Journal of Biomolecular NMR, **13:** 51–59, 1999.
KLUWER/ESCOM

Influence of non-bonded parameters on the quality of NMR structures: A new force field for NMR structure calculation

J.P. Linge & M. Nilges*
European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany

Received 15 June 1998; Accepted 17 September 1998

Key words: accuracy, force fields, Interleukin 4, precision, solution structure, structure validation

Abstract

The effects of different non-bonded parameters of force fields for NMR structure calculation on the quality of the resulting NMR solution structures were investigated using Interleukin 4 as a model system. NMR structure ensembles were calculated with an ab initio protocol using torsion angle dynamics. The calculations were repeated with five different non-bonded energy functions and parameters. The resulting ensembles were compared with the available X-ray structures, and their quality was assessed with common structure validation programs. In addition, the impact of torsion angle restraints and dihedral energy terms for the sidechains and the backbone was studied. The further improvement of the quality by refinement in explicit solvent was demonstrated. The optimal parameters, including those necessary for water refinement, are available in the new version of the PARALLHDG force field.

Abbreviations: CAD, contact area difference; IL4, Interleukin 4; MD, molecular dynamics; NOESY, nuclear Overhauser enhanced spectroscopy; TAD, torsion angle dynamics; vdW, van der Waals.

Introduction

Even the most complete NMR data set, which may include interproton distance restraints, torsion angle or coupling constant restraints, and orientational restraints derived from residual dipolar couplings, is by itself insufficient to calculate three-dimensional structures. The experimental information always has to be complemented by the empirical information contained in force fields, which are derived from small molecule and protein structure databases.

In NMR structure calculation by simulated annealing, an energy target function is used that describes experimental information and a priori knowledge about the system:

$$E = E_{\text{chem}} + w_{\text{NMR}} E_{\text{NMR}} \quad (1)$$

where E_{chem} is an energy function describing the covalent and non-bonded interactions, E_{NMR} is an energy function which represents the experimental data, and w_{NMR} is the weight of the NMR energy function. Because of the sparseness of the NMR data, force fields have a large impact on the quality and possibly also on the accuracy of NMR structures. Therefore one important goal is to improve the force fields.

The CSDX parameters (Engh and Huber, 1991), derived from the Cambridge Structural Database (Allen et al., 1979), are commonly used for X-ray crystal structure refinement, and are the reference parameters for structure validation programs (e.g., WHATIF (Vriend, 1990) and PROCHECK (Laskowski et al., 1996)). We have therefore decided to base the new PARALLHDG parameters on the CSDX force field. Recently (Adams et al., 1997), the non-bonded interactions in the CSDX force field, which used to be based on the CHARMM PARAM19 force field (Brooks et al., 1983), were replaced with those from the PROLSQ refinement program (Hendrickson, 1985). As Brünger et al. (1997) remarked, this provides uniformity among different crystallo-

*To whom correspondence should be addressed.

graphic refinement programs. In order to provide uniformity between X-ray and NMR refinement parameters, we have decided to base the non-bonded parameters in PARALLHDG also on the PROLSQ parameters. In this paper, we compare these parameters to several other parameter sets, including those used in previous versions of PARALLHDG. In addition, we investigate the effect of using standard dihedral potentials around rotatable bonds, which had been deliberately removed from previous versions of PARALLHDG, for consistency with distance geometry programs, and of a final short refinement in explicit solvent, using OPLS (Jorgensen and Tirado-Rives, 1988) non-bonded parameters.

We used Interleukin 4 as a model system for our study because we expected that non-bonded energy parameters are particularly important for the correct packing in helical structures. Interleukin 4 folds in a left-handed four helix bundle with up–up–down–down connectivity. Its sequence comprises 129 residues and contains three disulfide bridges (3–127, 24–65, 46–99). In the PDB there exist three X-ray structures (a 2.25 Å X-ray structure 1RCB (Wlodawer et al., 1992), a 2.35 Å X-ray structure 2INT (Walter et al., 1992), and a 2.6 Å X-ray structure 1HIK (Müller et al., 1995)), and a number of NMR structures from three different laboratories (PDB entries 1BBN, 1BCN, 1CYL, 1ITI, 1ITL, 2CYK (Powers et al., 1992a,b; Redfield et al., 1992; Smith et al., 1992; Müller et al., 1994)). A comparison of the structures 1RCB, 2INT, 1ITL, 1BBN, and 1ITI has been published (Smith et al., 1994). We decided to choose the data from a structure of intermediate quality (1BBN) rather than a high resolution structure (1ITI) because the effects of the force field should be more prominent.

Materials and methods

Structure calculation

Experimental data

We took the NOE peak list and torsion angle restraint list from the PDB entry 1BBN (Powers et al., 1992b) as a starting point for our structure calculations. The data contains 918 unambiguously assigned NOEs (329 long-range, 305 medium-range, and 284 short-range) and 174 torsion angle restraints (101 ϕ- and 73 ψ-restraints). Compared to the wild type, the amino acid sequence has four additional residues (Glu Ala Glu Ala) at the N-terminus and two Asn to Asp mutations for residues 38 and 105. The whole sequence comprises 133 residues.

Covalent parameters

Covalent interactions were always calculated with the latest implementation of the PARALLHDG.PRO force field (Version 5.0), which uses standard energy function terms of the form

$$E_{\text{chem}} = \sum_{\text{bonds}} k_b(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi cos(n\phi + d) + \sum_{\text{chiral, planar}} k_\omega(\omega - \omega_0)^2 \quad (2)$$

for maintaining correct bond lengths, angles, dihedral torsion angles, planarity and chirality. The dihedral angle term is usually not used.

The covalent parameters were derived from the CSDX (Engh and Huber, 1991) and previous PARALLHDG (Nilges et al., 1988; Kuszewski et al., 1992) force fields in an automated way by an iterative procedure. First, a peptide comprising all possible amino acids and all necessary combinations of amino acids and N- and C-terminal patches was extensively minimized in the CSDX force field. Hydrogens were then built in X-PLOR with the previous version of PARALLHDG. All parameters, including those involving hydrogens, were then derived from this structure with the PARAmeter LEARn command in X-PLOR. The inconsistent notation of the Pro H_δ protons was corrected by hand.

In contrast to the X-ray parameters, which use energy constants derived from the variance of the bond lengths and bond angles, PARALLHDG uses uniform energy constants: 500 kcal mol^{-1} rad^{-2} for the bond angles and improper dihedral angles, and 1000 kcal mol^{-1} Å^{-2} for the bonds. Dihedral angle energy terms were only used for rotatable bonds; all fixed angles were expressed as improper dihedral angles. For all side-chain dihedral angles, the three standard rotamers were defined; for the backbone three positions for ϕ (-60, 60, and 180) and four ψ (-135, -45, 45, and 135) were used. The energy constants were set to 5 kcal mol^{-1} rad^{-2} for the sidechains and 2 kcal mol^{-1} rad^{-2} for the backbone.

Non-bonded parameters

For the non-bonded interactions we took the parameters and energy functions from the previous versions of PARALLHDG, PARMALLH6, PROLSQ, and OPLS (OPLS is only used for the final refinement in explicit solvent).

PARALLHDG (Nilges et al., 1988; Kuszewski et al., 1992) uses a single repulsion term E_{repel} for the non-bonded interactions, which is given by:

$$E_{\text{repel}} = \begin{cases} 0 & , \quad r > r_{\min} \\ k_{\text{vdW}}(s^2 r_{\min}^2 - r^2)^2, & r \le r_{\min} \end{cases} \qquad (3)$$

where k_{vdW} is the energy constant, r is the distance between two atoms, and r_{min} is the sum of their vdW radii. The values of the vdW radii are adopted from DISGEO (Havel and Wüthrich, 1985) (Table 1). The default value of the scaling factor s is 0.78; for $s = 0.8$ the radii are identical to DISGEO except for the O and N atoms.

PARMALLH6 uses the same repel function as PARALLHDG, but the values of the vdW radii in PARMALLH6 are taken from the CHARMM (Brooks et al., 1983) empirical energy function and have been scaled by s to approximately match the DISGEO values (Table 1). The scaling factor s is normally set to a value of 0.8 (Nilges et al., 1988).

PROLSQ is widely used in X-ray structure refinement. The non-bonded parameters from PROLSQ are used in conjunction with the CSDX parameters in the X-ray refinement force field of CNS (Adams et al., 1997). It uses a different energy function for the non-bonded interactions

$$E_{\text{repel}} = \begin{cases} 0 & , \quad r > r_{\min} \\ k_{\text{vdW}}(s r_{\min} - r)^4, & r \le r_{\min} \end{cases} \qquad (4)$$

Since X-ray refinement is often done without hydrogens, these parameters describe extended atoms, i.e. they include the hydrogen parameters in the vdW radii of the heavy atom. We have used uniformly small hydrogen radii (0.8 Å) to avoid clashes between the aliphatic hydrogens and the extended atoms. The other distinction of the PROLSQ non-bonded parameters is the reduced radii for the 1–4 interactions (1–4 interactions are defined as interactions between atoms separated by three bonds). For all atom types the parameters r_{vdW} and r_{vdW14} are summarized in Table 1. The radius of the O atom is reduced in order to allow H-bonds to the O atom.

OPLS (Jorgensen and Tirado-Rives, 1988) uses the standard Lennard-Jones and Coulomb forms:

$$E_{\text{non-bonded}} = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) + \frac{Q_i Q_j}{\varepsilon_0 r} \qquad (5)$$

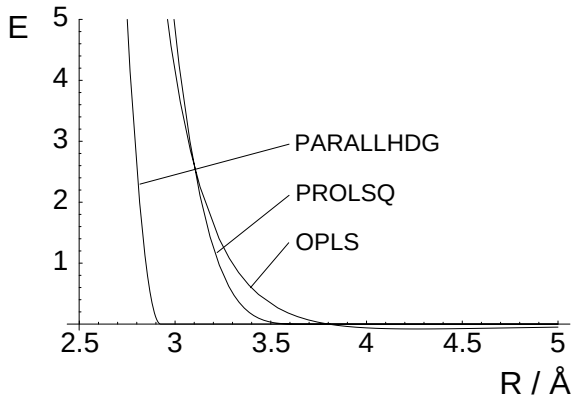

Figure 1. Different energy functions for non-bonded interactions (the atom type CH1E is taken as an example).

where ε is the well-depth, $r_{min} = 2^{1/6}\sigma$, Q_i is the charge on atom i, and ε_0 is the dielectric constant. The vdW parameters are derived from liquid simulations rather than from crystal data. The differences between the normal and the 1–4 interactions are realized by different values for ε and ε_{14} rather than for r_{vdW} and r_{vdW14} (Table 1). We merged the CSDX and the OPLS parameters without introducing new atom types. Some of the atoms in the OPLS force field do not have a direct match in the CSDX force field. These were replaced by the closest existing atom type, e.g. the sp^2 C atom in Arg is treated like a standard backbone C or the methyl groups of Val, Ile CD and Met are treated as those of Ala, Ile CG2, Leu and Thr.

In Figure 1 the PARALLHDG, PROLSQ, and OPLS repel functions for the non-bonded interactions are shown for the atom type CH1E (tetrahedral C atom with one H atom) as an example. The PROLSQ repel function is more similar to the repulsive part of the Lennard-Jones function. PARALLHDG has smaller radii due to the rationale that they are true all-atom representatives (similarly to DISGEO and DISMAN (Braun and Gō, 1985)). Furthermore, the 1–4 interactions were not treated separately in the PARALLHDG parameters, so that reduced radii were necessary to avoid 1–4 clashes.

Structure calculation

All structures were calculated with an ab initio simulated annealing protocol with X-PLOR 3.851 (Brünger, 1992). Similar to previous protocols using Cartesian MD (Nilges and O'Donoghue, 1998) and TAD simulated annealing (Stein et al., 1997), the protocol starts from random structures with good local geometry (Nilges et al., 1991) and consists of four

Table 1. Comparison of the vdW parameters of the force fields PARALLHDG (s = 0.78), PARMALLH6 (s = 0.8), PROLSQ and CSDX/OPLS. The shown atom types are explained in Engh and Huber (1991)

	HDG r_{vdW}	H6 r_{vdW}	PROLSQ		OPLS			
			r_{vdW}	r_{vdW14}	ε	r_{vdW}	ε_{14}	r_{vdW14}
H	0.9750	0.6400	0.80	0.65	0.050	0.2806	0.004	0.2806
HA	0.9750	1.1744	0.80	0.65	0.050	0.2806	0.004	0.2806
HC	0.9750	0.6400	0.80	0.65	0.050	0.2806	0.004	0.2806
C	1.4625	1.4400	1.70	1.55	0.105	2.1046	0.013	2.1046
C5	1.4625	1.4400	1.85	1.70	0.145	2.1046	0.145	2.1046
C5W	1.4625	1.4400	1.85	1.70	0.145	2.1046	0.018	2.1046
CF	1.4625	1.4400	1.85	1.70	0.110	2.1046	0.014	2.1046
CW	1.4625	1.4400	1.85	1.70	0.145	2.1046	0.018	2.1046
CY	1.4625	1.4400	1.85	1.70	0.110	2.1046	0.014	2.1046
CY2	1.4625	1.4400	1.85	1.70	0.105	2.1046	0.013	2.1046
CH1E	1.4625	1.4400	1.85	1.70	0.080	2.1327	0.010	2.1327
CH2E	1.4625	1.4400	1.85	1.70	0.118	2.1916	0.015	2.1916
CH2G	1.4625	1.4400	1.85	1.70	0.118	2.1327	0.015	2.1327
CH2P	1.4625	1.4400	1.85	1.70	0.118	2.1916	0.015	2.1916
CH3E	1.4625	1.4400	1.85	1.70	0.160	2.1944	0.020	2.1944
CR1E	1.4625	1.4400	1.85	1.70	0.110	2.1046	0.014	2.1046
CR1H	1.4625	1.4400	1.85	1.70	0.145	2.1046	0.018	2.1046
CR1W	1.4625	1.4400	1.85	1.70	0.110	2.1046	0.014	2.1046
CRHH	1.4625	1.4400	1.85	1.70	0.145	2.1046	0.018	2.1046
CRH	1.4625	1.4400	1.85	1.70	0.145	2.1046	0.145	2.1046
N	1.3163	1.2400	1.5	1.35	0.170	1.8240	0.021	1.8240
NC2	1.3163	1.2400	1.5	1.35	0.170	1.8240	0.021	1.8240
NH1	1.3163	1.2400	1.5	1.35	0.170	1.8240	0.021	1.8240
NH2	1.3163	1.2400	1.5	1.35	0.170	1.8240	0.021	1.8240
NH3	1.3163	1.2400	1.5	1.35	0.170	1.8240	0.021	1.8240
NR	1.3163	1.2400	1.5	1.35	0.170	1.8240	0.021	1.8240
O	1.2150	1.1856	1.45	1.3	0.210	1.6613	0.021	1.6613
OC	1.2150	1.1856	1.45	1.3	0.210	1.6613	0.021	1.6613
OH1	1.2150	1.1856	1.45	1.3	0.170	1.7230	0.021	1.7230
S	1.6398	1.5200	1.8	1.65	0.250	1.9924	0.031	1.9924
SM	1.6398	1.5200	1.8	1.65	0.250	1.9924	0.031	1.9924
SH1E	1.6398	1.5200	1.8	1.65	0.250	1.9924	0.031	1.9924

stages: a high-temperature TAD stage, two cooling stages, and a final minimization stage.

The masses are set uniformly to 100 amu and the friction coefficient f_{beta} (for the coupling to the external bath) is set to 20 ps^{-1}. The high-temperature TAD stage consists of 2000 TAD steps with a timestep of 0.045 ps at a bath temperature of 9000 K. The weight on the distance restraint energy term increases from 0.2 to 1 (10 to 50 kcal mol^{-1} $Å^{-2}$) during this stage; the energy constant for the experimental dihedrals is 150 kcal mol^{-1} rad^{-2}. A reduced non-bonded potential is used (Nilges and O'Donoghue, 1998), and the energy constant is increased from 0.001 to 0.02 kcal mol^{-1} $Å^{-4}$. The first cooling stage consists of 1000 TAD steps. The temperature decreases from 2000 K to 1000 K. The weight w_{vdW} of the vdW energy term E_{vdW} increases from 0.1 to 1. Non-bonded interactions are only evaluated between non-hydrogen atoms. The energy constant for the experimental dihedrals is 200 kcal mol^{-1} rad^{-2}. The second cooling stage consists of 1000 steps. All non-bonded interactions are treated explicitly. The temperature decreases from 1000 K to 50 K. The final minimization consists of 200 steps of restrained Powell minimization.

The energy constants were set to 50 kcal mol^{-1} $Å^{-2}$ for the NOEs and 200 kcal mol^{-1} rad^{-2} for the experimental dihedrals. A hundred structures were calculated. The 25 structures with lowest total energy were taken as an ensemble for the structure validation.

Structure calculations using only the NOE restraints

We calculated five structure ensembles with five different force fields using the NOE restraints, but without the torsion angle restraints and without the dihedral energy term (Table 2):

1. PARALLHDG with s = 0.78 (the standard value).
2. PARALLHDG with s = 0.75 (smaller vdW radii).
3. PARMALLH6 with s = 0.8 (the standard value). In the first calculation with the geometric energy function PARMALLH6, vdW radii were taken from the CHARMM force field and scaled by a factor of 0.8 to approximately match the parameters used in the distance geometry programs DISMAN and DISGEO.
4. PARMALLH6 with s = 0.89. We also performed a calculation with the scale factor 0.89, which corresponds to the zero-crossing of the Lennard-Jones function (Equation 5).
5. PROLSQ with s = 1.0

Structure calculations with NOE restraints, torsion angle restraints and dihedral energy functions

We calculated NMR structure ensembles using the PROLSQ force field and taking new energy function terms and additional restraints from J-couplings into account (Table 2):

6. PROLSQ with the dihedral energy function term for the sidechains.
7. PROLSQ with the dihedral energy term for the sidechains and including torsion angle restraints derived from J-couplings.
8. PROLSQ with the dihedral energy term for the sidechains and the backbone and including torsion angle restraints derived from J-couplings.

Water refinement

The 30 best structures (regarding total energy) from the calculations 5–8 were further refined in explicit solvent to remove artefacts due to the simple repulsive representation for the non-bonded energy term (calculations 9–12, Table 2). We used essentially the same protocol as previously (Prompers et al., 1995) with the new merged CSDX/OPLS force field.

9. water refinement of 5.
10. water refinement of 6.
11. water refinement of 7.
12. water refinement of 8.

Structure validation

The quality of the 25 best calculated structures was assessed with PROCHECK, PROSA and WHATIF.

We determined the content of residues with ϕ-ψ values in the most favoured, additional allowed, generously allowed and disallowed regions of the Ramachandran plot (Ramachandran et al., 1963) and the equivalent resolution (according to the distribution of ϕ-ψ angles) with PROCHECK-NMR (Laskowski et al., 1996).

We did the following checks with the program WHATIF (Vriend, 1990; Hooft et al., 1996): bond angle check (ANGCHK), backbone conformation normality check (BBCCHK), bump check (BMPCHK), bond length check (BNDCHK), buried unsatisfied H-bond donors/acceptors check (BPOCHK), χ_1-χ_2 check (C12CHK), torsion angle check (CHICHK), check for potentially flipped peptide planes (FLPCHK), check of the inside/outside distribution of residue types (INOCHK), check of the omega angle distribution (OMECHK), check of the side chain planarity (PLNCHK), packing quality control (QUACHK), calculation of the Ramachandran Z-score (RAMCHK) and check of the residue rotamers (ROTCHK). The Ramachandran Z-score of WHATIF, which is strongly correlated with the percentage of residues in the most favoured regions of the Ramachandran plot, is a good indicator for the overall quality of the ensembles (Hooft et al., 1997).

The program PROSA-II (Sippl, 1993) was used to determine mean force potentials. We averaged the mean force potential (determined with a window size of 1) over all residues. Then we averaged the mean force potential over all 25 models of the ensembles 1–12.

For the calculation of the average structures we used a protocol which iteratively determines the well-defined regions of a structure by excluding all residues for which the average C_α-distance from the average structure exceeds two standard deviations (Nilges et al., 1987). The average structures were refined against probability maps in order to get a more representative structure than a simple geometric average (DeLano and Brünger, 1994).

We used the program DSSP (Kabsch and Sander, 1983) to determine the exact location of the four α-

Table 2. Overview of the parameters used in the calculations 1–12

Calculation number	Non-bonded parameters	Scale factor s	Sidechain dihedrals	Torsion angle restraints	Backbone dihedrals
1	PARALLHDG	0.78	off	off	off
2	PARALLHDG	0.75	off	off	off
3	PARMALLH6	0.80	off	off	off
4	PARMALLH6	0.89	off	off	off
5	PROLSQ	1.0	off	off	off
6	PROLSQ	1.0	on	off	off
7	PROLSQ	1.0	on	on	off
8	PROLSQ	1.0	on	on	on
9	OPLS	–	off	off	off
10	OPLS	–	on	off	off
11	OPLS	–	on	on	off
12	OPLS	–	on	on	on

helices. The obtained positions of the four α-helices (6–19, 41–59, 70–94, 109–128) are almost identical to those published (Smith et al., 1994).

We calculated the RMSD between the average NMR structures and the 2.25 Å X-ray structure 1RCB (Wlodawer et al., 1992) or the 2.35 Å X-ray structure 2INT (Walter et al., 1992). As a second measure for the difference between two structures we calculated the contact area difference between the average structures and the 2.25 Å X-ray structure 1RCB or the 2.35 Å X-ray structure 2INT with the program CAD (Abagyan and Totrov, 1997).

Results and discussion

Comparison of the quality of the ensembles calculated with the force fields 1–5

In the comparison of the quality of the structural ensembles some general trends are obvious (Table 3, Figure 2).

Calculation 5 with PROLSQ gives the best quality indices (best Ramachandran quality, a small number of bumps, best backbone conformation Z-score by BBCCHK, best packing quality Z-score by QUACHK, good PROSA-II mean-force potential), with the exception of the number of unsatisfied H-bond donors and acceptors (BPOCHK) and the RMSD to the X-ray structure. The latter may be due to the fact that preliminary structures for the NOE assignment (Powers et al., 1992b) were originally calculated with vdW radii determined from the PARMALLH6 force field and that the NOE list is therefore optimized for use with this particular force field. In contrast, the CAD to the X-ray structures, which has been proposed as a better measure for the accuracy than the Cartesian RMSD (Abagyan and Totrov, 1997), indicates a better agreement between the X-ray structures and the structures from calculation 5 than those from calculation 3 with PARMALLH6. At present, we are investigating the influence of different distance calibrations and error-bounds on the structures, in combination with different force fields.

Calculation 4, with PARMALLH6 and $s = 0.89$, shows some special features. Due to the large vdW radii, it is not surprising that the number of too short interatomic distances (bumps) is very small, but interestingly, a good Ramachandran quality and a good χ_1–χ_2 Z-score is achieved. A possible explanation is that the conformational space is restricted by the large vdW radii to more standard conformations of the torsion angles. However, other indices like the QUACHK score and the PROSA energy are worse for $s = 0.89$.

The number of unusual backbone conformations per model and the backbone conformation Z-score, which is calculated by searching the WHATIF database for fragments of five C_α coordinates that are not represented by other structures, are distinctly better for PROLSQ than for the other force fields.

As we know from high-resolution X-ray structures, the number of unsatisfied hydrogen bond donors and acceptors (BPOCHK) should be close to zero. It might be useful to investigate the influence of an energy

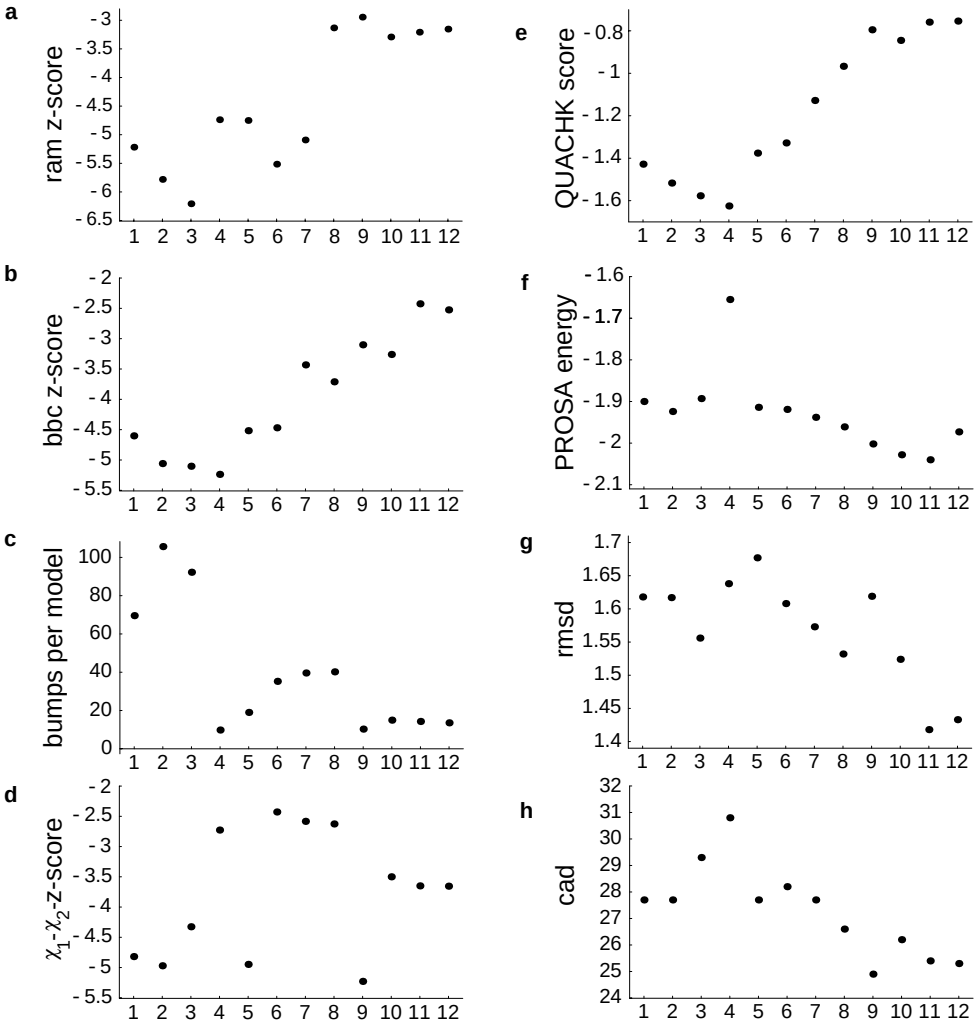

Figure 2. Quality indices for the structure ensembles of the calculations 1–12 (for an explanation of the abbreviations, see Table 3).

function term for the hydrogen bonds during the last cooling stage in order to decrease this number.

The χ_1–χ_2 Z-score is bad for all force fields except the PARMALLH6 force field with s = 0.89 (in this case the repulsion between aliphatic H atoms leads to more reasonable χ_1–χ_2 angles). However, one cannot expect to have good χ_1–χ_2 distributions because torsion angle restraints and dihedral energy terms were not used for the calculations 1–5. Some general problems, which occur in all calculations, are the too small bond angle, bond length, and ω-angle variability. This is an effect of a too strong restraining of the geometrical energy function terms. It would be useful to investigate the effects of weaker restraining on the overall accuracy and precision of the structures.

Improvement of the structures through dihedral energy terms and torsion angle restraints from J-couplings

As expected, the use of the dihedral energy term for the sidechains in calculation 6 results in a tremendous improvement of the χ_1–χ_2 Z-score. However, the additional restraining energy term leads to an increase of the number of bumps per model, the number of unsatisfied H-bond donors and acceptors and a decrease of the Ramachandran Z-score. The RMSD to the X-ray structures decreases whereas the CAD to the X-ray structures is, interestingly, almost not influenced.

More experimental data in the form of torsion angle restraints (calculation 7) give rise to structures of higher accuracy (better RMSD and CAD to the X-ray

Table 3. Quality indices of the calculated structure ensembles

Force field	PARALLHDG		PARMALLH6		PROLSQ				PROLSQ/OPLS			
Calculation	1	2	3	4	5	6	7	8	9	10	11	12
MostFav	65.6	65.0	61.0	65.8	**67.2**	66.3	72.7	75.3	74.5	73.9	76.7	76.4
AddAll	26.9	25.8	28.6	**25.3**	26.2	26.4	22.8	16.8	21.3	20.9	19.7	20.5
GenAll	5.2	6.8	8.0	6.5	**4.7**	5.3	3.7	6.6	2.7	3.5	2.7	2.0
Disall	2.3	2.4	2.5	2.4	**1.9**	2.0	0.9	1.4	1.5	1.7	0.9	1.1
RAMZ	−5.22	−5.78	−6.21	−4.75	**−4.74**	−5.52	−5.09	−3.13	−2.95	−3.29	−3.21	−3.15
EquivRes	3.0	3.1	3.2	3.0	**3.0**	3.0	2.7	2.6	2.6	2.6	2.5	2.5
UnBC	27.4	28.6	29.8	29.2	**25.1**	26.4	19.9	20.2	17.4	20.1	13.9	13.4
BBCZ	−4.60	−5.06	−5.10	−5.24	**−4.52**	−4.47	−3.43	−3.71	−3.10	−3.26	−2.43	−2.53
BMPCHK	69.6	105.6	92.2	**9.8**	19.0	35.2	39.6	40.2	10.3	15.0	14.3	13.6
BPOCHK	28.1	29.5	29.7	**27.5**	33.0	32.7	33.2	23.8	23.5	27.2	26.7	25.6
C12CHKZ	−4.82	−4.97	−4.33	**−2.73**	−4.95	−2.43	−2.58	−2.63	−5.23	−3.50	−3.65	−3.66
QUACHK	−1.43	−1.52	−1.58	−1.63	**−1.38**	−1.33	−1.13	−0.97	−0.80	−0.85	−0.76	−0.75
PROSA	−1.90	**−1.92**	−1.89	−1.66	−1.91	−1.92	−1.94	−1.96	−2.00	−2.03	−2.04	−1.97
RMSD *AV − 1RCB*	1.62	1.62	**1.56**	1.64	1.68	1.61	1.57	1.53	1.62	1.52	1.42	1.43
RMSD *AV − 2INT*	1.60	1.60	**1.54**	1.62	1.66	1.59	1.56	1.52	1.60	1.50	1.40	1.41
CAD *AV − 1RCB*	**27.7**	**27.7**	29.3	30.8	**27.7**	28.2	27.7	26.6	24.9	26.2	25.4	25.3
CAD *AV − 2INT*	28.4	28.2	29.6	30.8	**28.1**	28.2	28.0	27.2	25.0	26.1	25.8	25.6
RMSD *PM − 1RCB*	1.63	1.77	1.69	**1.56**	1.83	1.64	1.67	1.60	1.58	1.49	1.45	1.48
RMSD *PM − 2INT*	1.61	1.76	1.68	**1.55**	1.81	1.63	1.66	1.58	1.56	1.47	1.43	1.46
CAD *PM − 1RCB*	**26.6**	27.1	27.8	28.4	28.1	27.8	26.0	26.0	25.6	25.6	25.0	24.8
CAD *PM − 2INT*	**27.1**	27.8	28.3	27.9	28.4	27.4	27.1	26.0	25.5	25.0	24.9	24.4

MostFav, AddAll, GenAll, Disall: percentage of residues in most favoured, additionally allowed, generously allowed and disallowed regions of the Ramachandran plot; RAMZ: Ramachandran Z-score; EquivRes: equivalent resolution; UnBC: number of unallowed backbone conformations; BBCZ: backbone conformation Z-score; BMPCHK: number of bumps per model; BPOCHK: number of buried unsatisfied h-bond donors and acceptors per model; C12CHK: χ_1−χ_2 Z-score; QUACHK: quality packing Z-score; PROSA: mean force potential; RMSD: root mean square deviation; AV: average structure; PM: average structure refined against probability map. The best values of the calculations 1–5 are shown in bold.

structures) and lead to a better packing quality Z-score and PROSA-II mean force potential.

The main impact of the dihedral energy function for the backbone dihedral angles (calculation 8) is an improvement of the Ramachandran quality, which subsequently leads to a higher accuracy (lower RMSD and CAD values), a better packing quality Z-score and PROSA-II score.

Recently, Kuszewski et al. (1996, 1997) introduced a refinement against a conformational database potential. To some degree, the dihedral angle terms in the force field have a similar effect, as evidenced by the drastic improvement of the C12CHK score from WHATIF (see Table 3). However, the correlations between angles are not directly included. The dihedral angle terms can simply be used in the new PARALLHDG parameters, and in this paper, we restrict ourselves to a description of the properties of the new PARALLHDG parameters.

Improvement of the structures through water refinement

The effects of the water refinement with the CSDX/OPLS parameters are striking. Almost all quality indices and the accuracy, as measured by the RMSD and CAD to the X-ray structures, were significantly improved. The water refinement does not produce artefacts (e.g. unrealistic close packing) which may result from vacuum refinement with full vdW. In addition, the hybrid CSDX/OPLS force field avoids distortions in the water refinement of covalent interactions. The amount of CPU time needed on an R10000 processor is approximately 45 min per structure and therefore within an affordable range.

Conclusions

The development and improvement of the force fields used for X-ray and NMR structure refinement is an ongoing process. By comparing structural ensembles calculated with different energy functions and parameters for the non-bonded interactions, we have shown that the calculated structures obtained with the non-bonded energy function of PROLSQ achieve a higher quality than previous non-bonded representations. Our results clearly indicate that larger vdW radii than commonly employed in virtually all NMR structure calculation programs significantly improve the quality of the structures, if appropriate corrections for 1–4 interactions are employed.

The water refinement with the new merged parameter sets from the CSDX and OPLS force fields leads to a significant further improvement of the validation results and the accuracy of the structure ensembles in a computationally efficient way. The water refinement is carried out with covalent interactions identical to the structure generation. The refinement therefore does not lead to distortions in the covalent geometry. The OPLS parameters can be switched on by simply setting a flag in the parameter file.

The new parameter file PARALLHDG5.0.PRO, the new topology file TOPALLHDG5.0.PRO and the protocols for the refinement in explicit water can be downloaded from the following URL: www.nmr.embl-heidelberg.de/nmr/nilges.

Acknowledgements

We thank G. Marius Clore for making the NOE and torsion angle restraint lists available in the PDB, Gert Vriend for useful discussions and Seán O'Donoghue and Stephen Weeks for a critical reading of the manuscript. J.L. thanks the Boehringer Ingelheim Fonds for financial support through a Ph.D. fellowship.

References

Abagyan, R.A. and Totrov, M.M. (1997) *J. Mol. Biol.*, **268**, 678–685.

Adams, P.D., Pannu, N.S., Read, R.J. and Brünger, A.T. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 5018–5023.

Allen, F., Brice, M., Cartwright, B., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B., Kennard, O., Motherwell, W., Rodgers, J. and Watson, D. (1979) *Acta Crystallogr.*, **B35**, 2331–2339.

Braun, W. and Gō, N. (1985) *J. Mol. Biol.*, **186**, 611–626.

Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S. and Karplus, M. (1983) *J. Comput. Chem.*, **4**, 187–217.

Brünger, A.T., Adams, P.D. and Rice, L.M. (1997) *Structure*, **5**, 325–336.

Brünger, A. (1992) *X-Plor. A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

DeLano, W.L. and Brünger, A.T. (1994) *Proteins*, **20**, 105–123.

Engh, R. and Huber, R. (1991) *Acta Crystallogr.*, **A47**, 392–400.

Havel, T.F. and Wüthrich, K. (1985) *J. Mol. Biol.*, **182**, 281–294.

Hendrickson, W.A. (1985) *Methods Enzymol.*, **115**, 252–270.

Hooft, R.W., Vriend, G., Sander, C. and Abola, E.E. (1996) *Nature*, **381**, 272.

Hooft, R.W., Sander, C. and Vriend, G. (1997) *Comput. Appl. Biosci.*, **13**, 425–430.

Jorgensen, W. and Tirado-Rives, J. (1988) *J. Am. Chem. Soc.*, **110**, 1657–1666.

Kabsch, W. and Sander, C. (1983) *Biopolymers*, **22**, 2577–2637.

Kuszewski, J., Nilges, M. and Brünger, A.T. (1992) *J. Biomol. NMR*, **2**, 33–56.

Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1996) *Protein Sci.*, **5**, 1067–1080.

Kuszewski, J., Gronenborn, A. and Clore, G. (1997) *J. Magn. Reson.*, **125**, 171–177.

Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.

Müller, T., Dieckmann, T., Sebald, W. and Oschkinat, H. (1994) *J. Mol. Biol.*, **237**, 423–436.

Müller, T., Oehlenschläger, F. and Buehner, M. (1995) *J. Mol. Biol.*, **247**, 360–372.

Nilges, M. and O'Donoghue, S. (1998) *Progr. NMR Spectrosc.*, **32**, 107–139.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1987) *FEBS Lett.*, **219**, 17–21.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **239**, 129–136.

Nilges, M., Kuszewski, J. and Brünger, A.T. (1991) In *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy*, (Hoch, J.C., Poulsen, F.M. and Redfield, C., Eds) Plenum Press, New York, NY, pp. 451–455.

Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M. and Clore, G.M. (1992a) *Biochemistry*, **31**, 4334–4346.

Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M. and Clore, G.M. (1992b) *Science*, **256**, 1673–1677.

Prompers, J.J., Folmer, R.H., Nilges, M., Folkers, P.J., Konings, R.N. and Hilbers, C.W. (1995) *Eur. J. Biochem.*, **232**, 506–514.

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) *J. Mol. Biol.*, **7**, 95–99.

Redfield, C., Boyd, J., Smith, L.J., Smith, R.A. and Dobson, C.M. (1992) *Biochemistry*, **31**, 10431–10437.

Sippl, M.J. (1993) *Proteins*, **17**, 355–362.

Smith, L.J., Redfield, C., Boyd, J., Lawrence, G.M., Edwards, R.G., Smith, R.A. and Dobson, C.M. (1992) *J. Mol. Biol.*, **224**, 899–904.

Smith, L.J., Redfield, C., Smith, R.A., Dobson, C.M., Clore, G.M., Gronenborn, A.M., Walter, M.R., Naganbushan, T.L. and Wlodawer, A. (1994) *Nat. Struct. Biol.*, **1**, 301–310.

Stein, E.G., Rice, L.M. and Brünger, A.T. (1997) *J. Magn. Reson.*, **124**, 154–164.

Vriend, G. (1990) *J. Mol. Graph.*, **8**, 52–56.

Walter, M.R., Cook, W.J., Zhao, B.G., Cameron Jr., R.P., Ealick, S.E., Walter Jr., R.L., Reichert, P., Nagabhushan, T.L., Trotta, P.P. and Bugg, C.E. (1992) *J. Biol. Chem.*, **267**, 20371–20376.

Wlodawer, A., Pavlovsky, A. and Gustchina, A. (1992) *FEBS Lett.*, **309**, 59–64.