PROTEINS: Structure, Function, and Bioinformatics 55:483–486 (2004)

# SHORT COMMUNICATION

# DRESS: a Database of REfined Solution NMR Structures

**Sander B. Nabuurs,[1] Aart J. Nederveen,[2] Wim Vranken,[3] Jurgen F. Doreleijers,[4] Alexandre M.J.J. Bonvin,[2] Geerten W. Vuister,[5] Gert Vriend,[1] and Christian A.E.M. Spronk[1]***

[1]*Center for Molecular and Biomolecular Informatics, University of Nijmegen, Nijmegen, The Netherlands*
[2]*Department of NMR spectroscopy, University of Utrecht, Utrecht, The Netherlands*
[3]*EMBL-EBI, The European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom*
[4]*CESG/BioMagResBank, University of Wisconsin, Madison, Wisconsin*
[5]*Department of Biophysical Chemistry, University of Nijmegen, Nijmegen, The Netherlands*

***ABSTRACT*** **Several studies have shown that biomolecular NMR structures are often of lower quality when compared to crystal structures, and consequently they are often excluded from structural analyses. We present a publicly available database of re-refined NMR structures, exhibiting significantly improved quality. This database (available at http://www.cmbi.kun.nl/dress/) presents a uniformly refined and validated set of structural models that improves the value of these NMR structures as input for experimental and theoretical studies in many fields of research. Proteins 2004;55:483–486.**
**© 2004 Wiley-Liss, Inc.**

**Key words: structure refinement; protein structures; nuclear magnetic resonance (NMR); molecular dynamics; structure validation; explicit solvent**

## INTRODUCTION

In a recent survey of the quality of biomolecular structure models,[1] Roman Laskowski, author of the PROCHECK structure validation programs,[2,3] states the rule of thumb for selecting NMR models: "*Historically, the rule of thumb for selecting NMR structures for inclusion in structural analyses has been the simple one of excluding them altogether!*". Although this rule is referred to as an early prejudice, evidence has accumulated in literature over the past years supporting the notion that NMR-derived models are of lower quality than high-resolution structures derived by X-ray crystallography.[4,5] This finding is most evident in the analysis of non-optimal local geometry, packing quality and electrostatics of these structures.[4,6] Furthermore, NMR structures were recently shown to exhibit higher internal strain, and to diverge more rapidly during molecular dynamics simulations when compared to crystal structures.[7–9] As a result of this, NMR structures are often considered less useful in studies regarding biomolecular structures. However, in representative subsets of the Protein Data Bank (PDB),[10] culled at 20% sequence identity, around one in five structures stems from NMR experiments,[11] clearly illustrating that NMR-based structural models provide unique and important information not available from X-ray studies.

Many of the structural problems present in typical NMR structures appear to originate from the applied structure refinement protocol. For the sake of speed the nonbonded interactions are normally severely simplified, resulting in an unrealistic treatment of the electrostatic and van der Waals interactions that can lead to the artifacts described above. In the mid-nineties several papers appeared, demonstrating that biomolecular NMR structures can be significantly improved by the inclusion of explicit solvent molecules in a restrained molecular dynamics refinement.[12–14] The merits of different explicit water refinement approaches have been extensively discussed in literature since then,[5,15–17] with the results making their way into ARIA, one of the commonly used NMR structure calculation packages.[18] Unfortunately, not all newly deposited structures are refined using the latest protocols and several structures that are already present in the PDB are poorly refined according to today's standards.

For this reason, we have started an effort to re-refine the NMR structures deposited in the PDB in explicit solvent and make these refined structures available to the public. Here, we present the results for the first 100 structures that have been processed and illustrate the clear overall improvement of the structures in the DRESS database as compared to the currently available structures.

## MATERIAL AND METHODS

The one hundred structures presented here vary in size between 20 and 370 amino acids and were deposited

Grant sponsor: the European Community program NMRQUAL; Grant number: QLG2-CT-2000-0313; Grant sponsor: the Netherlands Foundation for Chemical Research (NWO/CW); Grant sponsor: the National Institute of General Medical Sciences, National Institutes of Health; Grant sponsor: P50-GM64598.

*Correspondence to: C. Spronk, Center for Molecular and Biomolecular Informatics, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands. E-mail: C.Spronk@cmbi.kun.nl

Received 8 January 2004; Accepted 19 January 2004

Published online 1 April 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20118

© 2004 WILEY-LISS, INC.

between 1993 and 2002, providing a representative subset of the NMR structures presently available from the PDB. For several of the structural models we observed a discrepancy between IUPAC nomenclature and the nomenclature as deposited in the PDB, possibly originating from nomenclature conversion steps. As nomenclature errors can result in an artificially high number of violations, all proton positions in the original structural models were corrected by a short minimization of the proton positions in X-PLOR[19] preceding the refinement procedure and violation analyses.

Indispensable to refine NMR structures are the experimental restraints used to calculate them. These are available from the PDB for about 60% of all deposited NMR structures, and were recently converted at the BioMagResBank into one unified format.[20] Using the FormatConverter, developed as part of the Collaborative Computing Project for the NMR Community (CCPN),[21] the one hundred experimental restraint sets were converted to the X-PLOR restraint format. The converted restraint sets were used to refine the corresponding ensembles, as deposited in the PDB, in a short restrained molecular dynamics simulation in explicit solvent, as described previously.[17] In short, the structure refinement consisted of the following steps: structures were immersed in a 7.0 Å shell of water molecules and energy-minimized. Subsequently, the systems were slowly heated from 100 to 500 K using 1000 steps of molecular dynamics, while applying harmonic position restraints on the protein that were slowly phased out during the heating stage. Refinement of the structures was then performed by 2000 steps of molecular dynamics at 500 K, followed by 4000 steps of slow cooling to 25 K and 200 steps of energy minimization. Scaling of the force constants for bonds, angles, impropers and omega angles during the cooling stage was slightly modified with respect to the original protocol to allow for the naturally occurring variation in these parameters as described by Engh and Huber.[22] Throughout the protocol the PARALLHDG 5.3 force field[17] was used with a full nonbonded representation including Lennard-Jones Van der Waals and electrostatic interactions from the OPLS force field.

The quality of the structure ensembles before and after refinement was judged both by their agreement with the experimental restraints and the quality scores as determined by the structure analysis programs PROCHECK[2] and WHAT CHECK.[23] Averages and standard deviations were calculated from the checks of the individual members of each ensemble.

## RESULTS AND DISCUSSION

The quality of protein structures with regard to a given quality criterion can be represented by a so-called Z-score, which is the deviation of that quality indicator from a database-derived average value, in units of the standard deviation of that database derived average. The database derived average, in our case derived from high-resolution crystal structures, will by definition have a Z-score of 0. Figure 1(A) through (C) shows the consistent improvement

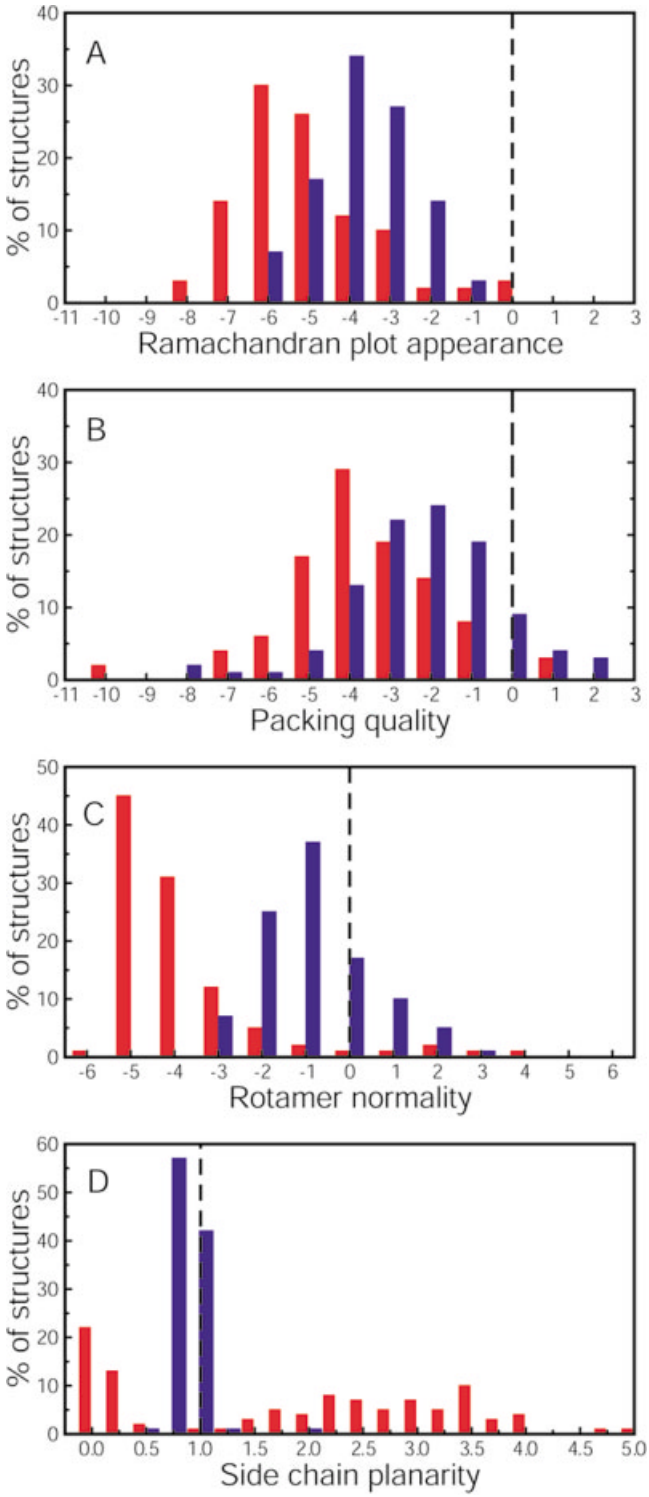


Fig. 1. Distribution of quality scores before and after refinement. Quality Z-scores distributions of the hundred selected NMR structures are presented for (**A**) Ramachandran plot appearance, (**B**) 2nd generation packing quality (all backbone and side chain contacts), and (**C**) $\chi_1$-$\chi_2$ rotamer normality. (**D**) RMS Z-score distribution for heavy atom side-chain planarity. For a discussion on Z-scores and how to interpret them, see Linge et al.[17] All distributions are shown before (red) and after (blue) refinement, values for the WHAT CHECK reference database are indicated with a dashed line.

**TABLE I. Average Quality Indicators Before and After Refinement for 100 NMR Structures**

| | Before refinement | After refinement |
|---|---|---|
| Rms violation of all distance restraints | 0.089 ± 0.083 | 0.029 ± 0.015 |
| Rms violation of all dihedral angle restraints | 4.44 ± 6.400 | 0.457 ± 0.367 |
| Number of consistently violating restraints[a] | 8 ± 17 | 0 ± 1 |
| Interatomic bumps per 100 residues[b] | 86 ± 61 | 11 ± 9 |
| PROCHECK results | | |
| Most favored | 65.0 ± 17.9 | 74.9 ± 15.7 |
| Allowed | 27.0 ± 12.0 | 19.4 ± 10.2 |
| Generously allowed | 4.7 ± 4.0 | 2.2 ± 2.0 |
| Disallowed | 3.3 ± 13.8 | 3.5 ± 13.8 |
| WHAT CHECK structure Z-scores[c] | | |
| Ramachandran plot appearance | −5.1 ± 1.7 | −3.6 ± 1.2 |
| 2nd generation packing quality[d] | −3.7 ± 1.8 | −2.2 ± 1.8 |
| $\chi1$-$\chi2$ rotamer normality | −3.8 ± 1.8 | −0.8 ± 1.2 |
| Backbone conformation | −4.7 ± 3.7 | −3.9 ± 2.7 |

[a]Defined as those restraints that violate (> 0.5 Å) in more than 50% of the members of a structure ensemble of at least ten structures.
[b]According to WHAT CHECK.
[c]For a detailed explanation of the different WHAT CHECK quality scores, see http://www.cmbi.kun.nl/gv/pdbreport/checkhelp/.
[d]Packing quality for all backbone and side chain contacts.

of both the overall and local quality Z-score distributions. The overall quality improvement is best illustrated by the Ramachandran plot and packing quality scores, which have been shown to be independent indicators for structure quality,[24,25] the local quality improvement by $\chi_1$-$\chi_2$ rotamer normality. The improvement of the Ramachandran plot quality can be mostly attributed to the newly parameterized backbone dihedral angles[17] and the larger freedom that we allow for the peptide plane angle ω, which is often too tightly restrained in NMR structures. The packing quality, the number of interatomic bumps and the hydrogen bonding network (data not shown) all benefit strongly from the better description of the electrostatic and van der Waals interactions in the applied refinement protocol.

The RMS Z-score distribution for the heavy atom side-chain planarity is shown in Figure 1(D). RMS Z-scores smaller than 1.0 indicate a tighter distribution and those larger than 1.0 a broader distribution of values than in the WHAT CHECK reference database. The distribution before refinement shows that side-chain planarity is often too tightly or too loosely restrained, with a local minimum there where X-ray structures are commonly found. After refinement nearly all structures have planarity RMS Z-scores comparable to high-resolution X-ray structures. Additionally, the re-refined structures show much better agreement with presently acceptable simple stereochemistry parameters such as bond lengths and bond angles (data not shown).

The agreement with the original experimental data and other parameters pertaining to structural quality are presented in Table I. Averaged over the set of 100 structures, the refinement in explicit water not only improves virtually all validation criteria, but, very importantly, also the agreement with the experimental restraints. The refined structures clearly illustrate that it is possible to bring NMR-derived structures closer to physical reality while at the same time improving their fit to the experimental data, as discussed before.[5,16,17] Markedly, the number of consistent violations is virtually reduced to zero after refinement. It is fair to note here that the high number of consistent violations in the input structures could partly be the result of nomenclature errors of untraceable origin.

In the analysis of Z-scores it is common practice to consider structural models with a Z-score bigger than +4 or smaller than −4 as outliers. In contrast to the original structural models, the refined structures now fall, on average, within the specified range. However, parameters related to overall structural quality still do not reach the level of high resolution X-ray structures. It should be kept in mind that this could be an indication that there are possibly additional problems with these structures that cannot be resolved with solely a refinement in explicit solvent. Despite this fact, we feel that the re-refined structures provide better starting points for further structural studies such as protein–protein and protein–ligand interactions.

## Conclusion

Our results show that the quality of the re-refined structures has significantly improved, both in terms of the agreement with the experimental input data and the quality as judged by the PROCHECK and WHAT CHECK programs. Therefore, we intend to expand the current dataset to comprise all NMR-derived structural models for which a refinement is feasible and make the refinement methods, the results, restraint analyses, and reports regarding the structural quality of each structure, publicly available at http://www.cmbi.kun.nl/dress/ for use in further research.

## REFERENCES


1. Laskowski RA. Structural quality assurance. Methods Biochem Anal 2003;44:273–303.
2. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–291.
3. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.
4. Doreleijers JF, Rullmann JA, Kaptein R. Quality assessment of NMR structures: a statistical survey. J Mol Biol 1998;281:149–164.
5. 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.
6. Doreleijers JF, Vriend G, Raves ML, Kaptein R. Validation of nuclear magnetic resonance structures of proteins and nucleic acids: hydrogen geometry and nomenclature. Proteins 1999;37:404–416.
7. Fan H, Mark AE. Relative stability of protein structures determined by X-ray crystallography or NMR spectroscopy: a molecular dynamics simulation study. Proteins 2003;53:111–120.
8. Maiorov V, Abagyan R. Energy strain in three-dimensional protein structures. Fold Des 1998;3:259–269.
9. Lee MR, Kollman PA. Free-energy calculations highlight differences in accuracy between X-ray and NMR structures and add value to protein structure prediction. Structure (Camb) 2001;9:905–916.
10. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
11. Wang G, Dunbrack RL, Jr. PISCES: a protein sequence culling server. Bioinformatics 2003;19:1589–1591.
12. Billeter M, Qian YQ, Otting G, Muller M, Gehring W, Wuthrich K. Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. J Mol Biol 1993;234:1084–1093.
13. Prompers JJ, Folmer RH, Nilges M, Folkers PJ, Konings RN, Hilbers CW. Refined solution structure of the Tyr41 → His mutant of the M13 gene V protein. A comparison with the crystal structure. Eur J Biochem 1995;232:506–514.
14. Kordel J, Pearlman DA, Chazin WJ. Protein solution structure calculations in solution: solvated molecular dynamics refinement of calbindin D9k. J Biomol NMR 1997;10:231–243.
15. 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.
16. 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.
17. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. Proteins 2003;50:496–506.
18. Linge JP, Habeck M, Rieping W, Nilges M. ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics 2003;19:315–316.
19. Brünger AT. X-PLOR, version 3.1. A system for X-ray crystallography and NMR. New Haven, CT: Yale University Press; 1992.
20. Doreleijers JF, Mading S, Maziuk D, Sojourner K, Yin L, Zhu J, Markley JL, Ulrich EL. BioMagResBank database with sets of experimental NMR constraints corresponding to the structures of over 1400 biomolecules deposited in the Protein Data Bank. J Biomol NMR 2003;26:139–146.
21. Fogh R, Ionides J, Ulrich E, Boucher W, Vranken W, Linge JP, Habeck M, Rieping W, Bhat TN, Westbrook J, Henrick K, Gilliland G, Berman H, Thornton J, Nilges M, Markley J, Laue E. The CCPN project: an interim report on a data model for the NMR community. Nat Struct Biol 2002;9:416–418.
22. Engh RA, Huber R. Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr A 1991;47:392–400.
23. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature 1996;381:272.
24. Hooft RWW, Sander C, Vriend G. Objectively judging the quality of a protein structure from a Ramachandran plot. Comput Appl Biosci 1997;13:425–430.
25. Vriend G, Sander C. Quality control of protein models: directional atomic contact analysis. J Appl Cryst 1993;26:47–60.