Rapid Identification of Ordered and Disordered Domains in NMR Structures

Craig E. Kundrot

Department of Chemistry and Biochemistry
University of Colorado
Boulder, Colorado 80309-0215

Received May 6, 1996

In contrast to X-ray crystal structures, NMR solution structures of biomacromolecules are reported as a family of 10 or more structures. All members of the family satisfy the experimental NMR constraints and retain good stereochemistry, but some regions of the structure are much better defined than others. An unsolved problem in the analysis of the structures produced by NMR spectroscopy is how to unambiguously determine which regions are well-defined and which are disordered. This communication presents a rapid, exhaustive, and unambiguous method for determining which regions of a structure are well-defined in a family of NMR structures.

The standard approach for identifying well-ordered domains is a trial and error process of superimposing members of the NMR family until one finds ranges of residues that superimpose well. This process involves a limited number of possibilities and many subjective decisions. Distance matrices[1−3] provide a method for comparing structures without superposition. A method has been described for identifying well-ordered domains using an iteratively filtered distance matrix projected into one dimension.[4] As described below, however, there are benefits to using an approach based on two-dimensional distance matrices.

A new, two-dimensional approach based on distance matrices has been implemented in the program ASDEM.[5] ASDEM produces a matrix, **A**, whose elements indicate the amount of structural variation between two residues in an NMR family of structures[6] (Figure 1). If two residues *m* and *n* belong to a well-ordered domain, their corresponding matrix element, A_{mn}, will have a small value. Well-ordered domains, therefore, give rise to submatrices that contain elements with small values (e.g., residues 11−13 in Figure 1). If, however, a residue is in a poorly defined domain, the rows and columns containing this residue will have large values (e.g., residue 1 in Figure 1). Lastly, if two well-defined domains are not ordered with respect to each other, they will produce small values for all intradomain elements but large values for the interdomain elements (e.g., residues 6−9 and 11−13 in Figure 1). In practice, the only adjustable parameter in this analysis is a "cutoff value" used to define well-ordered and disordered elements (disordered here means relative to the cutoff, not completely disordered). If A_{mn} is less than the cutoff, residues *m* and *n* are well-ordered with respect to each other. Conversely, if A_{mn} is greater than the cutoff, the residues are disordered with respect to each other.

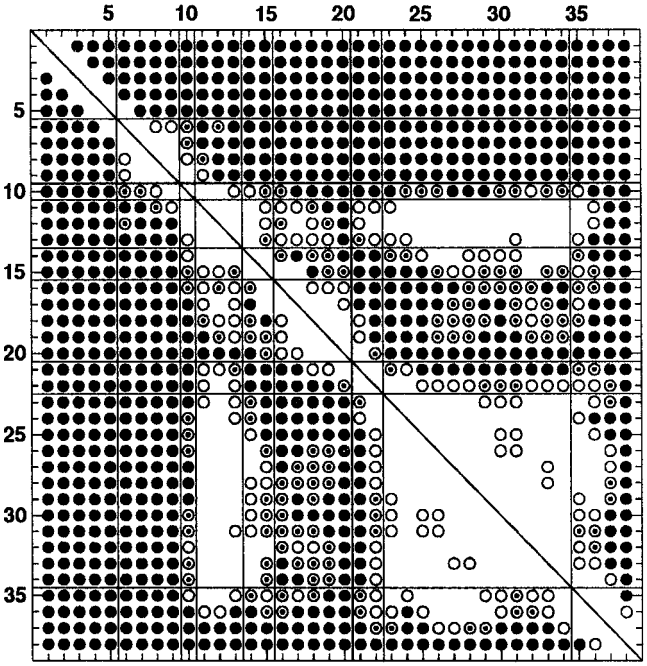

Figure 1. Average standard deviation matrix (ASDEM) for BNBD-12. The matrix contains elements A_{mn} with values 0.20 Å $< \sigma \leq$ 0.30 Å, 0.30 Å $< \sigma \leq$ 0.40 Å, and 0.40 Å $< \sigma$ indicated by open circles, circles containing dots, and black circles, respectively. White space indicates elements with $A_{mn} \leq$ 0.20 Å. The horizontal and vertical guide lines delineate the domains identified with a cutoff of 0.30 Å.

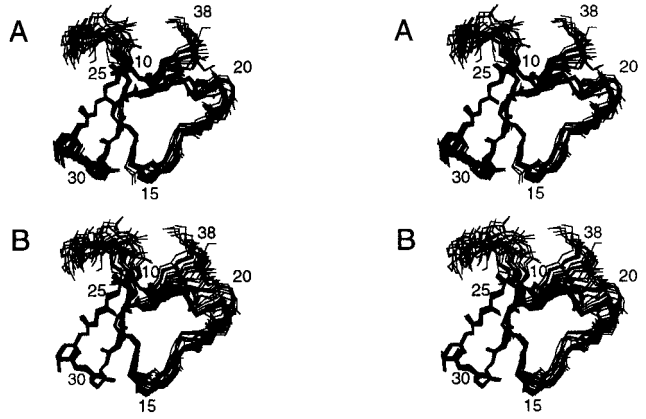

Figure 2. Superposition of the 20 BNBD-12 structures. The superpositions used the N, C_α, and C atoms of (A) domain 2 (10−13, 22−27, 32−36) from the original NMR analysis and (B) domain 2′ (11−13, 23−34) defined in this work using a cutoff level of 0.30 Å. Residues 1−5 are disordered and omitted for clarity.

The ASDEM method was applied to the 38-residue peptide bovine neutrophil β-defensin-12 (BNBD-12). The structure of BNBD-12 has been determined by NMR[7] and was found to contain two well-ordered domains: domain 1 (6−9) and domain 2 (10−13, 22−27, 32−36). The **A** matrix was calculated from the 20 BNBD-12 structures deposited in the Protein Data Bank.[8]

A cutoff value 0.30 Å shows three well-ordered domains: 1′ (6−9), 2′ (11−13, 23−34) and 3′ (16−20) (Figure 1). Domain 2′ is the same size as domain 2 in the original NMR analysis: 15 residues. However, eight of the residues differ. The average root-mean-square diameter (rmsd) for superimposing the 20 structures is 0.45 Å for the original domain 2 but only 0.29 Å

(1) Phillips, D. C. *Biochem. Soc. Symp.* **1970**, *31*, 11−28.
(2) Nishikawa, K.; Ooi, T.; Ysogai, Y.; Saito, N. *J. Phys. Soc. Jpn.* **1972**, *32*, 1331−1337.
(3) Richards, F. M.; Kundrot, C. E. *Proteins* **1988**, *3*, 71−84.
(4) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1987**, *219*, 11−16.
(5) ASDEM is written in FORTRAN 77 and is available from http://spot.Colorado.EDU/~kundrot/Home.html or from kundrot@colorado.edu.
(6) The elements in the average standard deviation matrix (ASDEM), A_{mn}, are calculated as follows. The distance matrix **D** (element D_{ij} equals the distance between atoms *i* and *j* of the molecule) is calculated for each family member. The standard deviation of element D_{ij}, σ_{ij}, is calculated over the *N* structures comprising the family of structures. A_{mn} is the average value of σ_{ij} calculated over all atoms *i* in residue *m* and all atoms *j* in residue *n*: $A_{mn} = \langle \sigma_{ij} \rangle$. In this example, only the N, C_α, and C atoms of each residue were used.
(7) Zimmermann, G. R.; Legault, P.; Selsted, M. E.; Pardi, A. *Biochemistry* **1995**, *34*, 13663−71.
(8) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F. J.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.

S0002-7863(96)01495-3 CCC: $12.00

for domain 2′ identified by the ASDEM method (Figure 2). Domains 1′ and 3′ have average pairwise rmsd values of 0.46 and 0.49 Å, respectively. Figure 1 also shows domain 1′, which corresponds exactly to the original domain 1, and domain 3′, which was not identified in the original analysis. Thus, compared to the original superposition analysis, the ASDEM method finds a large domain of the same size, but of better order, and it identifies an additional well-ordered domain.

The ASDEM method allows one to unambiguously assign each residue to a well-ordered or disordered domain for a given cutoff level. Using the 0.30 Å cutoff level, the assignments are 1−5, disordered; 6−9, domain 1′; 10, disordered; 11−13, domain 2′; 14−15, disordered; 16−20, domain 3′; 21−22, disordered; 23−34, domain 2′; 35−38 disordered.

If one changes the one adjustable parameter in this analysis, the cutoff value, then the domain boundaries change. For example, a cutoff value of 0.20 Å shrinks domain 2′ to 10 residues (11−12, 24−29, 33−34) (Figure 1). Another example is a cutoff of 0.40 Å. This level shows a new well-ordered domain: 4′ (36−38). Domains 1′ and 2′ would increase further to (6−10) and (11−13, 22−35), respectively, while domain 3′ would remain at (16−20). Even at the 0.40 Å cutoff, residues 1−5 are disordered with respect to themselves and the rest of the protein.

To summarize, the ASDEM method allows one to identify well-ordered and disordered domains within a family of NMR structures in a rapid, exhaustive, and unambiguous way. The major advantage of this method is that all pairwise comparisons of residues take place in a single matrix, and, given a cutoff level, the assignment of residues to well-ordered or disordered domains is unambiguous.

Acknowledgment. This work was funded by the Colorado RNA Center, W. M. Keck Foundation, and the National Science Foundation (MCB-9221307). I thank Art Pardi for many useful discussions.

JA961495O