*Journal of Biomolecular NMR*, **22:** 281–289, 2002.
*KLUWER/ESCOM*



# Improving the quality of protein structures derived by NMR spectroscopy**

Christian A.E.M. Spronk[a,c], Jens P. Linge[b], Cornelis W. Hilbers[a] & Geerten W. Vuister[a,*]
[a]*Department of Biophysical Chemistry, NSR-RIM Center, University of Nijmegen, The Netherlands;* [b]*Unité de Bioinformatique Structurale, Institut Pasteur, Paris, France;* [c]*Present adress: Centre for Molecular and Biomolecular Informatics, University of Nijmegen, The Netherlands*

Received 2 January 2002; Accepted 3 January 2002

*Key words:* NMR structures, proteins, structure refinement, structure validation

**Abstract**

Biomolecular structures provide the basis for many studies in several research areas such as homology modelling, structure-based drug design and functional genomics. It is an important prerequisite that the structure is reliable in terms of accurate description of the experimental data, and in terms of good quality of local- and overall geometry. Recent surveys indicate that structures solved by NMR-spectroscopy normally are of lower precision than high-resolution X-ray structures. Here, we present a refinement protocol that improves the quality of protein structures determined by NMR-spectroscopy to the level of those determined by high resolution X-ray crystallography in terms of local geometry. The protocol was tested on experimental data of the proteins IL4 and Ubiquitin and on simulated data of the protein Crambin. In almost all aspects, the protocol yielded better results in terms of accuracy and precision. Independent validation of the results for Ubiquitin, using residual dipolar couplings, indicates that the ensemble of NMR structure is substantially improved by the protocol.

## Introduction

High-resolution NMR spectroscopy was first used as a tool for structure determination of proteins and nucleic acids at atomic resolution some 15 years ago. Since then, approximately 2500 NMR-derived structures have been deposited at the Brookhaven Protein Data Bank (PDB). On a total of ∼15000 bio-macromolecular structures, determined by several experimental- and theoretical techniques, NMR-derived structures thus represent a significant fraction. Apart from structural information, NMR also provides information about the dynamics of proteins and nucleic acids. This unique combination of structure and dynamics makes NMR a powerful tool for the analysis of dynamic events, such as folding transitions in proteins, which are of major importance for biological function.

A major problem with NMR-derived structures of bio-macromolecules is, however, the lower quality of the structures as compared to high-resolution X-ray structures. This problem, which becomes manifest in the non-optimal local geometry, electrostatics and packing quality (Doreleijers et al., 1998, 1999a), arises mainly from the lower experimental data-content (Doreleijers, 1999; Doreleijers et al., 1999b) and from the type of structure refinement protocol that is used. With respect to the experimental data-content it can be expected that a number of recently developed methods, such as measurement of hydrogen-bonding patterns (Dingley and Grzesiek, 1998), the use of residual dipolar couplings for biomolecules aligned in a magnetic field (Tjandra and Bax, 1997) and novel techniques for automated NOE assignment and structure calculation (Nilges et al., 1997) will further increase the data-content and quality of NMR-derived structures. On the structure calcu-

*To whom correspondence should be addressed. E-mail: vuister@nmr.kun.nl

lation and refinement part recent work has shown that the quality of NMR structures depends on the type of non-bonded parameters used and can be significantly improved by refinement in explicit solvent (Linge and Nilges, 1999). Other important recent developments in structure refinement techniques involve the application of data base refinement techniques to remove physically unlikely and energetically unfavorable torsion angles in NMR structures of proteins and nucleic acids (Kuszewski and Clore, 2000).

In order to have an up to date picture of the main problems that occur in NMR-derived structures of proteins we have performed a statistical survey of the quality of protein NMR-structures deposited in the year 2001 at the Brookhaven PDB. Next, we present a simple method to improve the quality of NMR-derived protein structures by removing the main problems in local geometry, electrostatics and packing quality. The method, which is based on the protocol previously used to refine the structure of the *lac*-repressor headpiece in complex with DNA (Spronk et al., 1999), involves a short molecular dynamics simulation in water using the CHARMM22 force field for proteins and nucleic acids (MacKerell et al., 1992). We will refer this protocol as the 'CHARMM22 water refinement' and have tested it on three different protein structures. The first structure is that of interleukin 4 (IL4) of intermediate quality (pdb-entry 1BBN, Powers et al., 1992), which has been used in previous studies where the effect of non-bonded parameters on the quality of NMR-structures was investigated (Linge and Nilges, 1999). Second, in order to assess the accuracy of the calculated structures obtained with different calculation protocols, we chose to calculate the structure of the protein Crambin based on an artificial perfect set of NOE distance restraints derived from the atomic resolution (0.54 Å) X-ray structure (pdb-entry 1EJG, Jelsch et al., 2000). Finally, we have used the high quality NMR-data of the protein Ubiquitin (pdb-entry 1D3Z, Cornilescu et al., 1998) for independent validation of different calculation protocols against observed dipolar couplings.

## Materials and methods

### *Analysis of PDB entries*

For our survey of the quality of NMR-derived structures we analyzed 37 NMR-ensembles of protein structures deposited at the PDB in 2001 and used 13 X-ray structures with resolutions better then 1Å as a comparison set (Figure 1). For the analysis we made use of the PROCHECK (Laskowski et al., 1993) and WHAT IF (Vriend, 1990) programs, two widely used packages for protein-structure analysis. Here, we have focused our attention on the following quality criteria: (1) Analysis of unrealistically short inter-atomic distances. PROCHECK and WHAT IF report these as being 'bad contacts'. (2) Analysis of local geometry quality indicators as provided by the checking routines of the program WHAT IF. WHAT IF compares a number of parameters from the analyzed structure with those present in a database build on information from high-resolution X-ray structures. The program provides a Z-score or RMS Z-score as a quality indicator for each parameter. It suffices here to say that a Z-score should be close to 0, where negative and positive deviations generally indicate that the distribution of values of the particular parameter are worse or better than average, respectively. RMS Z-scores on the other hand should be close to 1.

### *Structure calculations of IL4, Crambin and Ubiquitin*

Interleukin 4 structures were calculated in X-PLOR (Brünger, 1992) as described in Linge and Nilges (1999) (calculation schemes 5 and 8, Table 2 of Linge and Nilges, 1999). The resulting structures served as input structures for both the water refinement as described in Linge and Nilges (1999) (calculation schemes 9 and 12, Table 2 of Linge and Nilges, 1999), and the CHARMM22 water refinement (see below).

Crambin and Ubiquitin structures were calculated using standard structure calculation protocols in ARIA 1.0, with PROLSQ force field parameters and including dihedral angle energy terms. Input distance restraints for Crambin were calculated from the X-ray structure (pdb-entry 1EJG) and contained all proton-proton distances smaller than 5 Å, upon which we added 20% error bounds. For Ubiquitin we used distance-, dihedral angle- and hydrogen-bond restraints taken from pdb-entry 1D3Z. Water refinements of Crambin and Ubiquitin were performed using both the protocol in ARIA 1.0 and the CHARMM22 water refinement.

### *CHARMM22 water refinement*

The CHARMM22 water refinement is similar to the protocol described in Spronk et al. (1999). The main differences with the previously described protocol are the length of the simulation, the weighting of the

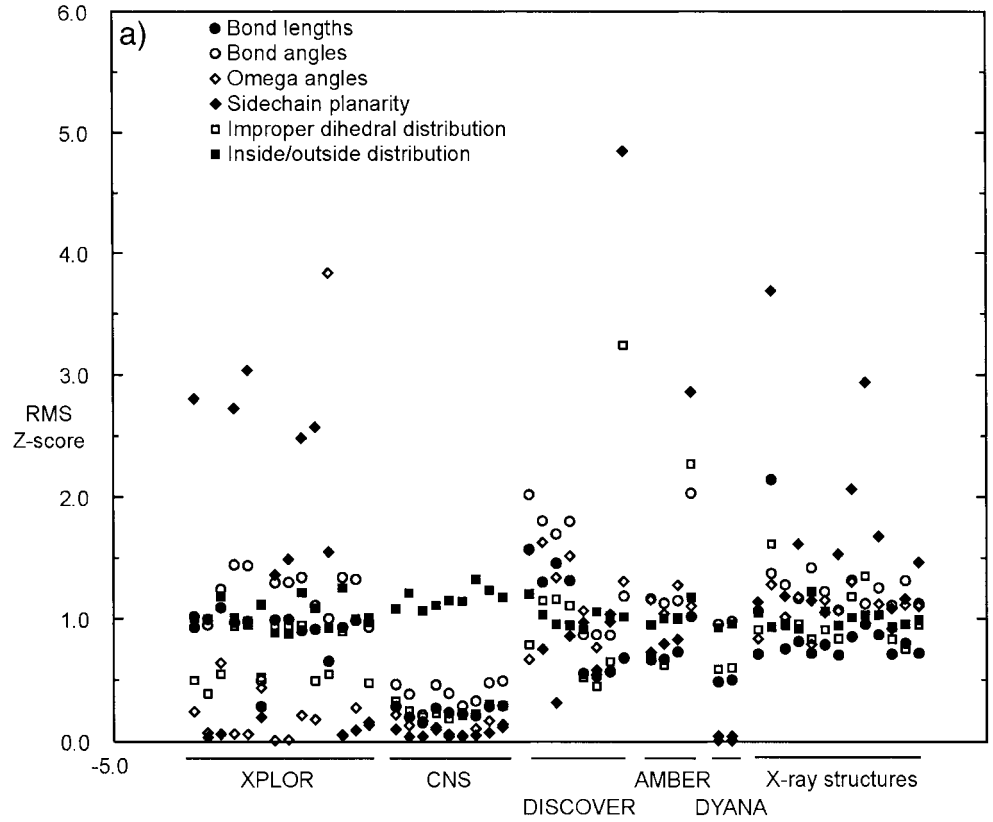


*Figure 1a.* RMS Z-scores for six parameters describing the local geometry of 37 NMR-ensembles of protein structures deposited at the Brookhaven PDB in 2001 (deposition codes: *1h8c, 1ha9, 1haj, 1hbw, 1hi7, 1hvw, 1hvz, 1hx2, 1hyi, 1hyj, 1hyk, 1hyw, 1hz3, 1hzk, 1hzl, 1i02, 1i0w, 1i11, 1i17, 1i1s, 1i25, 1i5h, 1i5j, 1i87, 1ib7, 1ic9, 1icl, 1ico, 1idh, 1idl, 1ig6, 1ija, 1ikc, 1imw, 1in2, 1in3 and 1j7q*) and 13 high resolution (<1 Å) X-ray structures (deposition codes: *1aho, 1bxo, 1byi, 1c75, 1et1, 1f94, 1gci, 1nls, 1rb9, 2fdn, 2pvb, 3pyp and 7a3h*). On the x-axis we have grouped the NMR structures according to the program used for the final refinement of the structures.

experimental restraints and the use of the original weights of the force constants for dihedral angles dealing with peptide planarity in the current protocol. Summarized the protocol consisted of the following steps: The protein structures were solvated in a rectangular water box (TIP3P water model, Jorgenson et al., 1983) with a minimum solute to wall distance of 10 Å. The systems were then neutralized and energy minimized, followed by the actual refinement of 3 ps of restrained MD at constant volume under periodic boundary conditions. All covalent bond lengths were constrained with the procedure SHAKE (Ryckaert et al., 1977) using a SHAKE-tolerance of $10^{-5}$. The initial temperature of the system was set to 300 K. During the first 0.1 ps of the simulation the solute was restrained using harmonic position restraints, followed by 0.5 ps of simulation at 300 K without harmonic position restraints. The system was then cooled in 4 steps of 50 K to a final temperature of 100 K. Force constants of potentials for experimental restraints were gradually lowered from 50 to 20 kcal $mol^{-1}$ $Å^{-2}$ and 200 to 80 kcal $mol^{-1}$ $Å^{-2}$ for NOE and torsion angle restraints, respectively. The resulting structures were finally energy minimized using 250 steps of energy minimization.

### *Structure validation*

For each molecule and structure calculation protocol 20 structures were used in the analysis. None of these structures contained violations of input distance- and dihedral angle restraints larger than 0.5 Å and 5°, respectively. All analyses of structures were done using PROCHECK (Laskowski et al., 1993), PROCHECK_NMR (Laskowski et al., 1996) and WHAT IF (Vriend, 1990). Averages and standard deviations were calculated from the checks of the individual structures of each ensemble. Unsatisfied hydrogen bond donors, -acceptors and inter atomic

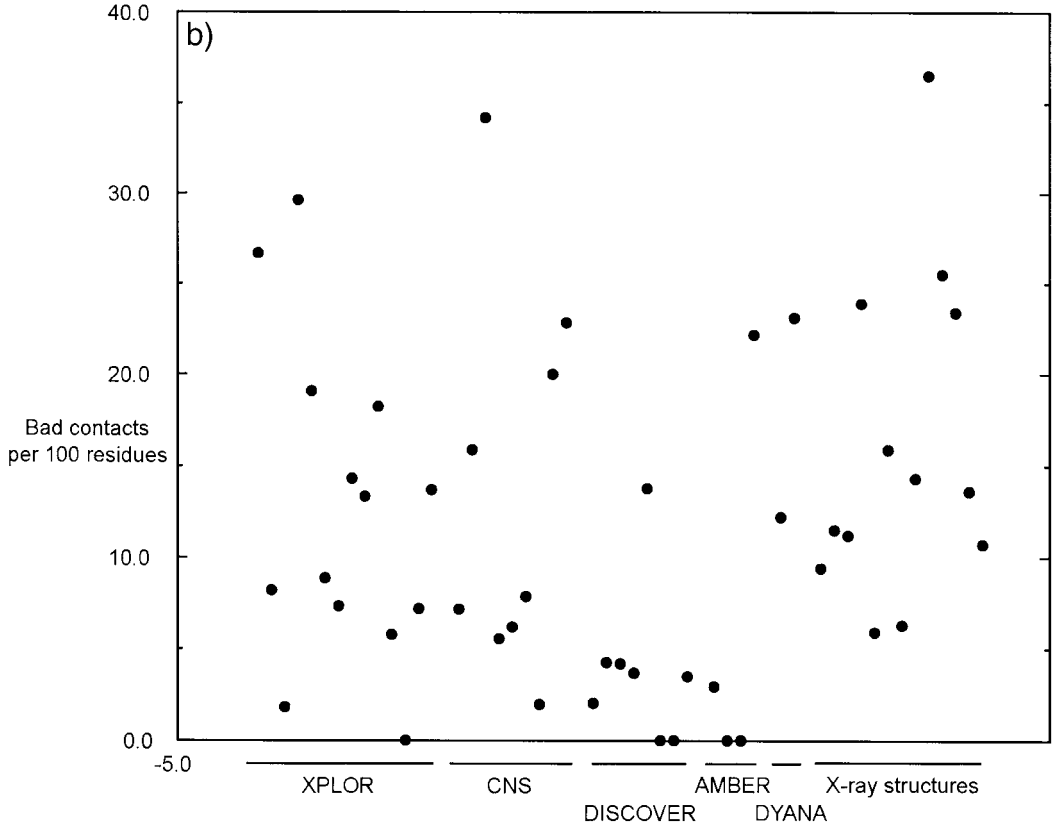


*Figure 1b.* Average number of bad contacts per 100 residues of the structures in Figure 1a from PROCHECK analyses.

bumps were analyzed by averaging the number of occurrences in the WHAT IF checking reports.

*Validation against dipolar couplings*

Dipolar couplings of Ubiquitin were previously measured in two different liquid crystalline phases for $C^{\alpha}$-$C^{\beta}$, $C^{\alpha}$-C′, N-C′, $H^{\alpha}$-$C^{\alpha}$, $H^{N}$-C′, $H^{N}$-N (Cornilescu et al., 2000). Validation of Ubiquitin structures against dipolar couplings was done using the program PALES (Zweckstetter and Bax, 2000) using the mode for obtaining the best fit of measured dipolar couplings to the 3D structures. All structures were individually analyzed using 1 set of dipolar couplings at a time. The final Cornilescu Q-factors are averages obtained for each ensemble. The X-ray structure of Ubiquitin (1UBQ) was protonated in X-PLOR prior to analysis.

## Results and discussion

*Quality of NMR-derived protein structures at the Brookhaven PDB*

In Figure 1a we have depicted the RMS. Z-scores for six different parameters describing the local geometry of the total of 50 different NMR and X-ray structures. We have classified the NMR structures according to the program used for the final refinement. It is important to note here that within a class of structures, different versions of programs and force fields often are used, which in some cases results in markedly different qualities of structures. Figure 1b shows the average number of bad contacts per 100 residues for all the structures.

It is clear from the results of our query that most of the recently deposited NMR structures have distributions of parameters describing the local geometry that do not correspond with those found for high-resolution X-ray structures. Further, there appears to be no consensus between the different software packages for the tightness of the different geometry restraints. The only

*Table 1.* Structure quality scores for Interleukin 4

| | Calculation 5[a] | | | Calculation 8[b] | | |
|---|---|---|---|---|---|---|
| | Unrefined | ARIA[c] | CHARMM22[d] | Unrefined | ARIA[c] | CHARMM22[d] |
| Structure Z-scores: | | | | | | |
| 2nd generation packing quality | $-3.68 \pm 0.25$ | $-1.94 \pm 0.30$ | $\mathbf{-1.74 \pm 0.41}$ | $-2.74 \pm 0.33$ | $-1.94 \pm 0.44$ | $\mathbf{-1.65 \pm 0.41}$ |
| Ramachandran plot appearance | $-4.65 \pm 0.39$ | $-2.95 \pm 0.41$ | $\mathbf{-2.36 \pm 0.42}$ | $-3.07 \pm 0.40$ | $-3.07 \pm 0.63$ | $\mathbf{-2.46 \pm 0.61}$ |
| Chi-1/chi-2 rotamer normality | $-4.83 \pm 0.21$ | $-4.78 \pm 0.28$ | $\mathbf{-1.46 \pm 0.39}$ | $-1.23 \pm 0.51$ | $-2.12 \pm 0.33$ | $\mathbf{-1.65 \pm 0.46}$ |
| Backbone conformation | $-4.16 \pm 0.65$ | $\mathbf{-3.02 \pm 0.75}$ | $-3.29 \pm 0.70$ | $-3.72 \pm 0.76$ | $\mathbf{-2.50 \pm 0.68}$ | $-2.81 \pm 0.79$ |
| RMS Z-scores: | | | | | | |
| Bond lengths | $0.24 \pm 0.01$ | $0.29 \pm 0.01$ | $\mathbf{1.03 \pm 0.01}$ | $0.26 \pm 0.01$ | $0.31 \pm 0.01$ | $\mathbf{1.03 \pm 0.01}$ |
| Bond angles | $0.39 \pm 0.01$ | $0.49 \pm 0.01$ | $\mathbf{1.21 \pm 0.02}$ | $0.42 \pm 0.01$ | $0.54 \pm 0.02$ | $\mathbf{1.20 \pm 0.02}$ |
| Omega angle restraints | $0.07 \pm 0.01$ | $0.19 \pm 0.01$ | $\mathbf{1.32 \pm 0.11}$ | $0.14 \pm 0.01$ | $0.22 \pm 0.02$ | $\mathbf{1.28 \pm 0.09}$ |
| Side chain planarity | $0.05 \pm 0.01$ | $0.24 \pm 0.02$ | $\mathbf{1.11 \pm 0.14}$ | $0.13 \pm 0.02$ | $0.35 \pm 0.03$ | $\mathbf{1.09 \pm 0.10}$ |
| Improper dihedral distribution | $0.16 \pm 0.01$ | $0.32 \pm 0.01$ | $\mathbf{0.96 \pm 0.05}$ | $0.23 \pm 0.01$ | $0.37 \pm 0.02$ | $\mathbf{0.94 \pm 0.03}$ |
| Inside/outside distribution | $0.96 \pm 0.02$ | $0.96 \pm 0.02$ | $\mathbf{0.97 \pm 0.03}$ | $0.97 \pm 0.03$ | $0.97 \pm 0.03$ | $\mathbf{0.98 \pm 0.03}$ |
| Inter-atomic bumps: | $19.10 \pm 4.73$ | $10.30 \pm 2.64$ | $\mathbf{1.75 \pm 1.48}$ | $39.05 \pm 6.72$ | $14.05 \pm 4.59$ | $\mathbf{2.00 \pm 1.41}$ |
| Unsatisfied H-bond donors: | $25.55 \pm 3.14$ | $16.50 \pm 2.76$ | $\mathbf{12.45 \pm 3.02}$ | $18.70 \pm 3.29$ | $19.45 \pm 3.30$ | $\mathbf{13.65 \pm 2.74}$ |
| Unsatisfied H-bond acceptors: | $9.85 \pm 3.22$ | $10.60 \pm 2.85$ | $\mathbf{3.75 \pm 1.68}$ | $8.70 \pm 2.76$ | $11.10 \pm 2.88$ | $\mathbf{4.30 \pm 1.66}$ |
| Ramachandran plot | | | | | | |
| Most favoured | 66.8 | 74.9 | **76.4** | 75.7 | 77.4 | **79.6** |
| Allowed | 26.8 | 21.2 | **17.3** | 16.7 | 19.2 | **16.0** |
| Generously allowed | 4.7 | **2.6** | 3.3 | 6.3 | **2.4** | 2.6 |
| Disallowed | 1.7 | **1.3** | 3.0 | 1.3 | **1.0** | 1.8 |
| Equivalent resolution | 3.0 | 2.6 | **2.5** | 2.6 | 2.5 | **2.3** |
| Rmsd to 1RCB (backbone/ | $2.00 \pm 0.26$ | $\mathbf{1.87 \pm 0.18}$ | $1.99 \pm 0.32$ | $1.77 \pm 0.17$ | $\mathbf{1.76 \pm 0.23}$ | $1.81 \pm 0.20$ |
| all heavy atoms)[e] | $2.95 \pm 0.23$ | $\mathbf{2.82 \pm 0.14}$ | $2.92 \pm 0.28$ | $2.78 \pm 0.20$ | $\mathbf{2.76 \pm 0.24}$ | $2.82 \pm 0.19$ |
| Rmsd to 2INT (backbone/ | $1.98 \pm 0.26$ | $\mathbf{1.84 \pm 0.18}$ | $1.98 \pm 0.32$ | $1.75 \pm 0.17$ | $\mathbf{1.75 \pm 0.23}$ | $1.80 \pm 0.21$ |
| all heavy atoms)[e] | $2.81 \pm 0.24$ | $\mathbf{2.67 \pm 0.16}$ | $2.80 \pm 0.28$ | $2.69 \pm 0.20$ | $\mathbf{2.67 \pm 0.24}$ | $2.73 \pm 0.20$ |

[a]Interleukin 4 structures calculated in X-PLOR using calculation scheme 5 as described in Linge and Nilges (1999).
[b]Idem using calculation scheme 8 as described in Linge and Nilges (1999).
[c]Water refinement as described in Linge and Nilges (1999).
[d]This study CHARMM22 water refinement.
[e]Average pairwise rmsd of the NMR structures to the X-ray structures for residues in α-helical regions (residues: 6–19, 41–59, 70–94, 109–128). The best values are shown in bold.

parameter that shows RMS-Z scores close to the expected value is the Inside/outside distribution, which is largely determined by the fold of the protein and not by the type of force field used. The other parameters, such as bond lengths, side chain planarity and omega angles are often too tightly restrained in NMR structures, with some exceptions, but not all, of the structures refined in DISCOVER or AMBER. Although for some parameters the effect of the tight restraining on the quality of the structure may not be severe, it is important that the geometry of the structures is properly restrained. For example, it has been noted that for omega angles there is a correlation between increased numbers of residues in most favored regions of the Ramachandran plot and increased standard deviations of the omega angle (Doreleijers et al., 1998).

The quality indicators for the X-ray structures show, apart from some exceptions, a clear consensus for the degree of geometry restraining with most of the RMS-Z scores being close to 1.0. In terms of inter-atomic overlaps there appears to be no significant difference in the quality of the NMR- and X-ray structures analyzed here, although the only structures free of such bumps have been determined by NMR.

Further, inspection of other important quality indicators, such as Ramachandran plots, packing quality, side chain-rotamers and backbone conformation indicates that in general NMR structures are significantly worse than the best available X-ray structures (Doreleijers et al., 1998, and data not shown). In addi-

*Table 2.* Structure quality scores for Crambin

| | Unrefined | ARIA[a] | CHARMM22[b] | 1EJG |
|---|---|---|---|---|
| Structure Z-scores: | | | | |
| 2nd generation packing quality | −0.77 ± 0.28 | −0.26 ± 0.40 | **0.08 ± 0.46** | 0.75 |
| Ramachandran plot appearance | −1.59 ± 0.38 | −3.02 ± 0.43 | **−0.86 ± 0.48** | 0.04 |
| Chi-1/chi-2 rotamer normality | **0.12 ± 0.26** | −1.53 ± 0.38 | −0.14 ± 0.48 | 0.44 |
| Backbone conformation | 0.23 ± 0.36 | −0.73 ± 0.52 | **0.99 ± 0.39** | 1.02 |
| RMS Z-scores: | | | | |
| Bond lengths | 0.22 ± 0.01 | 0.28 ± 0.01 | **1.08 ± 0.01** | 0.76 |
| Bond angles | 0.28 ± 0.01 | 0.37 ± 0.01 | **1.22 ± 0.05** | 1.04 |
| Omega angle restraints | 0.09 ± 0.01 | 0.16 ± 0.01 | **1.16 ± 0.12** | 0.99 |
| Side chain planarity | 0.41 ± 0.04 | 0.58 ± 0.08 | **1.04 ± 0.26** | 1.07 |
| Improper dihedral distribution | 0.24 ± 0.01 | 0.33 ± 0.02 | **0.75 ± 0.03** | 1.02 |
| Inside/outside distribution | **1.02 ± 0.01** | **1.02 ± 0.01** | **1.02 ± 0.01** | 1.01 |
| Inter-atomic bumps: | 6.45 ± 2.09 | 28.95 ± 3.97 | **0.70 ± 0.92** | 3 |
| Unsatisfied H-bond donors: | 2.20 ± 0.95 | **1.70 ± 0.73** | 2.40 ± 0.68 | 1 |
| Unsatisfied H-bond acceptors: | 3.60 ± 0.82 | 3.40 ± 1.31 | **3.95 ± 1.15** | 4 |
| Ramachandran plot | | | | |
| Most favoured | 87.3 | 87.1 | **94.9** | 97.2 |
| Allowed | 10.1 | 12.6 | **5.1** | 2.1 |
| Generously allowed | 2.6 | 0.3 | **0.0** | 0.0 |
| Disallowed | 0.0 | 0.0 | **0.0** | 0.0 |
| Equivalent resolution | | | | |
| Rmsd to 1EJG (backbone/ | 0.50 ± 0.10 | 0.59 ± 0.10 | **0.44 ± 0.06** | n.a. |
| all heavy atoms)[c] | 0.74 ± 0.09 | 0.86 ± 0.12 | **0.70 ± 0.08** | |

[a]Water refinement in ARIA 1.0.
[b]This study CHARMM22 water refinement.
[c]Average pairwise rmsd of the NMR structures to the X-ray structures (all residues). The analysis of the reference X-ray structure entry (PDB code: 1ejg) is added for comparison. The best values are shown in bold.

tion, it is not common practice to include electrostatic potentials in NMR structure calculations, which typically results in unrealistic charge distributions and unsatisfied internal hydrogen-bond donors and acceptors (*vide infra*).

In order to alleviate these common problems in NMR structures, we set out to explore how the calculation protocol could be modified to yield improved structures. We have found that this can be achieved by applying the CHARMM22 water refinement as a final refinement of structures calculated using standard calculation protocols. This water refinement is similar to that described by Linge and Nilges (1999), with the main differences being the use of the CHARMM22 force field and that potentials for NOE- and dihedral angle restraints are lowered in the course of the protocol. The CHARMM22 force field uses less tight restraining of the local geometry. Therefore, lowering the weights on the experimental terms is essential in order to keep a proper balance with the theoretical terms in the CHARMM22 force field. We have found that using constant high potentials on the experimental terms during the refinement does not have a significant effect on the quality of structures calculated with error free, i.e., simulated data, whereas in the case of real data the quality of structures worsens when using constant high potentials (data not shown).

### *Structure calculation and refinement of IL4, Crambin and Ubiquitin*

#### *Overall quality of the structures*

The results of the WHAT IF and PROCHECK analyses of the ensembles of structures of IL4, Crambin and Ubiquitin are given in Tables 1 to 3, respectively. For the quality indicators (RMS Z-scores) of the parameters describing the local geometry of the structures we see that structures refined with the CHARMM22 water refinement protocol have values that are well within the ranges found for high-resolution X-ray structures

*Table 3.* Structure quality scores for Ubiquitin

| | Unrefined | ARIA[a] | CHARMM22[b] |
|---|---|---|---|
| Structure Z-scores: | | | |
| 2nd generation packing quality | $-0.78 \pm 0.39$ | $\mathbf{0.03 \pm 0.39}$ | $-0.22 \pm 0.36$ |
| Ramachandran plot appearance | $0.10 \pm 0.44$ | $-0.26 \pm 0.34$ | $\mathbf{0.74 \pm 0.37}$ |
| Chi-1/chi-2 rotamer normality | $\mathbf{0.36 \pm 0.60}$ | $-1.08 \pm 0.39$ | $-0.19 \pm 0.63$ |
| Backbone conformation | $1.85 \pm 0.35$ | $1.11 \pm 0.34$ | $\mathbf{2.14 \pm 0.46}$ |
| RMS Z-scores: | | | |
| Bond lengths | $0.19 \pm 0.01$ | $0.32 \pm 0.01$ | $\mathbf{1.02 \pm 0.01}$ |
| Bond angles | $0.35 \pm 0.01$ | $0.44 \pm 0.01$ | $\mathbf{1.18 \pm 0.03}$ |
| Omega angle restraints | $0.15 \pm 0.01$ | $0.21 \pm 0.02$ | $\mathbf{1.12 \pm 0.10}$ |
| Side chain planarity | $0.14 \pm 0.05$ | $0.46 \pm 0.06$ | $\mathbf{1.12 \pm 0.15}$ |
| Improper dihedral distribution | $0.25 \pm 0.01$ | $0.37 \pm 0.02$ | $\mathbf{0.83 \pm 0.05}$ |
| Inside/outside distribution | $\mathbf{1.01 \pm 0.02}$ | $\mathbf{1.01 \pm 0.02}$ | $\mathbf{0.99 \pm 0.01}$ |
| Inter-atomic bumps: | $10.60 \pm 3.63$ | $29.65 \pm 9.54$ | $\mathbf{0.35 \pm 0.75}$ |
| Unsatisfied H-bond donors: | $4.65 \pm 1.23$ | $5.00 \pm 1.59$ | $\mathbf{2.95 \pm 1.40}$ |
| Unsatisfied H-bond acceptors: | $2.50 \pm 0.95$ | $2.30 \pm 1.03$ | $\mathbf{1.40 \pm 0.75}$ |
| Ramachandran plot | | | |
| Most favoured | 88.6 | 87.5 | **91.1** |
| Allowed | 9.4 | 10.7 | **8.3** |
| Generously allowed | 1.9 | 1.6 | **0.5** |
| Disallowed | **0.2** | **0.2** | **0.2** |
| Equivalent resolution | 1.4 | 1.6 | **1.0** |
| Rmsd to 1UBQ (bb/ | $0.60 \pm 0.06$ | $0.88 \pm 0.08$ | $\mathbf{0.58 \pm 0.05}$ |
| all heavy atoms)[c] | $1.34 \pm 0.08$ | $1.52 \pm 0.10$ | $\mathbf{1.27 \pm 0.09}$ |

[a]Water refinement in ARIA1.0.
[b]This study CHARMM22 water refinement.
[c] Average pairwise rmsd of the NMR structures to the X-ray structures for residues 1–70. The best values are shown in bold.

(Tables 1 to 3 and Figure 1). More important, however, is that the CHARMM22 water refined structures show clear overall improvements of the structure Z-scores, independent of the quality of the experimental input restraints. For example the Ramachandran Z-score, which is a good indicator of the overall quality of the ensembles (Hooft et al., 1997), shows consistent improvements in the structures. In addition, it is seen that the inter-atomic bumps are almost completely eliminated from the structures and that the network of buried hydrogen bonds is improved. This latter effect is most clearly seen in the interleukin 4 and Ubiquitin structures. For the Crambin structures these effects are less obvious since the positions of the hydrogen atoms in these structures are very restrained due to the use of the perfect NOE data set.

*Accuracy of the structures*

The atomic RMSD between X-ray and NMR structures is often used as an indicator of the accuracy of the NMR structures. In Tables 1 to 3 we have compared the results of the different calculation protocols with X-ray structures 1RCB (2.25 Å resolution, Wlodaver et al., 1992) and 2INT (2.35 Å resolution, Walter et al., 1992) of IL4, 1EJG of Crambin and 1UBQ (1.8 Å resolution, Vijay-Kumar et al., 1987) of Ubiquitin. Since real differences between X-ray structures and NMR structures cannot be excluded for IL4 and Ubiquitin, the effect of the calculation protocol on the accuracy of calculated structures is best reflected in the comparison of the artificial NMR structures of Crambin and its X-ray structure. From Table 2 we see that, although all structures of Crambin are good structures in terms of satisfying the input distance restraints, the structures closest to the real structure are those refined in the CHARMM22 refinement protocol. This is not only indicated by the atomic RMSD with the X-ray structure, but is also clearly reflected in most of the WHAT IF quality indicators.

Another indicator for the accuracy of the NMR structures is provided by independent validation against known dipolar couplings (Cornilescu et al.,

*Table 4.* Q-factors for dipolar couplings of Ubiquitin structures

| Dipolar coupling | | Unrefined | ARIA[a] | CHARMM22[b] | X-ray 1UBQ |
|---|---|---|---|---|---|
| $C^{\alpha}$-$C^{\beta}$ | (1) | **0.37 ± 0.02** | 0.38 ± 0.03 | **0.37 ± 0.05** | 0.23 |
| | (2) | – | – | – | – |
| $C^{\alpha}$ -$C'$ | (1) | 0.29 ± 0.02 | 0.30 ± 0.02 | **0.23 ± 0.02** | 0.14 |
| | (2) | 0.23 ± 0.02 | 0.25 ± 0.02 | **0.21 ± 0.03** | 0.19 |
| N- $C'$ | (1) | 0.29 ± 0.04 | 0.30 ± 0.04 | **0.21 ± 0.02** | 0.17 |
| | (2) | 0.36 ± 0.04 | 0.36 ± 0.07 | **0.25 ± 0.04** | 0.15 |
| $H^{\alpha}$-$C^{\alpha}$ | (1) | **0.26 ± 0.03** | 0.28 ± 0.04 | **0.26 ± 0.04** | 0.21 |
| | (2) | **0.28 ± 0.04** | 0.31 ± 0.04 | 0.32 ± 0.05 | 0.25 |
| $H^{N}$-$C'$ | (1) | **0.28 ± 0.02** | 0.31 ± 0.04 | **0.28 ± 0.02** | 0.26 |
| | (2) | 0.39 ± 0.02 | 0.44 ± 0.03 | **0.32 ± 0.02** | 0.29 |
| $H^{N}$-N | (1) | 0.39 ± 0.03 | 0.36 ± 0.08 | **0.30 ± 0.09** | 0.17 |
| | (2) | 0.40 ± 0.06 | 0.39 ± 0.06 | **0.28 ± 0.06** | 0.19 |

[a]Water refinement in ARIA1.0.
[b]This study CHARMM22 water refinement.
Q-factors for dipolar couplings were calculated as described in Cornilescu et al. using the program PALES. Validation was done against two sets of dipolar couplings from pdb-entry 1D3Z. All values are averages of individually fitted structures. The results for the 1.8 Å X-ray structure of Ubiquitin (1UBQ) are listed for comparison. The best values are shown in bold.

1998), which were not directly used in the calculation protocol. The agreement between calculated and measured dipolar couplings is expressed in the Q-factor, which is lower when the agreement is better. In Table 4 the results of this validation are shown for the different ensembles and compared to values obtained for the X-ray structure 1UBQ. When comparing the results of the NMR-structures, it is seen that significant improvements of the Q-factors of dipolar couplings involving the nuclei of the peptide bond ($H^{N}$, N and $C'$) are obtained after the CHARMM22 water refinement. The looser restraints on the planarity of the peptide bonds in the CHARMM22 force field are the likely reason for this effect and for the improved Ramachandran plots of the calculated structures. Further, it is clear that the X-ray structure 1UBQ still scores better than all the NMR ensembles for all six dipolar couplings used in the analysis. It will therefore be interesting to investigate what are the causes of these differences between the X-ray structure and the NMR structures, in order to further improve the methods used for structure refinement.

## Concluding remarks

Although NMR-based structure determination of biomolecules has become increasingly important over the past decades, the quality of these structures is normally still considerably less than those determined by high-resolution X-ray crystallography. An important factor determining the quality of the structures is the type of refinement of the structures at the final stage of the structure determination. Structure refinement protocols are still under continuous development, using different approaches and techniques. Here, we have applied a simple and generally applicable method to refine NMR-derived protein structures that yields structures with an overall improved quality and accuracy. Further, local geometries of these refined structures are of comparable quality as those found for high-resolution X-ray structures. Since high-resolution X-ray structures provide a widely accepted standard for comparison with newly determined biomolecular structures, we recommend that NMR structures are refined to the level of the structures described above, prior to submission to structure databases. In combination with recommendations regarding submission of structures to databases made elsewhere (Doreleijers et al., 1999a), this would be an important step forward to improving and standardizing the quality of structures at databases for biomolecules.

## Acknowledgements

We wish to thank Ad Bax for kindly providing the experimental data for Ubiquitin and useful sugges-

tions. We thank Gert Vriend, Michael Nilges and Jurgen Doreleijers for critical reading of the manuscript and discussions. Finally we thank Alexandre Bonvin for providing the original scripts for restrained Molecular Dynamics simulations in X-PLOR. The refinement protocol, implemented in X-PLOR, is freely available upon request by sending e-mail to C.Spronk@cmbi.kun.nl. The research of C. Spronk is financially supported by the Netherlands Organisation for Scientific Research (NWO).

## References

Brünger, A.T. (1992) X-PLOR. In *A system for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Cornilescu, G., Marquardt, J.L., Ottiger and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 6836–6837.

Dingley, A.J. and Grzesiek, S. (1998) *J. Am. Chem. Soc.*, **120**, 8293–8297.

Doreleijers, J.F. (1999) PhD thesis, Utrecht University, The Netherlands.

Doreleijers, J.F., Rullmann, J.A. and Kaptein, R. (1998) *J. Mol. Biol.*, **281**, 149–164.

Doreleijers, J.F., Vriend, G., Raves, M.L. and Kaptein, R. (1999a) *Proteins*, **37**, 404–416.

Doreleijers, J.F., Raves, M.L., Rullmann, T. and Kaptein, R. (1999b) *J. Biomol. NMR*, **14**, 123–132.

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

Jelsch, C., Teeter, M.M., Lamzin, V., Pichon-Pesme, V., Blessing, R.H. and Lecomte, C. (2000) *Proc. Natl. Acad. Sci USA*, **97**, 3171–3176.

Jorgenson, W.L., Chandrasekhan, J., Madura, J., Impley, R.W. and Klein, M.L. (1983) *J. Chem. Phys.*, **79**, 926–935.

Kuszewski, J. and Clore, G.M. (2000) *J. Magn. Reson.*, **146**, 249–254.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Cryst.*, **26**, 283–291.

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

Linge, J.P. and Nilges, M. (1999) *J. Biomol. NMR*, **13**, 51–59.

MacKerell, Jr. et al., and Karplus, M. (1992) *FASEB J.*, **6**, A143.

Nilges, M., Macias, M.J., O'Donoghue, S.I. and Oschkinat, H. (1997) *J. Mol. Biol.*, **269**, 408–422.

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

Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C. (1977) *J. Comp. Phys.*, **23**, 327–341.

Spronk, C.A.E.M., Bonvin, A.M.J.J., Radha, P.K., Melacini, G., Boelens, R. and Kaptein, R. (1999) *Structure*, **7**, 1483–1492.

Tjandra, N. and Bax, A. (1997) *J. Magn. Reson.*, **124**, 512–515.

Vijay-Kumar, S., Bugg, C.E., Cook, W.J. (1987) *J. Mol. Biol.*, **194**, 531–544.

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.

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

Zweckstetter, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 3791–3792.