Journal of Biomolecular NMR, **21:** 373–375, 2001.
KLUWER/ESCOM

Improving the efficiency of the Gaussian conformational database potential for the refinement of protein and nucleic acid structures

Philipp Neudecker*, Heinrich Sticht & Paul Rösch
Lehrstuhl für Biopolymere, Universität Bayreuth, D-95440 Bayreuth, Germany

Received 7 September 2001; Accepted 19 October 2001

Key words: conformational databases, major cherry allergen, NMR spectroscopy, potential of mean force, protein and nucleic acid structure determination, refinement, X-ray crystallography

Abbreviations: CPU, central processing unit; IEEE, Institute of Electrical and Electronics Engineers, Inc.; NIH, National Institutes of Health; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; RMSD, root mean square deviation

To improve the often poor stereochemical properties of protein or nucleic acid structures determined by NMR spectroscopy or X-ray crystallography, Kuszewski et al. (1996, 1997) proposed to bias the notoriously difficult parametrization of nonbonded interactions towards energetically more favorable and hence more abundant local conformations by using a potential of mean force for torsion angle conformations based on high-resolution crystal structure databases in the refinement process. This conformational database potential consists of several potential hypersurfaces each representing the distribution of a particular set of torsion angles in the database, e.g., of the Φ/Ψ angle combinations of the non-glycine residues (two-dimensional hypersurface). Kuszewski and Clore (2000) recently noted that the discontinuities of their discretely sampled original implementation can seriously affect convergence towards energetically favorable local conformations and sampling of the accessible conformational space, especially if the number of experimental restraints is very small, and proposed to fit a set of N Gaussian functions to every hypersurface, with N ranging from 18 to 60. Unfortunately, while yielding an analytical representation for the conformational database potential with continuous partial derivatives, this effectively multiplies the number of hypersurfaces that have to be evaluated by N, leading to a severalfold increase in CPU time. Gaussians decay very rapidly; their values 5.0, 7.5, and 10.0 standard deviations away from the center, e.g., are 3.7×10^{-6}, 6.1×10^{-13}, and 1.9×10^{-22}, respectively. As a consequence, the respective contributions of many of the Gaussians are completely negligible, and a considerable amount of CPU time can be saved by preventing their evaluation if the distance of the actual conformation from the center of the Gaussian exceeds a particular cutoff. This prompted us to modify the NIH version (Kuszewski and Clore, 2000) of X-PLOR 3.840 (Brünger, 1992) accordingly. By default, only Gaussians whose centers are less than 10.0 standard deviations away from the actual conformation are considered now, but the user can also specify his own set of cutoff values for the individual hypersurface classes and even change these cutoff values at any time in the course of the structure calculation.

As an example to investigate the effects of this modification, the calculation of the solution structure of the 159 amino acid major cherry allergen Pru av 1 was repeated with several different cutoff values. The 22 structures deposited with the PDB (access code: 1E09) are the result of the calculation of 60 structures with X-PLOR 3.851 (Brünger, 1992) based on a three-stage simulated annealing protocol using 2438 restraints derived from NMR experiments (Neudecker et al., 2001) supported by the original conformational database potential (Kuszewski et al., 1996) with a force constant factor of 3.0, which requires a total of 168.8 h of CPU time on a Sun Ultra Enterprise 450 server with four 400 MHz Ultra Sparc II processors, 13.2 h or 8% more than the calculation of 60 structures

*To whom correspondence should be addressed. E-mail: philipp.neudecker@uni-bayreuth.de

Table 1. Comparison of the structure calculations

X-PLOR version	3.851	3.851	3.840	3.840	3.840	3.840
Conformational database potential	none	discrete[a]	Gaussian[b]	Gaussian[b]	Gaussian[b]	Gaussian[b]
Force constant factor	n/a	3.0	1.0	1.0	1.0	1.0
Cutoff/standard deviations	n/a	n/a	none	10.0	7.5	5.0
No. of structures calculated	60	60	60	60	60	60
No. of structures accepted	24	22	22	22	25	29
Total energy[c,d]/kcal/mol	233 ± 9	244 ± 7	240 ± 8	240 ± 8	244 ± 8	241 ± 6
Backbone RMSD[c,e]/Å	0.61 ± 0.10	0.60 ± 0.09	0.59 ± 0.10	0.59 ± 0.10	0.62 ± 0.12	0.61 ± 0.10
Heavy atom RMSD[c,e]/Å	1.00 ± 0.12	0.93 ± 0.09	0.93 ± 0.11	0.93 ± 0.11	0.93 ± 0.12	0.95 ± 0.09
Residues in most favored regions of Ramachandran plot[f]/%	76.6	82.4	82.6	82.6	81.7	82.6
Residues in additional allowed regions of Ramachandran plot[f]/%	21.1	16.6	16.1	16.1	16.9	15.9
Residues in generously allowed regions of Ramachandran plot[f]/%	2.0	1.0	1.2	1.2	1.3	1.5
Residues in disallowed regions of Ramachandran plot[f]/%	0.2	0.0	0.1	0.1	0.1	0.1
Packing Z-score[c,g]	-1.49 ± 0.27	-0.54 ± 0.27	-0.82 ± 0.30	-0.82 ± 0.30	-0.79 ± 0.28	-0.80 ± 0.24
Backbone RMSD from PDB structures[h]/Å	0.65	0.00	0.41	0.41	0.43	0.35
Heavy atom RMSD from PDB structures[h]/Å	0.85	0.00	0.56	0.56	0.58	0.48
Backbone RMSD from no-cutoff structures[i]/Å	0.48	0.41	0.00	0.00	0.17	0.17
Heavy atom RMSD from no-cutoff structures[i]/Å	0.72	0.56	0.00	0.00	0.27	0.26
Total CPU time/h	155.6	168.8	925.3	714.6	567.1	412.6

[a] Kuszewski et al., 1996.
[b] Kuszewski and Clore, 2000.
[c] Average value over the accepted structures in the form average value ± standard deviation.
[d] Values of the target function excluding conformational database potential.
[e] Atomic RMSDs from the average structure.
[f] Determined with PROCHECK 3.4 (Laskowski et al., 1993).
[g] Determined with WHAT_CHECK, WHAT IF 19970704-1848 (Vriend and Sander, 1993).
[h] Atomic RMSD of the average structure from the average structure of the set of 22 structures calculated with the original conformational database potential (Kuszewski et al., 1996) and deposited with the PDB (access code: 1E09).
[i] Atomic RMSD of the average structure from the average structure of the set of 22 structures calculated with the Gaussian conformational database potential (Kuszewski and Clore, 2000) without cutoff.

without conformational database potential (Table 1). In contrast, the Gaussian conformational database potential (Kuszewski and Clore, 2000) with a force constant factor of 1.0, which is approximately equivalent due to different normalization, increases the CPU time by 769.7 h or 495%. With cutoff values of 10.0, 7.5, and 5.0 standard deviations, e. g., the increase in CPU time is cut down to 559.0 h or 359%, 411.5 h or 264%, and 257.0 h or 165%, respectively. Table 1 shows that every conformational database potential improves the stereochemical properties of the structures considerably, which is reflected by significantly higher proportions of residues in favorable regions of the Ramachandran plot and a marked improvement of the packing quality according to WHAT_CHECK (Vriend and Sander, 1993). The atomic RMSDs between the average structures are all below 0.65 Å for the backbone and 0.85 Å for all heavy atoms, which excludes that any serious artifacts have been introduced by either representation of the conformational database potential. In contrast, the differences in final target function values, precision, stereochemical properties, and atomic coordinates of the average structures among the structure calculations with conformational database potential are absolutely marginal, regardless of the representation and cutoff value. Interestingly, the 60 atom coordinate files resulting from the calculation with the default cutoff value of 10.0 standard deviations are identical to those resulting from the calculation without cutoff, i.e., more than 25% of the increase in CPU time can be saved without even influencing the results of the structure calculation. The choice of an appropriate cutoff value is a compromise between computational speed and quality of the final structures, but one should bear in mind that extremely low cutoff values may give rise to new discontinuities. Still, it is obvious that the error for the hypersurfaces caused by a reasonable cutoff is much smaller than the errors that arise from the inaccuracies of the crystal structures in the database, the discrete conformational sampling, and the approximation by a finite number of Gaussians. In addition, a finite cutoff adds to the stability of the calculation by preventing underflow exceptions, because the value of a Gaussian whose center is more than 37.64 standard deviations away from the actual conformation is smaller than 2.225×10^{-308}, the smallest positive normal number that can be encoded in double-precision floating point format complying with IEEE Standard 754–1985.

Acknowledgements

We thank R. Hofmann for expert technical assistance, and J. Kuszewski and G.M. Clore for making available the NIH version of X-PLOR, whose most recent releases include a modification similar to ours (http://nmr.cit.nih.gov). The modified source code files are available on request. P.N. gratefully acknowledges a fellowship from the Fonds des Verbandes der Chemischen Industrie in cooperation with the Bundesministerium für Bildung und Forschung (BMBF). This work was supported by grants from the Deutsche Forschungsgemeinschaft (Ro617/11-1) and the BMBF.

References

Brünger, A.T. (1992) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA.

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

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

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

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

Neudecker, P., Schweimer, K., Nerkamp, J., Scheurer, S., Vieths, S., Sticht, H. and Rösch, P. (2001) *J. Biol. Chem.*, **276**, 22756–22763.

Vriend, G. and Sander, C. (1993) *J. Appl. Cryst.*, **26**, 47–60.