Journal of Biomolecular NMR, **23:** 221–225, 2002.
KLUWER/ESCOM

Reweighted atomic densities to represent ensembles of NMR structures*

Charles D. Schwieters[a,**] & G. Marius Clore[b,**]
[a]*Computational Bioscience and Engineering Laboratory, Center for Information Technology, National Institutes of Health, Building 12A, Bethesda, MD 20892-5624, U.S.A.;* [b]*Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 5, Bethesda, MD 20892-0510, U.S.A.*

Received 1 May 2002; Accepted 31 May 2002

Abstract

A reweighted atomic probability density is introduced as a means of representing ensembles of NMR structures in a simple, concise and informative manner. This density is shown to give a better visual representation of molecular structure information than an unweighted density, and should provide a useful interactive graphics tool during the course of iterative NMR structure refinement. The approach is illustrated using several examples.

Unlike other techniques for biomolecular structure determination, NMR methods do not typically yield a unique configuration. Heterogeneity in NMR structures can be caused by a number of factors: (1) Underdetermined regions of the structure due to insufficient number or quality of experimental restraints; (2) actual solution-phase heterogeneity due to the presence of multiple conformations; and (3) thermal motion (Wüthrich, 1986; Clore and Gronenborn, 1989, 1991). NMR results can often be well-represented by the restrained regularized mean coordinates together with the average rms difference between the individual structures and the mean coordinates (placed in the B-factor column) (Clore et al., 1986; Nilges et al., 1988). However, this representation fails in cases of multiple conformers (i.e. structurally discrete clustering), which is particularly relevant for surface exposed side chains. In addition, the atomic rms to the mean gives no indication as to the nature of the distribution. When a single structure provides a poor representation for the ensemble of NMR structures (Sutcliffe, 1993; Doreleijers et al., 1999; Howe, 2001), multiple structures are submitted to the structure database.

However, displaying many structures is inconvenient in a practical sense, and is potentially misleading. Inconvenient in that it is awkward to load and manipulate many structures in the popular visualization programs. Misleading if too many or too few structures are loaded, in that representative configurations may be missed or obscured, respectively. Finally, a certain visual distraction can result when many structures are presented.

In this note we introduce a special atomic density which is appropriate for visualizing ensembles of NMR structures. In previous work (DeLano and Brünger, 1994; Bonvin and Brünger, 1995) density maps were generated from ensembles of structures, essentially by adding of the electron density due to all of the atoms of all the structures. We have found that a more satisfying representation of an ensemble can be obtained if the contribution due to each atom is rescaled depending on the distribution of its positions in the ensemble.

An evenly-weighted atom density at point $\mathbf{q}$ can be defined as

$$\rho_{\text{even}}(\mathbf{q}) = \frac{1}{N_s} \sum_{k=1}^{N_s} \sum_{l=1}^{N_a} \rho_a(\mathbf{q} - \mathbf{q}_{kl}),$$

where N_s is the number of structures in the ensemble, N_a the number of atoms in each structure, q_{kl} the position of atom l in structure k, and $\rho_a(\mathbf{q})$ a normalized distribution for a single atom. For computational convenience, we choose this density to be a

**To whom correspondence should be addressed. E-mail: Charles.Schwieters@nih.gov; mariusc@intra.niddk.nih.gov

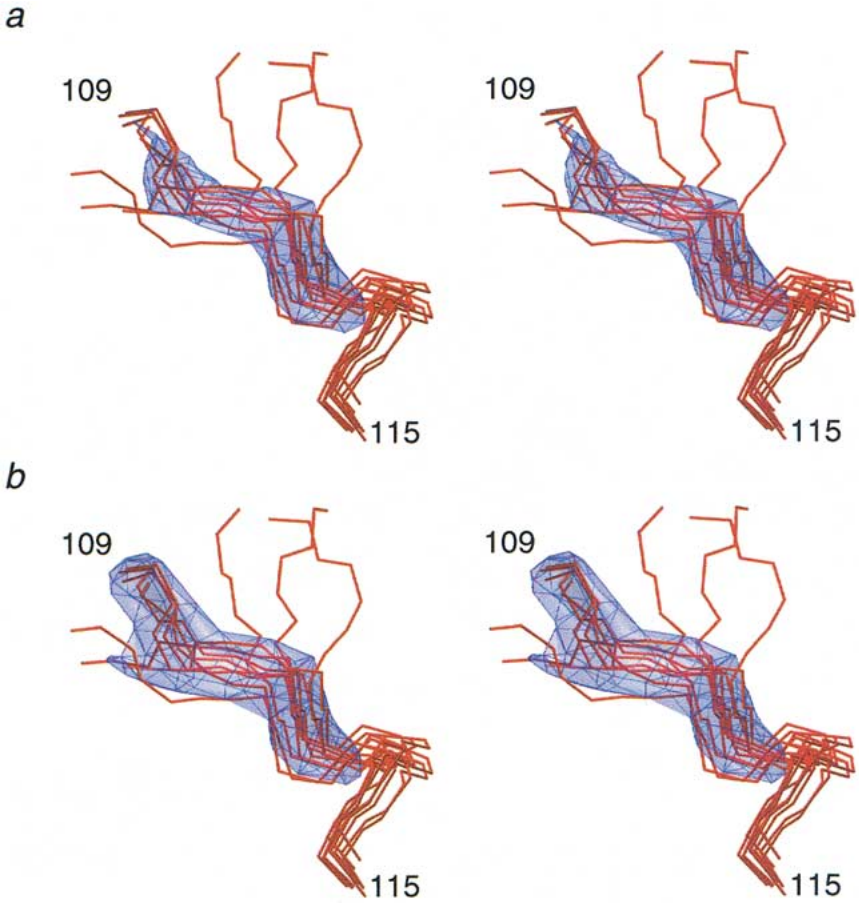

Figure 1. Eleven simulated annealing structures of the backbone atoms of the N-terminus (residues 109–115) of the LAP-2C domain (residues 111–152) of the constant region of the nuclear envelope protein LAP2 depicted in red. Transparent blue isosurfaces representing 40% of the maximum value were generated from residues 109–112. (a) Depicts $\rho_{even}(\mathbf{q})$ and (b) displays $\rho_{amp}(\mathbf{q})$.

piecewise function, whose nonzero region is a quartic polynomial:

$$\rho_a(\mathbf{q}) = \begin{cases} \dfrac{105}{32\pi a^7}(|\mathbf{q}|^2 - a^2)^2 & |\mathbf{q}| < a \\ 0 & \text{otherwise} \end{cases},$$

where $a[1 - \sqrt{2}/2]^{1/2}$ is the distance at which the density is one-half its maximum value. In the current study, this distance was chosen to be constant for all atoms, with a value of 1 Å.

Figure 1 depicts stereo images of the N-terminus (residues 109–115) of an ensemble of eleven simulated annealing structures calculated from NMR data for the LAP2-C domain of the constant region of the nuclear envelope protein LAP2 (Cai et al., 2000). The structures were best-fitted with respect to the backbone atoms of residues 111–152, and residues 109–110 are disordered.

In Figure 1a the ensemble of structures is overlayed with an isosurface of $\rho_{even}(\mathbf{q})$ generated from the backbone atoms of residues 109–112 drawn at 40% of the maximum of the density. The structural heterogeneity at the N-terminus causes the value of ρ_{even} to be reduced such that the isosurface is truncated relative to the structures from which it was constructed, and hence is not a good visual representation of the structures. By choosing a smaller value for the isosurface, more of the terminus can be enclosed, but this comes with an undesirable bloating of the conserved backbone region.

We would prefer that the probability density isosurface better represent the structures when inspected

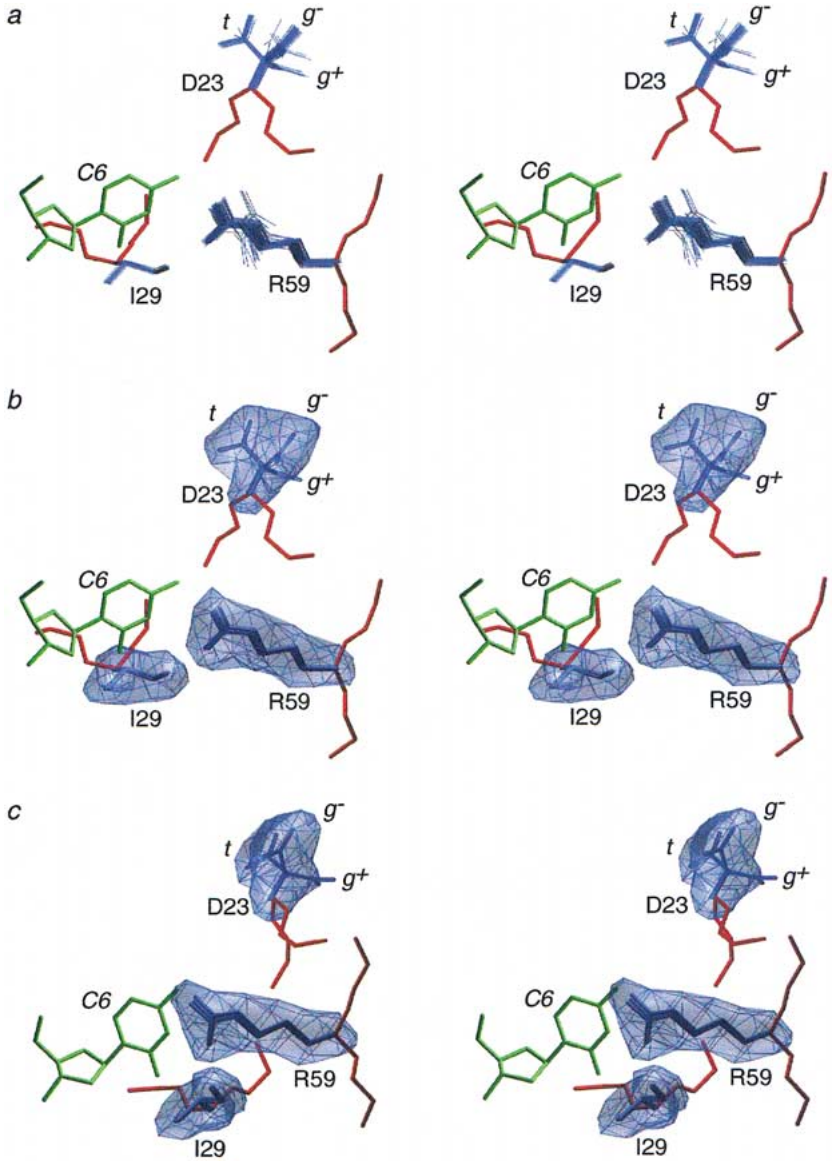

Figure 2. Stereoviews illustrating various representations of the side chains of Asp23, Ile29 and Arg59 of the hnRNPK KH3-ssDNA complex. (a) Superposition of the sidechains of 100 simulated annealing structures (blue). (b) and (c) Isosurface of the reweighted atomic density map for the three side chains drawn at a value of 20% of maximum; within the map, the coordinates of the sidechains of three representative structures are displayed in blue. The view shown in (b) is identical to that in (a). The coordinates of the protein backbone and the C6 nucleotide of the restrained regularized mean structure are shown in red and green, respectively.

visually. Towards this goal we define a density in which the contribution to each atom is reweighted

$$\rho_{\text{amp}}(\mathbf{q}) = \frac{1}{N_s} \sum_{l=1}^{N_a} w_l \sum_{k=1}^{N_s} \rho_a(\mathbf{q} - \mathbf{q}_{kl}),$$

where

$$1/w_l = \max \sum_k \rho_a(\mathbf{q} - \mathbf{q}_{kl}).$$

In this density, for a given atom, the relative probability of a given position is preserved, but the contribution of one atom relative to another is reweighted such that the maxima of each contribution is constant. [Note that the overall normalization is not preserved: $\int \rho_{\text{amp}}(\mathbf{q}) d\mathbf{q} \neq N_a$. This property is not reflected in the relative density maps considered in this paper.] A 40% of maximum amplitude isosurface of $\rho_{\text{amp}}(\mathbf{q})$

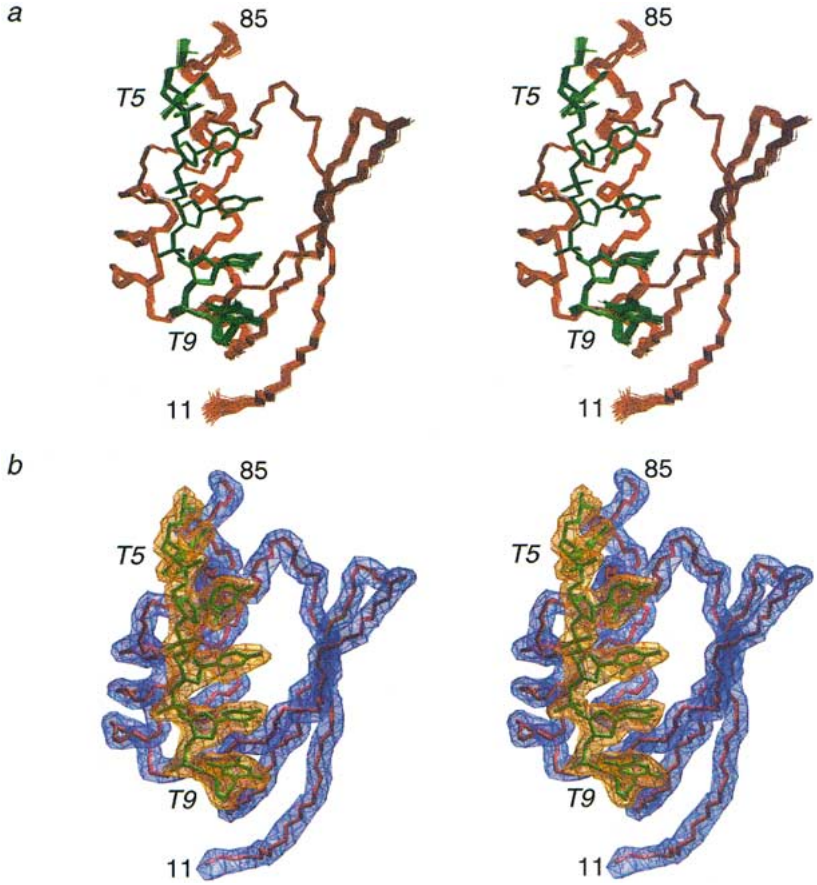

Figure 3. Stereoviews illustrating the backbone and DNA of the hnRNPK KH3-ssDNA complex. (a) Superposition of the ensemble of 100 simulated annealing structures with the protein backbone in red and the DNA heavy atoms in green. (b) Isosurface of the reweighted atomic density map for the protein backbone (blue) and DNA heavy atoms (orange) drawn at a value of 30% maximum amplitude. The restrained regularized mean coordinates for the protein backbone and the DNA are shown in red and green, respectively, within the density map. The structures are taken from (Braddock et al., 2002).

for the eleven-member ensemble of the N-terminus of LAP2-C is shown in Figure 1b. It is seen that this isosurface better represents the most likely structures. The surface appears to be a space-filled representation of the average structure, but it becomes spread out (or even bifurcated) when there is significant structural variation within the ensemble. These probability maps can be composed of as many structures as desired. Further, one might use this map for direct structural refinement, and while ρ_{even} would cause compression along the backbone, biasing towards a reduced radius of gyration, this effect would be minimal if ρ_{amp} were used.

The capability to create and display reweighted atomic density maps of ensembles of NMR structures has been incorporated into the VMD-XPLOR visualization package (Schwieters and Clore, 2001). VMD-XPLOR is based on the VMD visualization program (Humphrey et al., 1996) and is available on the world-wide web using the URL http://vmd-xplor.cit.nih.gov/

Examples

In this section we show examples of the use of atomic density maps to illustrate various aspects of the NMR structure of the hnRNPK KH3-ssDNA complex using data taken from Braddock et al. (2002). The structures were calculated on the basis of 1986 experimental restraints, including 73 intermolecular inter-proton distance restraints and 287 residual dipolar couplings, and the target function included a torsion angle database potential of mean force (Clore and Kuszweski, 2002). The coordinate precision (residues 11–85 of the protein and 5–9 of the ssDNA) is 0.23 ± 0.05 Å

for the protein backbone plus DNA heavy atoms, and 0.63 ± 0.05 Å for the protein heavy atoms plus DNA heavy atoms.

In Figure 2, stereoviews illustrate three representations of the side chains of residues Ile29, Arg59 and Asp23 of the hnRNPK KH3-ssDNA complex. Figure 2a depicts a superposition of 100 simulated annealing structures of the sidechains, while Figures 2b and 2c display a reduced number of structures along with isosurfaces of ρ_{amp} for the three side chains drawn at a threshold of 20% of maximum value. Ile29 exists in a single well-defined conformation (with χ_1 and χ_2 in the g^- and t rotamers, respectively). This is reflected in tightly clustered structures in Figure 2a, and a simple, compact isosurface in Figures 2b and 2c. For Arg59, the χ_1, χ_2 and χ_3 sidechain torsion angles are in the g^-, t and t conformations respectively; in the case of χ_4, 85% of the structures are in the t conformer, and 15% in the g^- conformer. The spread in the structures is evident in Figure 2a, and results in a slight broadening of the isosurfaces shown in Figures 2b and c. All three χ_1 rotamers are present for Asp23 with 56% in g^-, 34% in t and 10% in g^+; in the case of the χ_2 angle, 86% have a value of $0 \pm 20°$, and 14% around 90°. The isosurface representations in Figures 2b and c reflect this heterogeneity, but note that as the χ_1 g^+ rotamer for Asp23 is a minor confomer, the density only extends out to the C_γ atom at the displayed threshold of 20%; at a threshold of 30%, the density only extends to half-way along the C_β-C_γ bond.

Figure 3 shows two stereo representations of the protein backbone and DNA of the hnRNPK KH3-ssDNA complex. Since the structures are closely bunched, the isosurface forms tight shells about the two molecules, with some widening at the N-terminus of the protein chain where there is structural variability.

We do not anticipate that this reweighted density will replace the use of multiple structures in the visualization or analysis of ensembles of NMR structures. Rather, we view this density as a useful alternative representation in certain circumstances in which it is illuminating to replace a collection of many structures with a single isosurface. For instance, while the backbone may be well-represented by 10–20 structures, a good description of sidechain distributions often requires 50–100 structures. It is at these large numbers that the reweighted density becomes attractive. As in Figure 2, the reweighted density can represent the relative probability of a rotamer, while one can display a small number of structures which clearly show the atomic geometry.

Finally, since the probability map provides a simple and concise representation of a large ensemble of structures, we envision that it will also provide a versatile tool during the course of iterative NMR structure refinement, analogous to the use of structure databases in protein crystallography refinement (Jones et al., 1991). For example, when presented with a long side chain involved in an inter-residue interaction (intra or intermolecular), and whose direction is defined by the NOE data, one might use a generated density map to fit the side chain into its most probable conformations. This mechanism would take into account rotamer preferences, and would provide structural hints similar to those given by side chain database potentials of mean force (Clore and Kuszweski, 2002).

References

Bonvin, A.M.J.J. and Brünger, A.T. (1995) *J. Mol. Biol.*, **250**, 90–93.

Braddock, D.T., Baber, J.L., Levens, D. and Clore, G.M. (2002) *EMBO J.*, **21**, 3476–3480.

Cai, M.L., Huang, Y., Ghirlando, R., Wilson, K.L., Craigie, R. and Clore, G.M. (2001) *EMBO J.*, **20**, 4399–4407.

Clore, G.M. and Gronenborn, A.M. (1989) *CRC Crit. Rev. Biochem. Molec. Biol.*, **24**, 479–564.

Clore, G.M. and Gronenborn, A.M. (1991) *Science*, **252**, 1390–1399.

Clore, G.M. and Kuszweski, J. (2002) *J. Am. Chem. Soc.*, **124**, 2866–2867.

Clore, G.M., Brunger, A.T., Karplus, M. and Gronenborn, A.M. (1986) *J. Mol. Biol.*, **191**, 523–551.

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

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

Howe, P.W.A. (2001) *J. Biomol. NMR*, **20**, 61–70.

Humphrey, W., Dalke, A. and Schulten, K. (1996) *J. Mol. Graphics*, **14**, 33–38; available online at http://www.ks.uiuc.edu/Research/vmd/

Jones T.A., Zou J.Y., Cowan S.W. et al. (1991) *Acta Crystallogr.*, **A47**, 110–119.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317–324.

Schwieters, C.D. and Clore, G.M. (2001) *J. Magn. Reson.*, **149**, 239–244; available online at http://vmd-xplor.cit.nih.gov/

Sutcliffe, M.J. (1993) *Protein Sci.*, **2**, 936–944.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.