Journal of Biomolecular NMR, 3 (1993) 601–606
ESCOM

J-Bio NMR 145

The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination

Peter Güntert, Kurt D. Berndt* and Kurt Wüthrich**

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland

Received 19 July 1993
Accepted 23 July 1993

Keywords: Protein stucture determination; NMR structure; Assignment of NOE distance constraints; Structure refinement

SUMMARY

A new program, ASNO ('ASsign NOes'), for computer-supported NOE cross-peak assignments is described. ASNO is used for structure refinement in several rounds of NOESY cross-peak assignments and 3D structure calculations, where the preliminary structures are used as a reference to resolve ambiguities in NOE assignments which are otherwise based on the chemical shifts available from the sequence-specific resonance assignments. The practical use of ASNO for proteins is illustrated with the structure determination of Dendrotoxin K from *Dendroaspis polylepis polylepis*.

Assignment of interresidual cross peaks in ^{1}H NOESY spectra (Anil-Kumar et al., 1980) for the collection of NOE upper limits on ^{1}H–^{1}H distances is an essential part of the determination of 3D protein structures in solution by NMR (Wüthrich, 1986). Obtaining NOESY cross-peak assignments is usually a laborious endeavour, particularly in spectral regions where chemical-shift degeneracies result in excessive cross-peak overlap. Were it not for these inevitable chemical-shift degeneracies and the usually somewhat imprecise cross-peak positional information, all assignments could of course be made automatically, from knowledge of the chemical shifts resulting from the sequence-specific resonance assignments (Wüthrich, 1986). In practice, however, typically only a fraction of the NOESY cross peaks can be assigned in this direct way and subsequent-

*Present address: Karolinska institutet, Institutionen för Medicinsk fysik, Stockholm, Sweden.
**To whom correspondence should be addressed.

Abbreviations; Toxin K, dendrotoxin K (or 'trypsin inhibitor homologue K') from the venom of the black mamba *Dendroaspis polylepis polylepis*; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; REDAC, use of redundant dihedral angle constraints; RMSD, root-mean-square deviation.

ly used to generate a preliminary 'low-resolution' structure of the protein under investigation. Subsequently, these preliminary conformers may be used to reduce the number of heretofore ambiguous assignments by eliminating pairs of protons which have the chemical-shift coordinates of the cross peak considered but, on the basis of the preliminary solution structure, are further apart than a predetermined maximum distance cutoff for the observation of NOEs. The program ASNO uses this principle for automated removal of ambiguities arising from chemical-shift degeneracies, and thus supports the collection of an extensive input of NOE distance constraints in several rounds of NOESY cross-peak assignments and structure calculations.

For structure determination of proteins using 2D, 3D or 4D NMR experiments, ASNO has been implemented in the new interactive program package XEASY (C. Bartels, T.H. Xia, M. Billeter, P. Güntert and K. Wüthrich, to be published), which uses the X-window system (MIT) and replaces the previously described program package EASY (Eccles et al., 1991). The ASNO routine can be obtained from the authors as part of the XEASY package. Here, the practical use of the program ASNO is illustrated for dendrotoxin K, a small protein of 57 amino acid residues, for which the structure determination was based entirely on homonuclear 2D ^{1}H NMR experiments (Berndt et al., 1993).

The function of ASNO in the context of a complete protein structure determination in our laboratory is outlined in Fig. 1. The program XEASY supports the peak picking of the NMR spectra and the determination of sequence-specific resonance assignments. Once a complete (or in practice, nearly complete) list of ^{1}H chemical shifts has thus been obtained, and intraresidual cross peaks in the NOESY spectra are identified by comparison with COSY and TOCSY spectra, a limited number of interresidual NOESY cross peaks can be unambiguously assigned to pairs of hydrogen atoms on the basis of the chemical shifts (typically 30 to 50% of all cross peaks). The volumes of the assigned NOESY cross peaks are converted to upper bounds on ^{1}H–^{1}H distances using the program CALIBA (Güntert et al., 1991a,b). The resulting preliminary, incomplete list of NOE upper distance bounds is, where applicable, supplemented with additional conformational constraints, e.g., explicit upper and lower distance bounds for maintaining the covalent structure of disulfide bonds (Williamson et al., 1985), and constraints from spin–spin coupling constants (measured, for example, using the program INFIT (Szyperski et al., 1992); see Fig. 1). The program HABAS (Güntert et al., 1989) is used to analyse the local conformation on the level of dipeptide segments from the experimental coupling constants and the intraresidual and sequential NOE distance constraints, and to obtain stereospecific assignments for pairs of diastereotopic substituents. With this pretreated input a first set of conformers is generated with the program DIANA (Güntert et al., 1991a). Using this set of preliminary solution conformers as a reference, the program ASNO is used to help eliminate possible chemical shift-compatible cross-peak assignments involving pairs of protons that are further separated than a predetermined maximum distance cutoff for the observation of NOEs, and the program GLOMSA (Güntert et al., 1991a) is used for obtaining additional stereospecific assignments. The final, high-quality conformers obtained after several rounds of DIANA and ASNO calculations (Fig. 1) are subjected to a restrained energy minimization, e.g. with the program OPAL (P. Luginbühl, P. Güntert, M. Billeter and K. Wüthrich, to be published).

The input for the program ASNO consists of (i) a list of the proton chemical shifts from XEASY; (ii) a peak list from XEASY containing the chemical-shift coordinates of the cross-peaks in the NOESY spectrum; and (iii) a set of conformers calculated with DIANA using the previous,

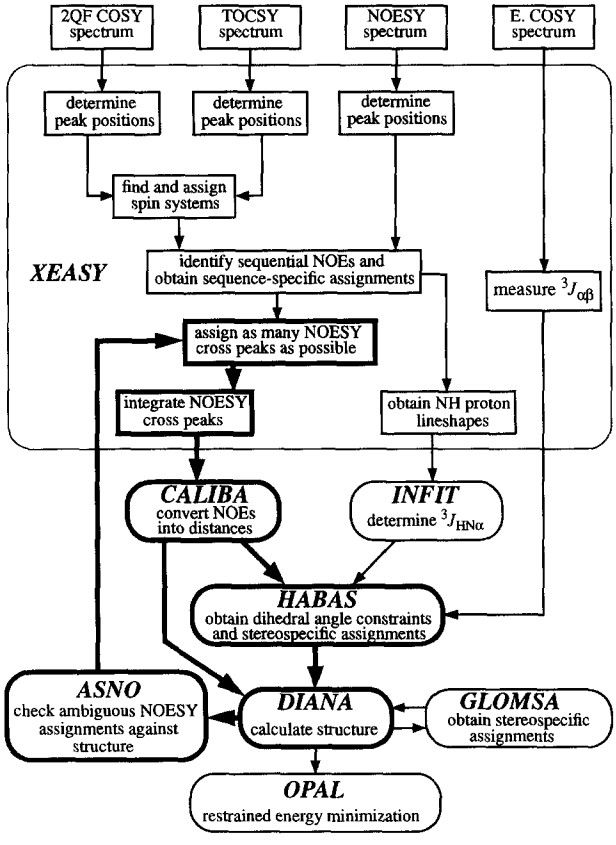

Fig. 1. Outline of the computer-supported procedure currently used in our laboratory for protein structure determination using the software packages XEASY, INFIT, CALIBA, HABAS, DIANA, GLOMSA, ASNO and OPAL. The refinement cycle including ASNO is emphasized with thicker lines (see text for details).

less complete input of NOE distance constraints. (This part of the input could also consist of the atom coordinates of a homologous protein, or the crystal structure of the protein under investigation; however, in such applications much care must be exercised to avoid possible bias in the imported reference data.) In addition, the user specifies the maximally allowed chemical-shift differences between corresponding cross-peak coordinates and proton chemical-shift values, $\Delta\omega_1$ and $\Delta\omega_2$, to be used for chemical shift-based assignments, the maximally allowed proton–proton distance in the molecular structure, d_{max}, that may give rise to an observable NOE, and the minimal number of conformers among those used to represent the solution structure, n_{min}, for which a given proton–proton distance must be shorter than d_{max} in order for a possible NOE assignment to be accepted. For a cross peak p with chemical-shift coordinates ω_1^p and ω_2^p, ASNO first determines the set of all possible chemical shift-based assignments (α, β), according to Eq. 1, where α and β denote entries in the proton list.

$$\left\{ (\alpha, \beta) \left| \left(\frac{\omega_1^p - \omega^\alpha}{\Delta\omega_1} \right)^2 + \left(\frac{\omega_2^p - \omega^\beta}{\Delta\omega_2} \right)^2 \leq 1;\ \alpha = 1, \ldots, n;\ \beta = 1, \ldots, n \right. \right\} \quad (1)$$

ω^{α} and ω^{β} are the chemical shifts of the protons α and β; n is the number of entries in the proton list. The resulting chemical shift-based assignments are checked against the corresponding ^1H–^1H distances in the available group of preliminary conformers and retained only if the distance between the two protons α and β is shorter than d_{max} in at least n_{min} conformers. If pseudo-atoms are used to represent groups of protons, the limit d_{max} is increased by the appropriate pseudoatom correction (Wüthrich et al., 1983). Unambiguously assigned cross peaks according to these criteria are used to derive additional distance constraints for the subsequent round of structure calculations. For Toxin K we used ASNO with $\Delta\omega_1 = \Delta\omega_2 = 0.015$ ppm, $d_{max} = 5.0$ Å and $n_{min} = 1$ out of 20 conformers.

Figures 2 and 3 illustrate the use of the program ASNO for the NMR structure determination of Toxin K (Berndt et al., 1993). Figure 2 shows how the chemical shift-based assignments and their ASNO assessment, based on the preliminary conformers, can be displayed and interactively checked or edited within the program XEASY. The overall effect of increasing the number of NOE upper distance constraints in successive rounds of structure refinement has been visualized in Fig. 3, where four groups of 10 Toxin K conformers each are displayed, which were selected at

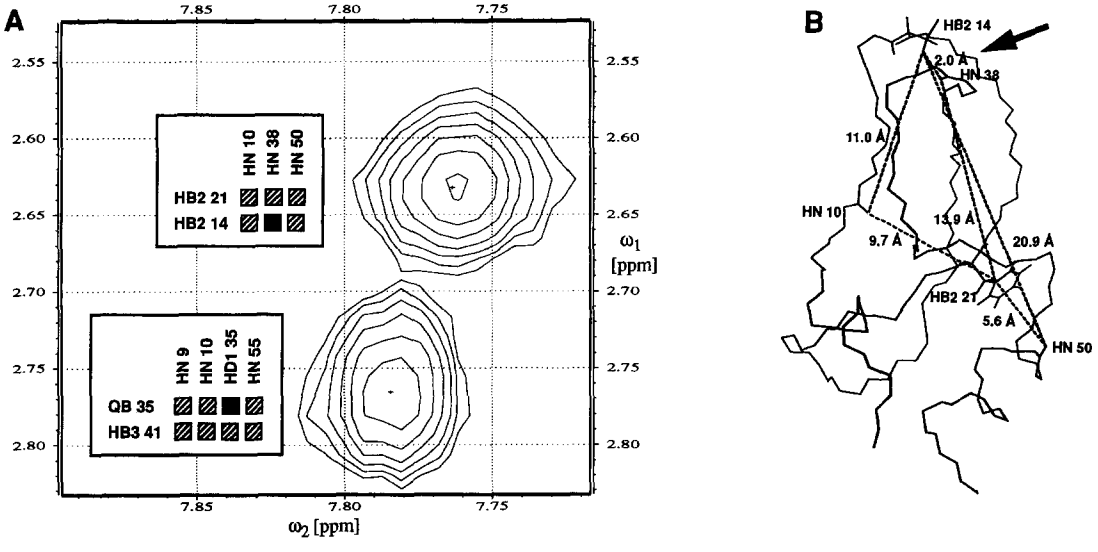

Fig. 2. Illustration of the use of the program ASNO in conjunction with the program package XEASY (C. Bartels, T.H. Xia, M. Billeter, P. Güntert and K. Wüthrich, to be published) for interactive spectral analysis. (A) Contour plot of a spectral region containing two cross peaks in the NOESY spectrum of Toxin K (Berndt et al., 1993) in H_2O solution. The cross peaks are displayed by XEASY together with the corresponding assignment matrices. In each matrix the rows and columns are identified with the type of hydrogen atom and the sequence position of the amino acid residue, the hatched squares denote chemical shift-based assignments, and black squares denote assignments by ASNO using the parameters $\Delta\omega_1 = \Delta\omega_2 = 0.015$ ppm, $d_{max} = 5.0$ Å and $n_{min} = 1$ out of 20 conformers (see text). The figure shows that ASNO yielded unique assignments for both peaks, whereas the chemical-shift information alone led to sixfold and eightfold degenerate assignments, respectively. (B) View of the Toxin K conformer with the lowest final DIANA target function value (Berndt et al., 1993), produced with the program MIDAS (Ferrin et al., 1988). All backbone atoms N, C^{α} and C′, the backbone amide protons of residues 10, 38 and 50, and the side chains of the residues Cys^{14} and Tyr^{21} have been drawn. All possible chemical shift-based assignments of the upper NOESY cross peak in A are identified by dashed lines, and the averages of the corresponding distances in the 20 final DIANA conformers are indicated. The arrow points to the distance corresponding to the unique assignment by ASNO.

different stages of the structure refinement process (Berndt et al., 1993). The initial dataset, obtained using only chemical-shift information for the assignment of NOESY cross peaks, contained a total of 322 NOE upper distance constraints (excluding NOEs corresponding to distances that are independent of the conformation), i.e., approximately 40% of the final number. Using this data combined with nine upper and nine lower distance constraints for the three disulfide bonds (Williamson et al., 1985) and the hitherto identified dihedral angle constraints from the program HABAS, a structure calculation was started with 100 randomized starting conformers, using the program DIANA. Although the resulting 10 best conformers do not form a tight bundle, the global polypeptide fold is nonetheless clearly defined (Fig. 3A). These conformers

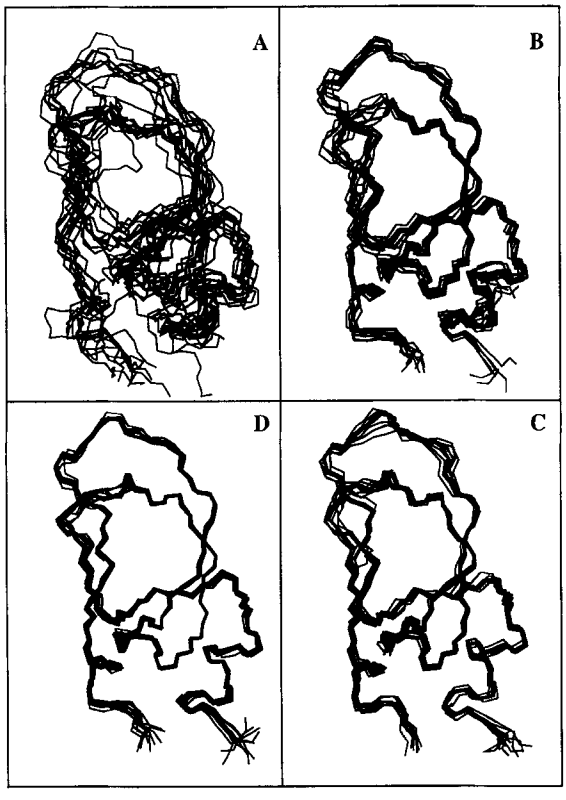

Fig. 3. View of the polypeptide backbone of the 10 best DIANA conformers of Toxin K at four stages of structure refinement using the program ASNO (Fig. 1). (A) Structure calculated on the basis of 322 NOE upper distance constraints, assigned from chemical-shift information only (213 intraresidual and sequential; 109 medium-range and long-range; average RMSD (McLachlan, 1979) for the backbone atoms N, C^{α} and C' to the mean structure = 1.66 ± 0.40 Å). (B) Structure calculated from an input of 657 NOE upper distance constraints assigned using ASNO with the structure from (A) (359 intraresidual and sequential, 298 medium-range and long-range backbone; RMSD to the mean structure = 0.65 ± 0.10 Å). (C) Structure calculated from an input of 747 relevant NOE upper distance constraints assigned using ASNO with the structure from (B) (373 intraresidual and sequential, 365 medium-range and long-range; backbone RMSD to the mean structure = 0.38 ± 0.09 Å). (D) 'Final' structure calculated from 809 NOE upper distance constraints (398 intraresidual and sequential, 411 medium-range and long-range; backbone RMSD to the mean structure = 0.32 ± 0.07 Å). At all four stages (A) – (D), additional constraints for the three disulfide bonds as well as the available dihedral angle constraints were added to the input (see text).

were then used as input to the program ASNO to assign additional NOESY cross peaks with heretofore ambiguous assignments due to chemical-shift degeneracies. The resulting 'second cycle' dataset now contained a total of 657 constraints from NOESY cross peaks. With the exception of some chain-terminal residues, the resulting structure was already quite well defined (Fig. 3B). Further improvement of the structure by additional cycles of refinement (Figs. 3C and D) is mostly confined to the polypeptide segments linking the regular secondary-structure elements. For the final DIANA calculation with 809 NOE upper distance limits, the REDAC strategy (Güntert and Wüthrich, 1991) was used, which resulted in 47 out of 50 starting conformers having a final target function value of less than 2.0 $Å^2$ (Berndt et al., 1993).

ACKNOWLEDGEMENT

We acknowledge financial support from the Schweizerischer Nationalfonds (project 31.32033.91) and thank R. Marani for the careful processing of the manuscript.

REFERENCES

Anil-Kumar, Ernst, R.R. and Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.*, **95**, 1–6.
Berndt, K.D., Güntert, P. and Wüthrich, K. (1993) *J. Mol. Biol.*, in press.
Eccles, C., Güntert, P., Billeter, M. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 111–130.
Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Langridge, R. (1988) *J. Mol. Graphics*, **6**, 13–27.
Güntert, P. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 447–456.
Güntert, P., Braun, W. and Wüthrich, K. (1991a) *J. Mol. Biol.*, **217**, 517–530.
Güntert, P., Braun, W., Billeter, M. and Wüthrich, K. (1989) *J. Am. Chem. Soc.*, **111**, 3997–4004.
Güntert, P., Qian, Y.Q., Otting, G., Müller, M., Gehring, W. and Wüthrich, K. (1991b) *J. Mol. Biol.*, **217**, 531–540.
McLachlan, A.D. (1979) *J. Mol. Biol.*, **128**, 49–79.
Szyperski, T., Güntert, P., Otting, G. and Wüthrich, K. (1992) *J. Magn. Reson.*, **99**, 552–560.
Williamson, M.P., Havel, T.F. and Wüthrich, K. (1985) *J. Mol. Biol.*, **182**, 295–315.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids,* Wiley, New York.
Wüthrich, K., Billeter, M. and Braun, W. (1983) *J. Mol. Biol.*, **169**, 949–961.