98 torsion angle restraints (63 ϕ and 35 χ_1 restraints). We did not include the residual dipolar couplings as restraints for the simulated annealing refinement, but used them afterward to calculate Q factors as an independent validation measure. The 1.8 Å X-ray structure 1UBQ[15] served as a reference structure for the calculation of the RMSD values.

We calculated structures of the 46-residue protein crambin solely with a list of 1517 distance restraints derived from the 0.54 Å X-ray structure 1EJG[16] (data stem from Spronk et al.[2]). We defined the disulphide bonds between cysteines 3–40, 4–32, and 16–26 in the topology and did not employ any hydrogen bonds or torsion angle restraints.

PARALLHDG Force Field

Since version 5.0 of the PARALLHDG parameters,[1] the covalent parameters have been based on the CSDX parameters developed for X-ray crystal structure refinement[4] with additional parameters involving hydrogen atoms.

PARALLHDG includes the nonbonded interaction from PROLSQ[17] for the use in simulated annealing and gas-phase refinement, the OPLS parameters[5] with some minor

modifications, and the TIP3P water model[18] for refinement in explicit solvent. In version 5.3, we added parameters for DMSO stemming from Lin et al.[19]

We enforced planarity for the guanidino group of arginine side-chains since version 5.0. The absence of this term resulted in bad statistics of side-chain planarity in NMR structures calculated with X-PLOR in the analysis of Doreleijers et al.,[20] who had not distinguished between the different planar side-chains. We note that there still is no uniform treatment of arginine planarity between different structure refinement packages. In version 5.3 we also added a restraint for tyrosine hydroxyl, which tends to be planar with the ring because of sp^2 hydridization. Staggered conformations with respect to bonded groups are now preferred for CH_3 and NH_3 hydrogens. We modified torsions at the N-terminus to reflect the effects of the NH to NH_3 substitution. All of these changes reflect the conformations energetically preferred by the local chemistry around a particular bond. By inspecting peptide structures in the Cambridge Structural Database,[21] we obtained a few new bond angles for C- and N- terminal groups.

The main changes in the current release (5.3) concern the dihedral angles introduced in version 5.0. We modified (or even introduced for the first time) virtually all parameters involving triply substituted carbons bonded to doubly substituted carbons in respect of periodicity, position of minima, and force constant. The aim of the new backbone dihedral angle parameters is to reflect local energy barriers of the backbone yet avoid emphasising particular (long-range) preferences for secondary structure. In devising the new parameters we took note of the analysis of the ϕ and ψ angles of high-resolution protein crystal structures subdivided by secondary structure type, including coil regions.[22] Regardless of secondary structure, the consistent features of the distributions are the unlikelihood of finding ψ of (60 to 80) or (-100 to -120) and ϕ of (-40 to 40) and (120 to 180). These features arise because of local chemistry; in this two- to three-fold junction the three-fold carbon carries two large substituents. Single cosine functions for ϕ and ψ cannot adequately represent this pattern of dominant maxima. We introduced new dihedral angle terms for the ϕ and ψ backbone angles C_β-C_α-C-N, C_β-C_α-C-O, C-N-C_α-C_β, and C-N-C_α-C. The *cis* conformations of each of these are the dominant energetic maxima for ϕ and ψ rotations. Each dihedral angle term has an associated energy function with a single narrow maximum at the *cis* position. This function (shown in Fig. 1) is a composite of six cosine terms:

$$E_{\text{dih}} = k(0.5(1 + \cos \alpha) + 0.4(1 + \cos 2\alpha) + 0.3(1 + \cos 3\alpha)$$

$$+ 0.2(1 + \cos 4\alpha) + 0.1(1 + \cos 5\alpha) + 0.05(1 + \cos 6\alpha)),$$

$$(1)$$

where k is the energy constant. Figure 2 depicts the resultant of the four dihedral angle terms with k in the ratio of 1:0.5:1:0.75. This reflects the relative height of the energy barrier due to each *cis* conformer. The parameters are suitable for both L and D amino acids. The composite cosine function excludes the known sparsely populated
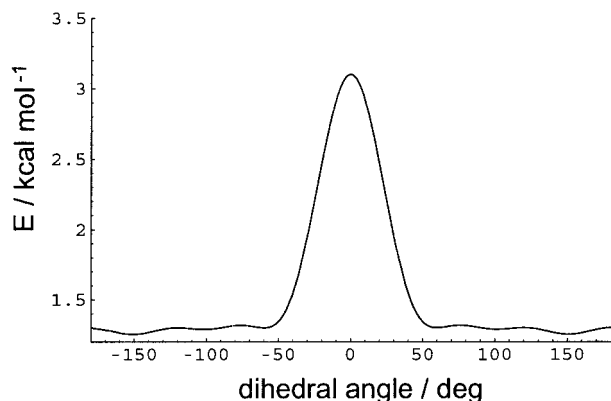
Fig. 1. Functional form of the composite cosine term in eq. 1 showing the single maximum at the *cis* position.
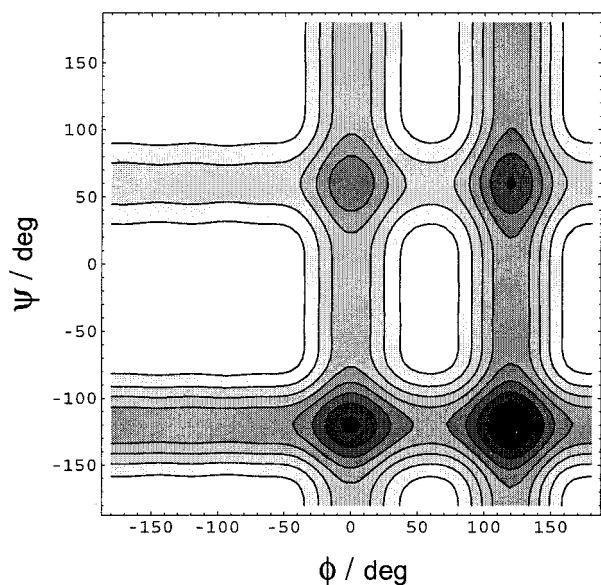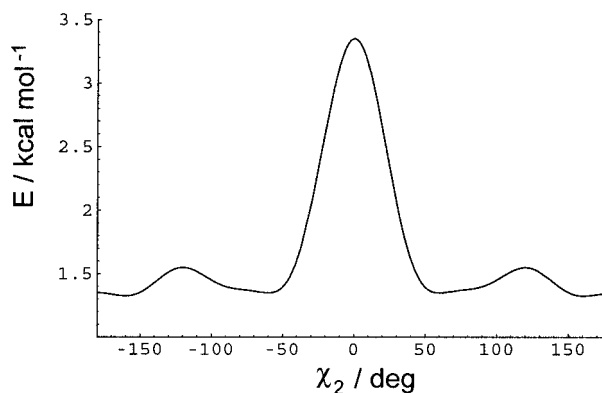

Fig. 3. Functional form of the composite cosine term in eq. 2 representing the barriers to rotation around an S—S bond. The principle maximum occurs when C_β atoms are close and the subsidiary maxima when C_β and H_β are close.

Fig. 2. Ramachandran map illustrating the aggregate effect of the four dihedral function for C_β-C_α-C-N, C_β-C_α-C-O, C-N-C_α-C_β, and C-N-C_α-C on the energy of the backbone conformation. Contours at 0.4 kcal mol^{-1}; intervals with darker shading representing higher energy.

term for C_α-C-N-C_α represents the angle ω. The new *cis* peptide patch changes the atom type. This makes it easy to reduce the ω energy constant during refinement.

Simulated Annealing Protocol for Structure Calculation

We calculated structures with the standard protocol of ARIA 1.2.[11] The simulated annealing protocol consisted of four stages: a high-temperature torsion angle simulated annealing phase[24,25] at 10,000 K (1100 steps), a first torsion angle dynamics cooling phase from 10,000 K to 2000 K (550 steps), a second Cartesian dynamics cooling phase from 2000 K to 1000 K (5000 steps), and a third Cartesian dynamics cooling phase from 1000 K to 0 K (2000 steps). We used a timestep of 45 fs for the torsion angle dynamics and 5 fs for the Cartesian dynamics. During the high-temperature stage we included only two atoms per residue in the nonbonded interactions,[26] during the first and second cooling stages we did not explicitly take hydrogens into account, and during the third cooling stage we included all atoms. In the second cooling stage, we reduced the asymptote of the soft-square potential to 0.1. We employed the energy term for the dihedral angles only in the third cooling stage. The masses were 100 amu uniformly; the "friction coefficient" f_β[27] for the coupling to the external bath was 20 ps^{-1}. Note that energy constants, masses, temperatures, and molecular dynamics time-scales are not independent; dynamics at 100 K with uniform masses of 100 amu and a timestep of 5 fs corresponds to dynamics at $100/15 \cdot 100$ K $\approx$ 667 K with uniform masses of 15 amu and a timestep of $\sqrt{15/100} \cdot 5$ fs $\approx$ 1.94 fs. We used a floating chirality approach[28] when stereospecific assignments were not available. After the simulated annealing protocol, we minimized the structures with 200 steps of restrained Powell minimization. We calculated 100 structures and sorted the structures with respect to total energy. From the ensemble of 100 structures, we chose the 25 best structures regarding total energy as starting structures for the water refinement. Because of the sparsity of data and the larger size of IL-4, we used a longer protocol for IL-4 (twice the number of steps).

regions of the Ramachandran map, but does not tightly constrain the backbone to the highly populated α or β regions as those particular preferences are the result of long-range effects.

We also modified the disulphide dihedral angle. The angle distributions for disulphides are notoriously bad in NMR structures,[23] probably because there are rarely any direct experimental restraints on this angle. A composite cosine series with a large maximum corresponding to the C_β—S—S—C_β *cis* conformation and much smaller maxima at ±120° (see Fig. 3) represents the new parameters

$$E_{dih} = (0.5(1 + \cos \chi_2) + 0.4(1 + \cos 2\chi_2) + 0.4(1 + \cos 3\chi_2)$$
$$+ 0.2(1 + \cos 4\chi_2) + 0.1(1 + \cos 5\chi_2) + 0.075(1 + \cos 6\chi_2)).$$
$$(2)$$

We completely revised the terms for the peptide group planarity and the *cis* peptide patch. A single improper

TABLE I. Overview of the Restrained Molecular Dynamics Protocol Used During the Refinement in Explicit Solvent

Stage	N_{steps}	Temperature (K)	K_{angles} (kcal mol^{-1} rad^{-2})	K_{improper} (kcal mol^{-1} rad^{-2})	K_{ω} (kcal mol^{-1} rad^{-2})	K_{NOE} (kcal mol^{-1} Å^{-2})
Minimization	120					
Heating	1000	$100 \rightarrow 500$	50	5	5	50
Refinement	1000	500	50	5	5	50
Cooling	2000	$500 \rightarrow 25$	$50 \rightarrow 500$	$5 \rightarrow 500$	$5 \rightarrow 100$	50
Minimization	200					

Water Refinement

We used a full nonbonded representation during the water refinement, including Lennard–Jones vdW and electrostatic interactions from the OPLS force field with minor modifications. Our protocol updated the nonbonded list every time an atom had moved more than 0.5 Å, with a cutoff for the nonbonded list generation of 9.5 Å. We calculated the Coulomb interaction with a potential shift,[29] which brought the energy to zero at a distance of 8.5 Å.

In the beginning of the refinement, we immersed the structures in a 7.0 Å shell of water molecules with a minimum distance between a protein heavy atom and a water oxygen of 4.0 Å. We chose this large minimum distance to avoid water molecules being inserted into holes in structures, which can be a consequence of bad packing in NMR structures due to low restraint density. During a short minimization (120 steps) a parabolic restraint[30] with an energy constant of 10 kcal mol^{-1} Å^{-2} kept all nonhydrogen atoms (including side-chains and solvent) close to their starting positions.

We modified the water refinement protocol from Linge and Nilges[1] without increasing the total simulation time significantly (summarized in Table I). It consisted of three stages, a heating stage from 100 to 500 K in steps of 100 K with 200 steps of molecular dynamics at each temperature, a short refinement stage with 1000 steps at 500 K, and a cooling stage from 500 K to 25 K in steps of 25 K with 100 steps of molecular dynamics at each temperature. During the heating, we phased out the positional restraints. In the heating and high-temperature phase, we reduced the bond angle and improper torsion angle energy constants to 50 and 5 kcal mol^{-1} rad^{-2}, respectively, to obtain a flexibility similar to standard molecular dynamics force fields. Without this reduction in the high-temperature phase, the structures would remain close to their starting positions. During cooling, we brought the constants back to their normal values of 500 kcal mol^{-1} rad^{-2}, with the exception of the improper torsion angle term for the backbone angle ω, which was only 100 kcal mol^{-1} rad^{-2}. For ubiquitin, we also used values of 20 and 500 kcal mol^{-1} rad^{-2} to investigate whether the ω energy constant has a significant effect on the overall structure (see Results). We increased the weight on the dihedral angle energy term to 500 kcal mol^{-1} rad^{-2} during cooling to maintain the side-chain quality parameters approximately at the same value during refinement. Two-hundred steps of conjugate gradient minimization followed after cooling to 25 K. We kept all energy constants for experimental energy terms at exactly the values that they had at the end of the second cooling phase of the simulated annealing protocol.

DMSO Refinement

A cubic box containing 208 molecules built the solvent shell for the DMSO refinement, with a box size of 29.096 Å, resulting in the correct experimental density of 1.0955 g cm^{-3} for DMSO at 298 K. We equilibrated this box for 1 ns at 300 K under NVT conditions using the DMSO parameters from GROMOS.[19] The total energy of the system was -10197 kJ mol^{-1}. Because DMSO is larger than water, we used a 12.5 Å solvation shell with a 2.4 Å protein–DMSO minimum distance to generate the solvent shell. Apart from that, the refinement protocol is exactly the same as for the water refinement. ARIA 1.2[11] can automatically refine the final structures in an explicit layer of DMSO after the last iteration of NOE assignment and structure calculation.

Structure Validation

From the ensemble of 100 calculated structures, we chose the 20 best structures regarding total energy for structure validation. After the water refinement, we validated the 20 best structures regarding NOE energy from the ensemble of 25 water-refined structures. We used the program WHATCHECK[6] to calculate Z scores for the Ramachandran map (RAMCHK), the packing quality (QUACHK and NQACHK), the χ_1–χ_2 correlation (C12CHK), and the backbone conformation (BBCCHK). We determined root mean square (RMS) Z scores for the bond lengths (BNDCHK), bond angles (ANGCHK), omega angles (OMECHK), side-chain planarity (PLNCHK), improper dihedral angles (HNDCHK), and the inside/outside distribution (INOCHK). We also counted the number of interatomic bumps (BMPCHK), unsatisfied H-bond donors (BH2CHK), and acceptors (BA2CHK).

To discuss the Z scores and RMS Z scores[31] in more detail (see Results), we briefly summarize their definitions. The Z score is the deviation of a value α from the database (db)-derived average value $\langle \alpha_{\text{db}} \rangle$ (in units of the standard deviation $\sigma(\alpha_{\text{db}})$ of α in the database):

$$Z = \frac{\alpha - \langle \alpha_{db} \rangle}{\sigma(\alpha_{db})}. \tag{3}$$

For example, consider the bond lengths b_{ij} in a given protein, where each bond of type I ($i \in I$) occurs J ($j \in J_i$) times. J_i is the number of bonds of type i in the protein.

TABLE II. Quality Indices for IL-4

Performed checks	Without force field dihedrals		With force field dihedrals	
	Unrefined	Water refined	Unrefined	Water refined
WHATCHECK Z scores				
1st-generation packing quality (QUACHK)	-2.2 ± 0.2	-1.1 ± 0.3	-2.0 ± 0.2	-1.0 ± 0.3
2nd-generation packing quality (NQACHK)	-3.3 ± 0.3	-2.1 ± 0.4	-3.3 ± 0.3	-1.8 ± 0.5
Ramachandran plot appearance (RAMCHK)	-4.5 ± 0.4	-2.9 ± 0.4	-4.8 ± 0.4	-3.0 ± 0.5
χ_1–χ_2 rotamer normality (C12CHK)	-4.7 ± 0.2	-4.9 ± 0.2	-1.2 ± 0.4	-1.6 ± 0.3
Backbone conformation (BBCCHK)	-4.8 ± 0.9	-4.2 ± 1.0	-4.8 ± 1.0	-4.4 ± 1.0
WHATCHECK RMS Z scores				
Bond lengths (BNDCHK)	0.210 ± 0.003	0.268 ± 0.007	0.214 ± 0.003	0.282 ± 0.007
Bond angles (ANGCHK)	0.333 ± 0.005	0.426 ± 0.015	0.339 ± 0.005	0.457 ± 0.013
Omega angles (OMECHK)	0.06 ± 0.01	0.68 ± 0.05	0.08 ± 0.01	0.82 ± 0.10
Side-chain planarity (PLNCHK)	0.04 ± 0.01	0.28 ± 0.04	0.06 ± 0.01	0.39 ± 0.05
Improper dihedral angles (HNDCHK)	0.14 ± 0.01	0.33 ± 0.02	0.16 ± 0.01	0.37 ± 0.01
Inside/outside distribution (INOCHK)	0.95 ± 0.02	0.96 ± 0.02	0.96 ± 0.03	0.97 ± 0.02
Interatomic bumps (BMPCHK)	22.4 ± 4.6	18.3 ± 5.1	28.3 ± 6.1	22.3 ± 4.7
Unsatisfied H-bond donors (BH2CHK)	25.3 ± 3.2	14.6 ± 3.5	27.2 ± 3.7	16.6 ± 3.5
Unsatisfied H-bond acceptors (BA2CHK)	0.8 ± 0.7	0.8 ± 0.6	0.8 ± 0.8	0.6 ± 0.8
PROCHECK results				
Most favored	68.1 ± 2.5	74.1 ± 3.2	67.7 ± 2.8	75.0 ± 3.3
Allowed	26.6 ± 3.2	20.1 ± 2.5	26.4 ± 3.5	19.8 ± 3.1
Generously allowed	4.1 ± 1.8	3.5 ± 2.0	4.1 ± 1.6	2.7 ± 1.3
Disallowed	1.2 ± 0.9	2.3 ± 1.1	1.8 ± 1.3	2.5 ± 1.2
PROSA-II mean force energy	-1.96 ± 0.10	-2.03 ± 0.11	-1.96 ± 0.11	-2.00 ± 0.12
RMSD to 1RCB (backbone)	1.93 ± 0.18	1.88 ± 0.22	1.94 ± 0.17	1.86 ± 0.19
RMSD to 1RCB (all heavy atoms)	2.85 ± 0.20	2.84 ± 0.25	2.91 ± 0.21	2.85 ± 0.21
RMSD to 2INT (backbone)	1.90 ± 0.18	1.81 ± 0.22	1.91 ± 0.17	1.82 ± 0.19
RMSD to 2INT (all heavy atoms)	2.78 ± 0.20	2.75 ± 0.24	2.82 ± 0.19	2.76 ± 0.20
RMSD from mean structure (backbone, 2nd)	0.90 ± 0.18	1.00 ± 0.31	0.99 ± 0.16	1.05 ± 0.22
RMSD from mean structure (heavy atoms, all)	2.21 ± 0.31	2.30 ± 0.34	2.25 ± 0.32	2.31 ± 0.31
NOE violations > 0.1 Å	1.3 ± 0.9	7.3 ± 2.7	2.4 ± 1.2	7.8 ± 2.2

The deviation (measured in standard deviations) from the database-derived average value $\langle b_i^{db} \rangle$ for bond j of type i is

$$z_{ij} = \frac{b_{ij} - \langle b_i^{db} \rangle}{\sigma(b_i^{db})}, \tag{4}$$

in which $\sigma(b_i^{db})$ is the standard deviation of b_i in the database. The ideal value and its standard deviation stem from a set of high-resolution structures.[4] One can define an RMS Z score for a bond of type i in the given protein:

$$\mathrm{RMSZ}(b_i) = \sqrt{\frac{\sum_{j=1}^{J_i} z_{ij}^2}{J_i}}. \tag{5}$$

The RMS Z score for all bond types is then

$$\mathrm{RMSZ}(b) = \sqrt{\langle \overline{\mathrm{RMSZ}(b_i)^2} \rangle} = \sqrt{\frac{1}{I} \sum_{i=1}^{I} \frac{\sum_{j=1}^{J_i} z_{ij}^2}{J_i}}. \tag{6}$$

In the same manner, one can calculate RMS Z scores for bond angles, impropers, side-chain planarity, etc.

For comparison, in particular to Xia et al.,[3] we also quote the content of residues with ϕ–ψ values in the most favored, additional allowed, generously allowed, and disallowed regions of the Ramachandran plot[32] as determined with PROCHECK.[7] We used the program PROSA–II[33] to determine mean force potentials, averaged over all residues in a structure. We calculated the RMS differences to

the X-ray structures with CNS[34] (N, C_α, C atoms for the backbone or all heavy atoms). For fitting the structure, we chose the amino acid ranges 6–19, 41–59, 70–94, and 109–126 for IL-4, 1–70 for ubiquitin, and 1–46 for crambin. We used the two sets of residual dipolar couplings deposited in the 1D3Z Protein Data Bank (PDB) entry to calculate Q factors as described in Cornilescu et al.[14] with the program PALES.[35]

We validated all structures separately and averaged their validation results over the ensemble of 20 structures. The number quoted in Tables II–VI are simply the average values and standard deviations of the ensemble.

RESULTS AND DISCUSSION
Refinement in Water

The refinement in explicit water improves almost all validation criteria of IL-4, ubiquitin, and crambin (see data in Tables II–VI). Most importantly, the Ramachandran plot (as evidenced by the RAMCHK Z score and the Procheck results), the packing quality (QUACHK), the backbone conformation (BBCCHK), the number of unsatisfied H-bond donors and acceptors (BH2CHK and BA2CHK), and the number of interatomic bumps (BMPCHK) get better. For ubiquitin, the Q factors for the residual dipolar couplings also ameliorate, especially for the N-C and the H_N—N couplings. There are, however, three exceptions. First, the precision decreases slightly (larger RMSD to the

TABLE III. Quality Indices for Crambin[†]

Performed checks	Unrefined	Water refined	DMSO refined
WHATCHECK Z scores			
1st-generation packing quality (QUACHK)	-1.4 ± 0.2	-0.1 ± 0.1	-0.0 ± 0.1
2nd-generation packing quality (NQACHK)	-0.8 ± 0.3	0.3 ± 0.3	0.2 ± 0.2
Ramachandran plot appearance (RAMCHK)	-3.5 ± 0.4	-0.5 ± 0.5	-0.7 ± 0.5
χ_1–χ_2 rotamer normality (G12CHK)	-2.5 ± 0.5	-3.1 ± 0.4	-3.3 ± 0.5
Backbone conformation (BBCCHK)	0.1 ± 0.3	0.6 ± 0.3	0.6 ± 0.3
WHATCHECK RMS Z scores			
Bond lengths (BNDCHK)	0.213 ± 0.001	0.249 ± 0.005	0.256 ± 0.007
Bond angles (ANGCHK)	0.236 ± 0.002	0.310 ± 0.009	0.316 ± 0.006
Omega angles (OMECHK)	0.03 ± 0.01	0.51 ± 0.04	0.50 ± 0.04
Side-chain planarity (PLNCHK)	0.03 ± 0.01	0.33 ± 0.06	0.31 ± 0.09
Improper dihedral angles (HNDCHK)	0.18 ± 0.01	0.28 ± 0.02	0.28 ± 0.01
Inside/outside distribution (INOCHK)	1.01 ± 0.01	1.01 ± 0.01	1.01 ± 0.01
Inter/atomic bumps (BMPCHK)	4.6 ± 1.7	3.9 ± 1.5	4.4 ± 1.7
Unsatisfied H-bond donors (BH2CHK)	3.6 ± 1.1	2.1 ± 1.2	1.9 ± 0.7
Unsatisfied H-bond acceptors (BA2CHK)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
PROCHECK results			
Most favored	88.3 ± 3.1	90.9 ± 2.9	90.7 ± 2.0
Allowed	11.7 ± 3.1	9.1 ± 2.9	9.3 ± 2.0
Generously allowed	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Disallowed	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
PROSA-II mean force energy	-0.68 ± 0.05	-0.59 ± 0.07	-0.58 ± 0.06
RMSD to 1EJG (backbone)	0.49 ± 0.06	0.47 ± 0.07	0.47 ± 0.08
RMSD to 1EJG (all heavy atoms)	0.70 ± 0.08	0.75 ± 0.10	0.76 ± 0.09
RMSD from mean structure (backbone, 2nd)	0.23 ± 0.07	0.30 ± 0.06	0.27 ± 0.06
RMSD from mean structure (heavy atoms, all)	0.34 ± 0.07	0.38 ± 0.05	0.35 ± 0.06
NOE violations > 0.1 Å	0.3 ± 0.5	1.1 ± 0.9	1.2 ± 1.0

[†]Calculations are without force field dihedrals.

average structure). This is often regarded as disappointing in the NMR literature, although it can be argued that it is a positive feature as the sampling of conformational space improves. Second, the χ_1–χ_2 rotamer normality Z score gets slightly worse for IL-4 and crambin. Third, the number of NOE violations increases slightly. Note that we quote small violations of 0.1 Å; there were no violations larger than 0.3 Å. This is a consequence of competition with the additional energy terms in the force fields. We cannot expect a reduction of the violations as seen in Spronk et al.[2] because the covalent parameters and relative weights between experimental data and force field are the same during calculation and refinement.

Some commonly used validation criteria are too insensitive to distinguish between refined and unrefined structures. First, the inside−outside distribution Z score from WHATCHECK depends strongly on the overall fold of the protein. Thus, this value is usually close to one, regardless of the structure refinement used. Second, PROSA-II energies mostly depend on the overall fold and are thus insensitive to the subtle differences between unrefined and water-refined structures.

Most validation criteria are slightly better than for the CHARMM22-refined structures and significantly better than those of ARIA 1.0-refined structures (compare with data in Linge and Nilges[1] and Spronk et al.[2]).

Refinement in DMSO

The refinement in DMSO has a similar effect on the crambin structures as the refinement in water (see Tables III and IV). Obviously, it is of special interest for the increasing number of NMR experiments performed on DMSO samples. Other solvents often used in NMR such as TFE will be included in the future.

Dihedral Angle Force Field Term

The single most important effect of the dihedral angle force field function is the improvement of the χ_1–χ_2 rotamer normality Z score. Where existant, the other dihedral angle distributions (χ_2, χ_3, etc.) also improve dramatically, whereas other validation results change only slightly. Thus, we employ this dihedral angle energy term in version 5.3 of the PARALLHDG force field by default. We note that full molecular dynamics force fields standardly include such a term.

The strong influence of the force field dihedral terms on sidechain behavior reflects the fact that NMR data usually restrain sidechains only weakly, in particular at the protein surface.

Z Scores Revisited

It is common practice to regard Z scores close to 0 as good and Z scores larger than 4 or smaller than -4 as outliers. Ramachandran Z scores of, for examples -0.2 and 0.8 mean that both structures exhibit a Ramachandran plot that is within one standard deviation of the average in the database. The value of 0.8 corresponds to a Ramachandran map that is more regular than the one belonging to the value of -0.2 (more angles in the allowed regions of the Ramachandran map). The value of -0.2

TABLE IV. Quality Indices for Crambin[†]

Performed Checks	Unrefined	Water refined	DMSO refined
WHATCHECK Z-scores:			
1st generation packing quality (QUACHK)	-1.5 ± 0.2	0.0 ± 0.1	-0.1 ± 0.2
2nd generation packing quality (NQACHK)	-0.6 ± 0.3	0.7 ± 0.4	0.4 ± 0.3
Ramachandran plot appearance (RAMCHK)	-4.0 ± 0.4	-1.0 ± 0.6	-1.2 ± 0.3
$\chi_1-\chi_2$ rotamer normality (C12CHK)	0.8 ± 0.4	0.5 ± 0.5	0.4 ± 0.5
Backbone conformation (BBCCHK)	-0.1 ± 0.3	0.4 ± 0.3	0.4 ± 0.2
WHATCHECK RMS Z-scores:			
Bond lengths (BNDCHK)	0.220 ± 0.002	0.250 ± 0.004	0.255 ± 0.009
Bond angles (ANGCHK)	0.241 ± 0.002	0.315 ± 0.011	0.322 ± 0.012
Omega angles (OMECHK)	0.04 ± 0.01	0.54 ± 0.03	0.54 ± 0.05
Side chain planarity (PLNCHK)	0.07 ± 0.03	0.33 ± 0.10	0.34 ± 0.11
Improper dihedral angles (HNDCHK)	0.19 ± 0.01	0.28 ± 0.01	0.29 ± 0.02
Inside/outside distribution (INOCHK)	1.02 ± 0.01	1.01 ± 0.01	1.01 ± 0.01
Inter-atomic bumps (BMPCHK)	5.4 ± 1.7	3.1 ± 1.2	4.3 ± 0.9
Unsatisfied H-bond donors (BH2CHK)	4.1 ± 0.8	2.3 ± 0.9	2.2 ± 0.7
Unsatisfied H-bond acceptors (BA2CHK)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
PROCHECK results:			
Most favoured	87.7 ± 2.6	92.4 ± 3.5	89.4 ± 2.3
Allowed	12.3 ± 2.6	7.6 ± 3.5	10.6 ± 2.3
Generously allowed	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Disallowed	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
PROSA-II mean force energy	-0.65 ± 0.04	-0.59 ± 0.05	-0.60 ± 0.05
RMSD to 1EJG (backbone)	0.45 ± 0.07	0.44 ± 0.07	0.45 ± 0.09
RMSD to 1EJG (all heavy atoms)	0.71 ± 0.09	0.76 ± 0.09	0.76 ± 0.08
RMSD from mean structure (backbone, 2nd)	0.26 ± 0.12	0.32 ± 0.08	0.29 ± 0.05
RMSD from mean structure (heavy atoms, all)	0.36 ± 0.10	0.43 ± 0.06	0.41 ± 0.04
NOE violations > 0.1 Å	0.0 ± 0.0	0.4 ± 0.7	0.8 ± 1.0

[†]Calculations are with force field dihedrals.

would be closer to the database-derived average value. It could thus be argued that the value of 0.8 is less likely and thus "worse." However, we considered a structure as having better validation results if the value gets more positive. This is the more intuitive interpretation also used in Spronk et al.[2] It is consistent with the notion that more φ–ψ pairs in the most favored regions correspond to a better structure. On the other hand, it is important to remember that variations in the database are not only caused by differences in quality but also by fundamental differences in structures. Thus, one has to be careful not to overinterpret these Z scores.

Usefulness of RMS Z Scores

Ideal RMS Z scores are close to 1. RMS Z scores smaller than 1.0 point out a tighter distribution and those larger than 1.0 a broader distribution of values than in the WHATCHECK database. All RMS Z scores for bond lengths, bond angles, omega angles, impropers, and side-chain planarity increase (closer to one) during the refinement in explicit solvent. This is a consequence of the larger nonbonded forces acting during the refinement, and, for the angle ω, of the reduced energy constant. However, the RMS Z scores are still significantly closer to zero than in the CHARMM22 refinement.[2] Partly, this is due to systematic differences between the CHARMM22 parameters and the WHATCHECK database; in contrast, the PARALL-HDG parameters are consistent with the WHATCHECK database.

The influence of systematic differences between the refinement force field and database parameters is most obvious if one uses constraints to maintain the values (e.g., the SHAKE[36] constraint used in Spronk et al.[2] for the bond lengths). To clarify this point, we compare the distribution and RMS Z scores obtained for the C_α—C and N—C bond lengths in the calculated structures before and after refinement in water with those after a short minimization using CHARMM22 parameters (Fig. 4). The RMS Z scores for the distributions in Figure 4 are 0.181 and 0.266 (see Table V) before and after water refinement, respectively, while the δ–function-like distribution for CHARMM22 corresponds to an RMS Z score of 1.02. In this example, the δ–function-like distribution gives rise to the "best" RMS Z score because its center is accidentally around one standard deviation away from the mean value. If one looks only at RMS Z scores, one could mistakenly argue that this distribution has the highest quality.

RMS Z scores smaller than 1.0 (in conjunction with a Z score near zero) mean that the distribution is tighter than expected from high-resolution X-ray crystal structures. Considering the completeness and quality of experimental NMR data, we think it is better to restrain the geometry tightly because it is clearly impossible to determine any deviations from ideal covalent parameters. Thus, we see no advantage in lowering the energy constants for covalent energy terms, apart from loosening the peptide plane angle ω. The observed deviation of ω in high-resolution X-ray structures is substantial—up to 6° from the mean

TABLE V. Quality Indices for Ubiquitin

Performed checks	Without force field dihedrals		With force field dihedrals	
	Unrefined	Water refined	Unrefined	Water refined
WHATCHECK Z scores				
1st-generation packing quality (QUACHK)	-0.2 ± 0.2	0.5 ± 0.1	-0.0 ± 0.1	0.5 ± 0.2
2nd-generation packing quality (NQACHK)	-0.9 ± 0.2	-0.4 ± 0.3	-0.7 ± 0.3	-0.3 ± 0.3
Ramachandran plot appearance (RAMCHK)	-0.1 ± 0.4	0.6 ± 0.3	-0.3 ± 0.3	0.4 ± 0.3
$\chi_1-\chi_2$ rotamer normality (C12CHK)	-3.5 ± 0.3	-3.5 ± 0.4	0.1 ± 0.5	0.1 ± 0.5
Backbone conformation (BBCCHK)	1.7 ± 0.2	2.4 ± 0.4	1.8 ± 0.2	2.3 ± 0.3
WHATCHECK RMS Z scores				
Bond lengths (BNDCHK)	0.176 ± 0.001	0.258 ± 0.006	0.181 ± 0.002	0.266 ± 0.006
Bond angles (ANGCHK)	0.312 ± 0.002	0.380 ± 0.007	0.324 ± 0.003	0.408 ± 0.010
Omega angles (OMECHK)	0.11 ± 0.01	0.67 ± 0.04	0.13 ± 0.01	0.74 ± 0.03
Side-chain planarity (PLNCHK)	0.04 ± 0.01	0.32 ± 0.05	0.07 ± 0.02	0.44 ± 0.08
Improper dihedral angles (HNDCHK)	0.17 ± 0.01	0.28 ± 0.01	0.18 ± 0.01	0.32 ± 0.02
Inside/outside distribution (INOCHK)	0.99 ± 0.02	1.01 ± 0.02	1.00 ± 0.00	1.00 ± 0.02
Interatomic bumps (BMPCHK)	6.6 ± 2.7	1.9 ± 2.1	8.4 ± 2.6	3.7 ± 1.9
Unsatisfied H-bond donors (BH2CHK)	4.6 ± 1.4	3.0 ± 1.4	5.1 ± 1.4	2.9 ± 1.2
Unsatisfied H-bond acceptors (BA2CHK)	1.1 ± 0.4	1.0 ± 0.0	1.5 ± 0.6	1.1 ± 0.2
PROCHECK results				
Most favored	91.3 ± 1.8	91.2 ± 1.4	91.4 ± 1.1	90.7 ± 2.0
Allowed	7.2 ± 2.2	8.4 ± 1.6	7.4 ± 1.7	9.2 ± 1.9
Generously allowed	1.4 ± 1.0	0.2 ± 0.6	1.0 ± 0.9	0.0 ± 0.0
Disallowed	0.1 ± 0.3	0.2 ± 0.6	0.2 ± 0.6	0.1 ± 0.3
PROSA-II mean force energy	-1.74 ± 0.05	-1.65 ± 0.08	-1.73 ± 0.08	-1.69 ± 0.11
RMSD to 1UBQ (backbone)	0.51 ± 0.05	0.56 ± 0.07	0.55 ± 0.05	0.54 ± 0.04
RMSD to 1UBQ (all heavy atoms)	1.21 ± 0.06	1.24 ± 0.06	1.23 ± 0.07	1.20 ± 0.07
RMSD from mean structure (backbone, 2nd)	0.24 ± 0.04	0.35 ± 0.06	0.26 ± 0.05	0.31 ± 0.06
RMSD from mean structure (heavy atoms, all)	1.02 ± 0.15	1.07 ± 0.13	1.01 ± 0.13	1.05 ± 0.12
NOE violations > 0.1 Å	1.1 ± 0.8	3.6 ± 1.7	3.1 ± 1.1	7.7 ± 1.7
Q factors				
$C_\alpha - C_\beta$ (1)	0.359 ± 0.018	0.369 ± 0.034	0.380 ± 0.018	0.376 ± 0.023
$C_\alpha - C$ (1)	0.241 ± 0.015	0.216 ± 0.028	0.245 ± 0.016	0.227 ± 0.021
$C_\alpha - C$ (2)	0.210 ± 0.012	0.208 ± 0.026	0.209 ± 0.015	0.220 ± 0.022
$N - C$ (1)	0.245 ± 0.017	0.193 ± 0.029	0.219 ± 0.017	0.192 ± 0.018
$N - C$ (2)	0.308 ± 0.019	0.198 ± 0.036	0.270 ± 0.015	0.194 ± 0.014
$H_\alpha - C_\alpha$ (1)	0.241 ± 0.022	0.254 ± 0.023	0.246 ± 0.023	0.255 ± 0.028
$H_\alpha - C_\alpha$ (2)	0.269 ± 0.028	0.304 ± 0.022	0.270 ± 0.017	0.306 ± 0.029
$H_N - C$ (1)	0.304 ± 0.019	0.277 ± 0.031	0.298 ± 0.013	0.271 ± 0.025
$H_N - C$ (2)	0.371 ± 0.016	0.327 ± 0.032	0.359 ± 0.015	0.324 ± 0.018
$H_N - N$ (1)	0.400 ± 0.019	0.236 ± 0.064	0.368 ± 0.021	0.216 ± 0.031
$H_N - N$ (2)	0.346 ± 0.025	0.246 ± 0.029	0.317 ± 0.022	0.266 ± 0.024

value.[37] Small variations in a backbone torsion angle may influence the other backbone angles and thus the structure accuracy. Further, it could be argued that the ω angle depends on many more data points than a single covalent bond and so can be influenced by the NMR data. However, our test calculation with ubiquitin (see Table VI) shows that the influence of this value on other structural features is minimal; Q factors even improve slightly with tightened geometry for the ω angle. We have made 100 kcal mol^{-1} rad^{-2} the default value for the ω angle force constant as a compromise between the expected flexibility and the improved Q factors.

Comparison with Other Refinement Schemes

Our refinement takes approximately 15 min for IL-4, 9 min for ubiquitin, and 5 min for crambin on a single-processor AthlonXP 1.5 GHz Linux machine. It is thus much more CPU efficient than other refinement schemes, for example, an explicit water calculation in a solvent box or a generalized Born calculation as in Xia et al.[3] (200 h or 15–90 h, respectively, for glutaredoxin-2 on an SGI Origin 2400 and 3800 server using R12000 CPUs).

Kuszewski et al. proposed a different refinement idea, termed "database refinement".[38,39] Their refinement does not aim at improving the physical properties of the force field but directly seeks to make the NMR structure more similar to the existing protein structures by using potential energy terms derived from the statistics of high-resolution X-ray crystal structures. This is in marked difference to our approach. The force field parameters used in our refinement are not derived from protein structures but from small molecules; Engh and Huber obtained their parameters[4] by analyzing statistics over amino acid analogs in the Cambridge database of small molecular compounds,[21] the dihedral angle parameters reflect local chemistry, and Jorgensen and Tirado-Rives optimized the nonbonded OPLS parameters by comparing calculated

**TABLE VI. Quality Indices for Water-Refined Ubiquitin (From Three Calculations
with ω Energy Constants of 20, 100, and 500 kcal^{-1} mol^{-1} rad^{-2})**

Performed checks	20	100	500
WHATCHECK Z scores			
1st-generation packing quality (QUACHK)	0.4 ± 0.2	0.5 ± 0.2	0.6 ± 0.2
2nd-generation packing quality (NQACHK)	-0.4 ± 0.3	-0.3 ± 0.3	-0.3 ± 0.3
Ramachandran plot appearance (RAMCHK)	0.2 ± 0.2	0.4 ± 0.3	0.6 ± 0.4
$\chi_1-\chi_2$ rotamer normality (C12CHK)	0.2 ± 0.4	0.1 ± 0.5	-0.1 ± 0.4
Backbone conformation (BBCCHK)	2.1 ± 0.4	2.3 ± 0.3	2.5 ± 0.3
WHATCHECK RMS Z-scores:			
Bond lengths (BNDCHK)	0.264 ± 0.006	0.266 ± 0.006	0.272 ± 0.006
Bond angles (ANGCHK)	0.387 ± 0.010	0.408 ± 0.010	0.445 ± 0.012
Omega angle restraints (OMECHK)	1.45 ± 0.11	0.74 ± 0.03	0.25 ± 0.01
Side-chain planarity (PLNCHK)	0.44 ± 0.07	0.44 ± 0.08	0.45 ± 0.08
Improper dihedral distribution (HNDCHK)	0.31 ± 0.02	0.32 ± 0.02	0.34 ± 0.01
Inside/outside distribution (INOCHK)	1.00 ± 0.02	1.00 ± 0.02	1.00 ± 0.02
Interatomic bumps (BMPCHK)	5.7 ± 2.5	3.7 ± 1.9	3.0 ± 1.8
Unsatisfied hydrogen donors (BH2CHK)	2.5 ± 1.1	2.9 ± 1.2	3.1 ± 1.6
Unsatisfied hydrogen acceptors (BA2CHK)	1.1 ± 0.2	1.1 ± 0.2	1.1 ± 0.2
PROCHECK results			
Most favored	90.2 ± 2.2	90.7 ± 2.0	91.2 ± 1.7
Allowed	9.7 ± 2.0	9.2 ± 1.9	8.5 ± 1.7
Generously allowed	0.1 ± 0.3	0 ± 0.0	0.2 ± 0.6
Disallowed	0.0 ± 0.0	0.1 ± 0.3	0.1 ± 0.3
PROSA-II mean force energy	-1.66 ± 0.09	-1.69 ± 0.11	-1.67 ± 0.09
RMSD to 1UBQ (backbone)	0.56 ± 0.04	0.54 ± 0.04	0.53 ± 0.04
RMSD to 1UBQ (all heavy atoms)	1.22 ± 0.06	1.20 ± 0.07	1.19 ± 0.07
RMSD from mean structure (backbone, 2nd)	0.34 ± 0.06	0.31 ± 0.06	0.30 ± 0.06
RMSD from mean structure (heavy atoms, all)	1.06 ± 0.11	1.05 ± 0.12	1.03 ± 0.12
NOE violations > 0.1 Å	7.5 ± 1.8	7.7 ± 1.7	9.7 ± 2.4
Q-factors			
$C_\alpha - C_\beta$ (1)	0.388 ± 0.031	0.376 ± 0.023	0.370 ± 0.022
$C_\alpha - C$ (1)	0.247 ± 0.022	0.227 ± 0.021	0.219 ± 0.017
$C_\alpha - C$ (2)	0.252 ± 0.015	0.220 ± 0.022	0.203 ± 0.019
$N - C$ (1)	0.207 ± 0.019	0.192 ± 0.018	0.191 ± 0.017
$N - C$ (2)	0.201 ± 0.017	0.194 ± 0.014	0.197 ± 0.016
$H_\alpha - C_\alpha$ (1)	0.272 ± 0.031	0.255 ± 0.028	0.243 ± 0.021
$H_\alpha - C_\alpha$ (2)	0.331 ± 0.035	0.306 ± 0.029	0.286 ± 0.023
$H_N - C$ (1)	0.281 ± 0.023	0.271 ± 0.025	0.267 ± 0.020
$H_N - C$ (2)	0.326 ± 0.020	0.324 ± 0.018	0.329 ± 0.015
$H_N - N$ (1)	0.222 ± 0.028	0.216 ± 0.031	0.216 ± 0.039
$H_N - N$ (2)	0.274 ± 0.028	0.266 ± 0.024	0.262 ± 0.027

properties of simulations of small organic molecules with experimental data.[5]

CONCLUSION

We described a short and efficient refinement protocol for protein structures in explicit solvent (water and DMSO).

The use of few waters in a thin layer around the protein rather than a large box with periodic boundary conditions is obviously sufficient to account for the most important effects of water, providing polar and nonpolar interactions on the surface of the protein to prevent unrealistic packing of the side-chains (cf. Beglov and Roux[40]). Hence, the refinement in explicit solvent is CPU efficient and sufficient to improve structures in a similar way as more time-consuming schemes.[2,3]

Whereas we have not shown results for protein structures from homology modeling or threading approaches, it is obvious that refinement in explicit solvent is also applicable for these structures.

The force field used in the refinement is consistent with that used for structure calculation and validation. Thus, there are no systematic differences between the force field and the database that could influence the validation results. In contrast to other proposed refinement schemes, we do not use a standard molecular dynamics force field but parameters specially developed for structure refinement. This is an important distinction because the relative weighting between different terms in force fields such as CHARMM22 or AMBER94[41] are different from those found, for example, in the Engh and Huber parameters.[4] The former are parameterized to calculate realistic energetic and dynamic properties of a structure and the latter to produce idealized average local geometry that is consistent and in balance with the strong restraints derived from experimental data.

Our results for IL-4, crambin, and ubiquitin demonstrate that a force field dihedral angle energy term improves the $\chi_1-\chi_2$ distribution without interfering with
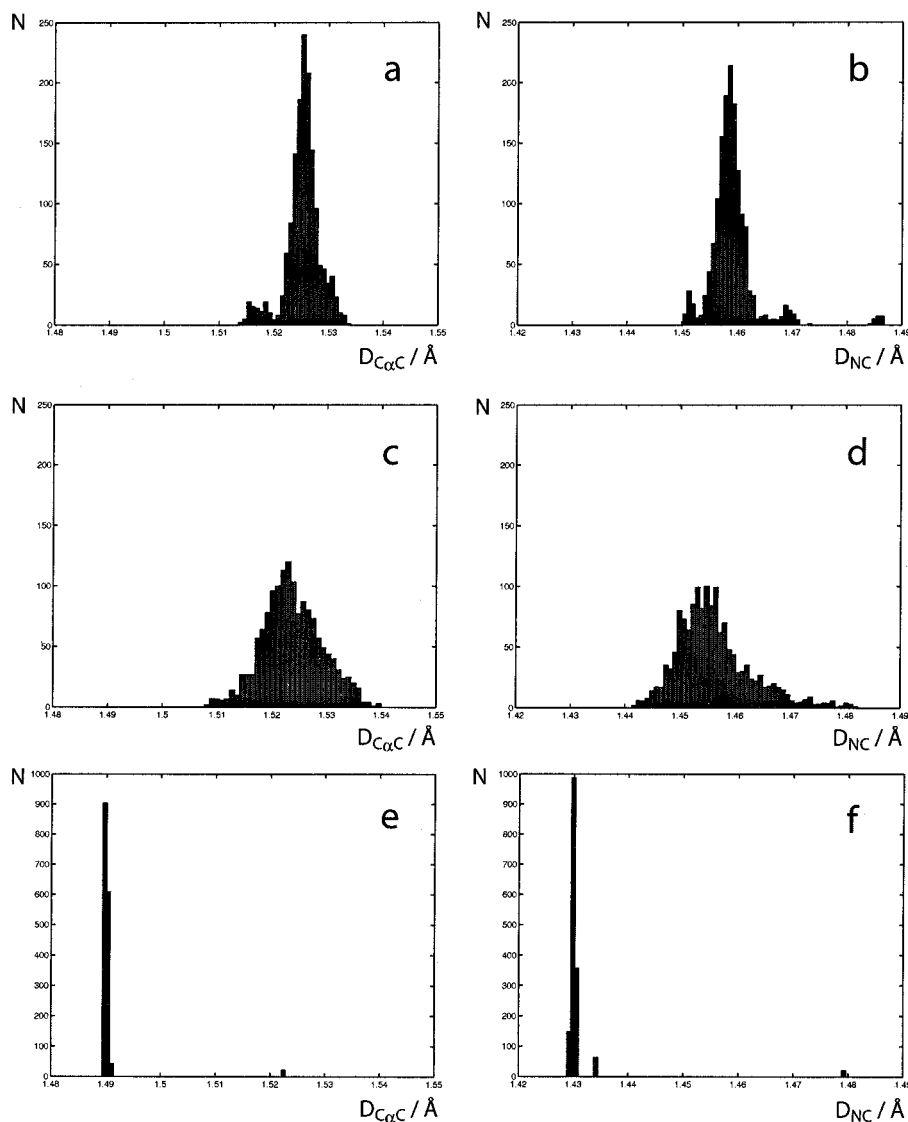
Fig. 4. Distribution of C_α—C and N—C bond lengths of ubiquitin. (**a**) C_α—C unrefined. (**b**) N—C unrefined (the small peak around 1.486 Å is due to the different parameterization at the N-terminus). (**c**) C_α—C after water refinement. (**d**) N—C after water refinement. (**e**) C_α—C after short minimization with the CHARMM22 force field using SHAKE (the small peak around 1.522 Å is due to the different parameterization at the C-terminus). (**f**) N—C after short minimization with the CHARMM22 for field using SHAKE (the small peak around 1.479 Å is due to the different parameterization at the N-terminus).

other quality indices like the Ramachandran plot or the packing quality. Thus, we will use the proposed dihedral angle energy function by default. RMS Z scores for covalent parameters such as bond lengths and bond angles are not a useful criterion of the overall structural quality because their influence on the global structure is small and direct experimental evidence for most covalent parameters is sparse. Therefore, it is beneficial to restrain the covalent parameters for which there are no experimental data. Even ω can be restrained tightly; in the case of ubiquitin, it actually improves the Q factors. As shown, variation in ω on other validation indices is minimal.

In general, it is encouraging that various water refinement schemes[2,3] give similar results, although these methods used different force fields (PARALLHDG, OPLS, CHARMM22, AMBER), programs (CNS, X-PLOR, AMBER, OPAL[42]), and refinement concepts (thin layer, solvent box, generalized Born).

Our protocol for refinement in explicit solvent is part of ARIA 1.2,[11] available from: http://www.pasteur.fr/recherche/unites/Binfs/.

REFERENCES

1. Linge JP, Nilges M. Influence of non-bonded parameters on the quality of NMR structures: a new force-field for NMR structure calculation. J Biomol NMR 1999;13:51–59.
2. 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.
3. Xia B, Tsui V, Case DA, Dyson HJ, Wright PE. Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. J Biomol NMR 2002;22:317–331.
4. Engh RA, Huber R. Accurate bond and angle parameters for X-ray structure refinement. Acta Crystallogr A 1991;47:392–400.
5. Jorgensen WI, Tirado-Rives J. The OPLS potential function for proteins. Energy minimization for crystals of cyclic peptides and crambin. J Am Chem Soc 1988;110:1657–1666.
6. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature 1996;381:272.
7. 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.
8. Oldfield TJ. SQUID: a program for the analysis and display of data from crystallography and molecular dynamics. J Mol Graph 1992;10:247–252.
9. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiber WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watan-abe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
10. Powers R, Garrett DS, March CJ, Frieden EA, Gronenborn AM, Clore GM. Three-dimensional solution structure of human interleukin-4 determined by multidimensional heteronuclear magnetic resonance spectroscopy. Science 1992;19:1673–1677.
11. Linge JP, Habeck H, Rieping W, Nilges M. ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics 2003. In Press.
12. Walter MR, Cook WJ, Zhao BG, Cameron RP, Ealick SE, Walter RL, Reichert P, Nagabhushan TL, Trotta PP, Bugg CE. Crystal structure of recombinant human interleukin-4. J Biol Chem 1992;267:20371–20376.
13. Wlodawer A, Pavlovsky A, Gustchina A. Crystal structure of human recombinant human interleukin-4 at 2.25 Å resolution. FEBS Lett 1992;309:59–64.
14. 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.
15. Vijay-Kumar S, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8 Å resolution. J Mol Biol 1987;194:531–544.
16. Jelsch C, Teeter MM, Lamzin V, Pichon-Pesme V, Blessing RH, Lecomte C. Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. Proc Natl Acad Sci USA 2000;97:3171–3176.
17. Hendrickson WA. Stereochemically restrained refinement of macromolecular structures. Meth Enzymol 1985;115:252–270.
18. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
19. Liu H, Müller-Plathe F, van Gunsteren WF. A molecular dynamics simulation study with a combined quantum mechanical and molecular mechanical potential energy function: solvent effects on the conformational equilibrium of dimethoxyethane. J Chem Phys 1995;102:1722–1730.
20. Doreleijers JF, Rullmann JA, Kaptein R. Quality assessment of NMR structures: a statistical survey. J Mol Biol 1998;281:149–164.
21. Allen FHS, Brice MD, Cartwright BA, Doubleday A, Higgs H, Hummelink T, Hummelink-Peters BG, Kennard O, Mother-well WDS, Rodgers JR, Watson DG. The Cambridge Crystallographic Data Centre: computer-based search, retrieval analysis and display of information. Acta Crystallogr B 1979;35:2331–2339.
22. Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM. Analysis of main chain torsion angles in proteins: prediction of coupling constants for native and random coil conformations. J Mol Biol 1996;255:494–506.
23. MacArthur MW, Thornton JM. Conformational analysis of protein structures derived from NMR data. Proteins 1993;17:232–251.
24. Rice LM, Brünger AT. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. Proteins 1994;19:277–290.
25. Stein EG, Rice LM, Brünger AT. Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. J Magn Reson 1997;124:154–164.
26. Nilges M, O'Donoghue SI. Ambiguous NOEs and automated NOESY assignment. Prog NMR Spectrosc 1998;32:107–139.
27. 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.
28. Folmer RH, Hilbers CW, Konings RN, Nilges M. Floating stereospecific assignment revisited: application to an 18 kDa protein and comparison with J-coupling data. J Biomol NMR 1997;9:245–258.
29. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
30. Bruccoleri RE, Karplus M. Spatially constrained minimization of macromolecules. J Comput Chem 1986;7:165–175.
31. Hooft RW, Sander C, Vriend G. Objectively judging the quality of a protein structure from a Ramachandran plot. Comput Appl Biosci 1997;13:425–430.
32. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol 1963;7:95–99.
33. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins 1993;17:355–362.
34. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J.-S., Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. Acta Crystallogr D 1998;54:905–921.
35. Zweckstetter M, Bax AJ. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. J Am Chem Soc 2000;122:3791–3792.
36. Ryckaert JP, Ciocotti G, Berendsen HJC. Numerical integration of Cartesian equations of motion of a system with constraints—Molecular dynamics of N-alkanes. J Comput Phys 1977;23:327–341.
37. MacArthur MW, Thornton JM. Deviations from planarity of the peptide bond in peptides and proteins. J Mol Biol 1996;264:1180–1195.
38. Kuszewski J, Gronenborn AM, Clore GM. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. Protein Sci 1996;5:1067–1080.
39. Kuszewski J, Gronenborn AM, Clore GM. Improvements and extentions 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.
40. Beglov D, Roux B. Dominant solvation effects from the primary shell of hydration: approximation for molecular dynamics simulations. Biopolymers 1995;35:171–178.
41. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer D, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
42. Luginbühl P, Güntert P, Billeter M, Wüthrich K. The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. J Biomol NMR 1996;8:136–146.