*Journal of Biomolecular NMR*, **12:** 543–548, 1998.
*KLUWER/ESCOM*



# Conformational analysis of protein and nucleic acid fragments with the new grid search algorithm FOUND

Peter Güntert[a], Martin Billeter[a,*], Oliver Ohlenschläger[b], Larry R. Brown[b] and Kurt Wüthrich[a,**]
[a]*Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland;* [b]*Institut für Molekulare Biotechnologie, Abteilung Molekulare Biophysik/NMR-Spektroskopie, Postfach 100813, D-07745 Jena, Germany*

Received 23 April 1998; Accepted 2 June 1998

*Key words:* NMR structure determination, distance restraints, torsion angle restraints, stereospecific assignment, local conformation analysis, grid search

**Abstract**

The new computer algorithm FOUND, which is implemented as an integrated module of the DYANA structure calculation program, is capable of performing systematic local conformation analyses by exhaustive grid searches for arbitrary contiguous fragments of proteins and nucleic acids. It uses torsion angles as the only degrees of freedom to identify all conformations that fulfill the steric and NMR-derived conformational restraints within a contiguous molecular fragment, as defined either by limits on the maximal restraint violations or by the fragment-based DYANA target function value. Sets of mutually dependent torsion angles, for example in ribose rings, are treated as a single degree of freedom. The results of the local conformation analysis include allowed torsion angle ranges and stereospecific assignments for diastereotopic substituents, which are then included in the input of a subsequent structure calculation. FOUND can be used for grid searches comprising up to 13 torsion angles, such as the backbone of a complete α-helical turn or dinucleotide fragments in nucleic acids, and yields a significantly higher number of stereospecific assignments than the precursor grid search algorithm HABAS.

*Abbreviations:* CPU, central processing unit; DYANA, dynamics algorithm for NMR applications; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy.

In the course of a structure determination by NMR (Wüthrich, 1986) one determines the allowed conformation space of the molecule, which is restricted by stereochemical, steric and experimental contraints. Due to the size and complexity of the conformation space, as well as the large number of experimental constraints, this is a formidable calculation for typical proteins and nucleic acid fragments. However, the structure determination can be simplified by considering limited fragments of the complete molecule which are amenable to systematic grid searches. A large fraction of the experimental restraints can thus be checked for inconsistencies before the actual structure calculation is started. Furthermore, stereospecific assignments of pairs of diastereotopic protons or isopropyl groups can be obtained by performing separate local grid searches for both possible stereospecific assignments (Güntert et al., 1989; Nilges et al., 1990; Polshakov et al., 1995), and experimental data such as scalar coupling constants and short range NOEs can be converted into direct restraints on torsion angles. This may improve the success rate of the subsequent structure calculations and the quality of the resulting structure. Grid search methods are especially useful for nucleic acids, where long-range NOEs are scarce and hence short-range restraints become more important (Wüthrich, 1986). This paper describes a new

*Present address: Biochemistry and Biophysics, Box 462, Göteborg University, SE 405 30 Göteborg, Sweden.
**To whom correspondence should be addressed.

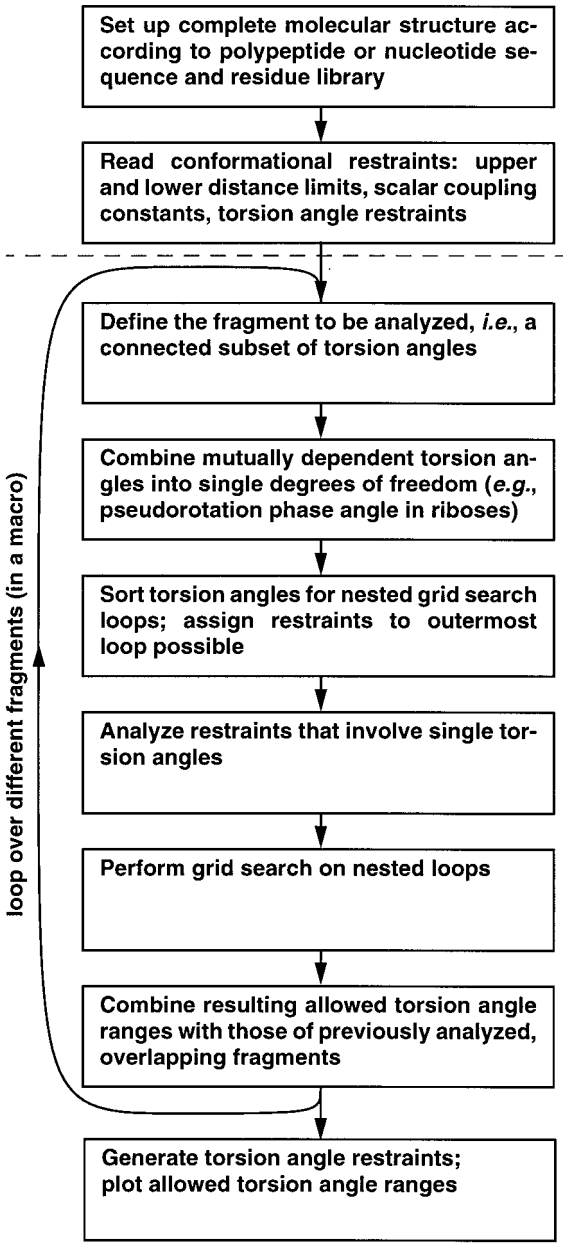


*Figure 1.* Flow chart of the FOUND algorithm. The initial two steps are performed by general commands of the DYANA program (Güntert et al., 1997).

grid search algorithm, FOUND, which permits us to study larger protein and nucleic acid fragments than previously available procedures, and yields a high percentage of stereospecific assignments for pairs of diastereotopic hydrogen atoms or methyl groups.

Grid searches have so far been used predominantly for the local conformation analysis of protein fragments involving the three torsion angles $\phi$, $\psi$ and $\chi^1$ of an amino acid residue with the aims of obtaining stereospecific assignments of the β-protons and restraints for the torsion angles $\phi$, $\psi$ and $\chi^1$ (Güntert et al., 1989; Nilges et al., 1990). In contrast to these earlier approaches, which were specific for $\phi$–$\psi$–$\chi^1$ fragments in proteins, FOUND enables one to study arbitrary fragments, defined as connected subsets of torsion angles in a molecular structure. Covalent geometry parameters such as bond lengths and bond angles, and Karplus relations (Karplus, 1959) are taken from the standard DYANA residue library (Güntert et al., 1997). This allows, in principle, the treatment of any type of molecule, the fragment size being limited only by the computing power available. All upper and lower limit distance restraints, scalar couplings and angle restraints available within the chosen molecular fragment are used as input for the search. Stereospecific assignments be determined can also for other groups than βCH2, and for more than one pair of diastereotopic substituents per fragment. Often it is advisable, for improved efficiency, to perform multiple grid searches with relatively small molecular fragments. Results from different grid searches for

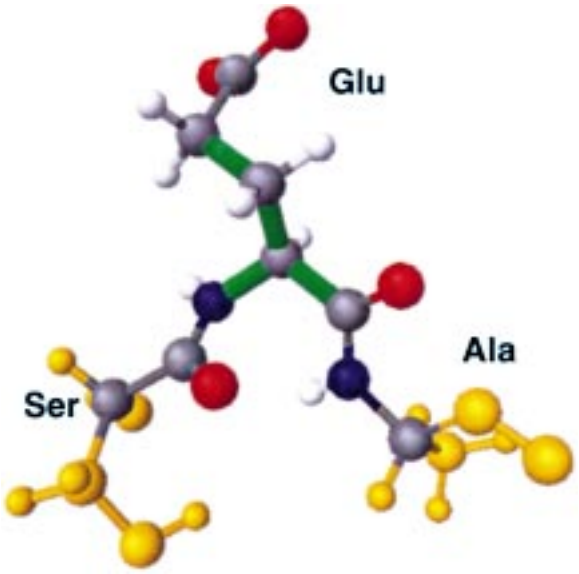


*Figure 2.* Ball-and-stick representation of a 'dipeptide fragment' defined by the torsion angles $\phi$, $\psi$, $\chi^1$ and $\chi^2$ of the central Glu amino acid residue. Atoms included in the dipeptide fragment are colored gray (carbon), purple (nitrogen), red (oxygen) and white (hydrogen), and those not included are yellow. The four variable torsion angles are identified by green coloring of the corresponding bonds; the gray bonds are non-rotatable for the grid search, where the $\omega$ torsion angles are assumed to be *trans* (180°). This figure and Figure 4 were created with the program MOLMOL (Koradi et al., 1996).

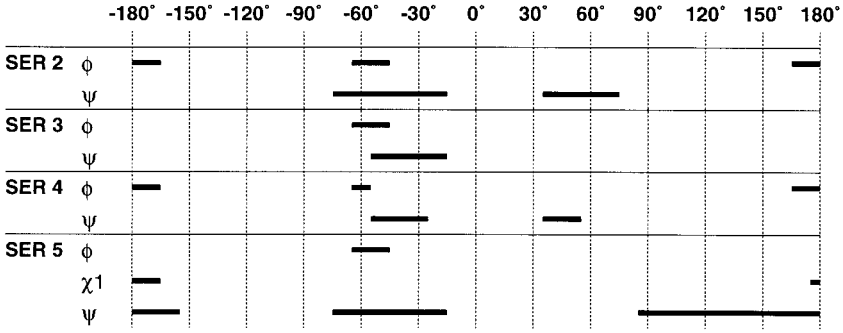


*Figure 3.* Allowed torsion angle ranges (horizontal bars) obtained in a 9-dimensional grid search for a tetrapeptide fragment $-\mathrm{Ser}_2-\mathrm{Ser}_3-\mathrm{Ser}_4-\mathrm{Ser}_5-$, with the $\phi$ and $\psi$ backbone torsion angles of the four serine residues 2–5 and the $\chi^1$ side chain torsion angle of Ser 5 as variables. A simulated NMR data set for an $\alpha$-helical peptide was used as input for the grid search (see text).

overlapping fragments can then be combined to generate the narrowest possible torsion angle restraints before the start of the actual structure calculation.

The FOUND algorithm is shown schematically in Figure 1. Since FOUND is an integrated module of the DYANA program (Güntert et al., 1997), the standard DYANA data structures and commands can be used to represent the molecule under investigation on the basis of its sequence. The DYANA residue libraries contain the covalent structures of amino acid residues and nucleotides according to the ECEPP (Momany et al., 1975) or AMBER (Cornell et al., 1995) force fields. The molecule is represented as a tree structure consisting of rigid units, i.e., groups of atoms with invariant relative positions that are connected by rotatable bonds. The degrees of freedom are exclusively variations of torsion angles about single bonds. Thereby, sets of mutually dependent torsion angles are treated as a single degree of freedom. For example, the relations of the ribose ring torsion angles $\nu_0$, $\nu_1$, $\nu_2$, $\nu_3$ and $\nu_4$ in nucleic acids with the pseudorotation phase angle, $P$, are given by (Saenger, 1984),

$$\nu_k = \nu_{\max} \cos\left(P + \frac{4\pi}{5}(k-2)\right)$$
$$(P \in [0, 2\pi];\ \nu_{\max} = 40^\circ;\ k = 0, \ldots, 4), \qquad (1)$$

and the parameter $P$ is used as a single degree of freedom in the grid search. This ensures that the relations among the mutually dependent torsion angles are always fulfilled, and the efficiency of the algorithm is significantly increased. Conformational restraints, i.e., upper and lower distance limits, scalar coupling constants and torsion angle restraints, are read into the program using the corresponding DYANA commands, and steric lower limits are set up automatically by the program (Güntert et al., 1991).

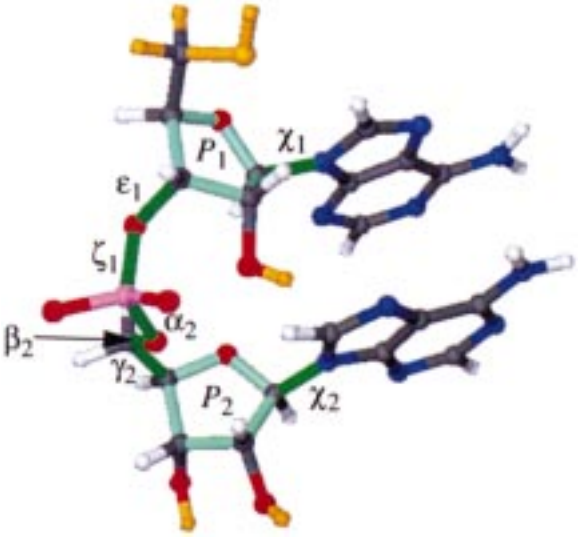


*Figure 4.* RNA-dinucleotide ApA. The fragment considered by FOUND in a test calculation with a simulated NMR input (see text) is drawn with the same color code as Figure 2, and with pink color for $^{31}$P. Independent rotations are possible about the dark green bonds, i.e., each of these bonds represents one dimension in the grid search. Light green bonds are those forming sugar rings, where each sugar ring corresponds to one grid search dimension represented by the pseudorotation phase angle (Saenger, 1984).

The molecular fragment to be analyzed in the grid search is defined by selecting a connected subset of torsion angles (Figure 2). The program then extracts from the input data for the complete molecule the subset of conformational restraints located within the chosen fragment, and evaluates them in a multidimensional grid search. The grid search can be illustrated by $n$ nested loops, each of which corresponds to a degree of freedom. The actual implementation can handle any number of degrees of freedom (a maximum is specified at compile time). Since a restraint is evaluated

as soon as all relevant torsion angles are set, it is in general not necessary to completely build each conformation that corresponds to a grid point. Allowed partially or completely built conformations are characterized by a DYANA target function value (Güntert et al., 1991, 1997) below a user-defined threshold, $f_{\max}$. If, alternatively, thresholds on the sizes of individual violations are applied, the same algorithm can be used with $f_{\max} = 0$; a contribution of 1 is then added to the target function whenever a violation exceeds its threshold.

In routine applications the results of the grid search comprise the number of allowed conformations, *N*, and the sets of allowed values for each torsion angle in the molecular fragment chosen for the analysis (Figure 3). The result yields the important information that the set of experimental restraints for a given fragment is compatible with at least one conformation. Inconsistencies among the local restraints can thus be detected right at the start of a NMR structure determination. If the fragment under investigation contains *M* pairs of diastereotopic substituents for which the stereospecific assignment is not known, $2^M$ grid searches will be performed, one for each combination of possible stereospecific assignments, and their results will be combined. By checking whether the restraints are compatible with only one of the two possible stereospecific assignments of a diastereotopic pair, the algorithm can also be used to obtain stereospecific assignments on the basis of local restraints. In practice, a complete local conformation analysis of a macromolecule includes many grid searches for (possibly overlapping) fragments, and as a result restraints are generated for all torsion angles that formed part of at least one of the fragments studied. For each torsion angle a restraint in the form of either a single allowed interval that includes all allowed values, or of several smaller intervals can be created. With the use of a single interval some information may be lost if more than one contiguous allowed region was found in the grid search. The use of multiple intervals, on the other hand, may add extra local minima to the target function in the subsequent structure calculation.

As an illustration of the function of FOUND we simulated NMR data for a helical tetrapeptide within a protein, $-\mathrm{Ser}_2-\mathrm{Ser}_3-\mathrm{Ser}_4-\mathrm{Ser}_5-$, where the atoms $C^\alpha$, $C'$ and O of the preceding residue and the atoms N, $H^N$ and $C^\alpha$ of the following residue are also part of the fragment considered (Figure 2). The following input data were defined in order to enforce a helical conformation: For residues 2–5 the coupling constants $J_{\mathrm{HN}\alpha}$ were set to 4.0 ± 1.5 Hz, and the sequential distances $d_{\mathrm{NN}}$ for the residue pairs 2–3, 3–4 and 4–5 were restrained by upper limits of 2.8 Å (Wüthrich, 1986). In addition, the two coupling constants $^3J_{\alpha\beta}$ of Ser 5 were set to 3.5 ± 1.5 Hz and 10.5 ± 1.5 Hz, respectively. This allows *trans* and *gauche* – $\chi^1$-rotamer conformations, depending on the stereospecific assignment of the β-protons. Finally, two medium-range upper distance limits between the α-proton of Ser 2 and the two β-protons of Ser 5 (Wüthrich, 1986) were set to 3.6 Å. A nine-dimensional grid search with a step size of 10° involving the torsion angles φ and ψ of residues 2–5 and $\chi^1$ of Ser 5 was performed. Conformations were only accepted when no violation was present. The result comprises 7042 different conformations for one possible stereospecific assignment of the two β-protons of Ser 5, and zero conformations for the other. Thus, although five of the nine torsion angles have two allowed ranges and the ranges on all ψ-angles include at least 40° (Figure 3), an unambiguous stereospecific assignment of the β-methylene group of Ser 5 was obtained. The total calculation time was 32 seconds on a Silicon Graphics Indigo2 computer with an R10000 processor.

The performance of FOUND in the determination of stereospecific assignments in globular proteins was assessed on the basis of the previously published experimental NMR data set for the 135-residue pathogenesis-related protein P14a from tomato leaves (Fernández et al., 1997). This data set comprises 2435 upper distance limits derived from NOESY cross peaks, and the following spin–spin coupling constants: 111 $^3J_{\mathrm{HN}\alpha}$, 51 $^3J_{\alpha\beta}$, 87 $^3J_{\mathrm{N}\beta}$ and 93 $^3J_{C'\beta}$ . The number of upper distance limits corresponds to those actually measured and differs from the final number given by Fernández et al. (1997), who modified the data set to account for missing stereospecific assignments (Güntert et al., 1991). The accuracy of the coupling constant values was assumed to be ±1.5 Hz for $^3J_{\mathrm{HN}\alpha}$, ±2.0 Hz for $^3J_{\alpha\beta}$ and $^3J_{\mathrm{N}\beta}$, and ±3.0 Hz for $^3J_{C'\beta}$. FOUND was applied to all 'dipeptide' fragments defined by the torsion angles φ, ψ, $\chi^1$ and $\chi^2$ of residue *i*, which also includes the atoms $C^\alpha$, $C'$ and O of the preceding residue, and the atoms N, $H^N$ and $C^\alpha$ of the following residue, since the ω torsion angles of the peptide bonds are fixed in either the *trans* or *cis* conformation (Figure 2). The target function threshold for allowed conformations was set to 0.02 $Å^2$, and the step size of the grid search was 10°. The reduction of the allowed conformation space in the four-dimensional grid searches is mostly due to

the experimental restraints: As a typical example, the theoretical total number of conformations for Leu 29, $36^4 = 1\ 679\ 616$ is reduced by the steric part of the target function to 193 655, whereas the corresponding number of allowed conformations drops to 874 upon inclusion of the experimental restraints. The grid searches yielded stereospecific assignments for 41 β-methylene groups, two additional proline methylene groups and the isopropyl groups of 7 valines and 1 leucine. FOUND also detected inconsistencies among the local input restraints for two residues. The total calculation time for all dipeptide fragments in the protein was 111 seconds on a Silicon Graphics Indigo2 computer with an R10000 processor.

The number of stereospecific assignments obtained with FOUND significantly exceeds the 30 assignments previously obtained with HABAS. To better rationalize the improved performance we performed additional test calculations with P14a, using different conditions. First, the same criteria for defining allowed conformations were used as in the HABAS program, with thresholds of 0.1 Å on violations of individual experimental and steric distance restraints, and 0.5 Hz on coupling constants. A slightly lower number of stereospecific assignments was obtained than when using the target function threshold, i.e., 34 for β-methylene groups and 10 for other groups, and inconsistencies were detected for three residues. This shows that the target function criterion for allowed conformations is not only consistent with the ranking of individual conformers in a subsequent DYANA structure calculation (Güntert et al., 1997), but gives also a higher number of unambiguous stereospecific assignments. Second, in another calculation where the $\chi^2$ torsion angle was not included in the dipeptide fragments, nearly the same number of stereospecific assignments was obtained as in the calculation including $\chi^2$; only the stereospecific assignment of the leucine isopropyl group was lost, since it is not part of the $\phi$–$\psi$–$\chi^1$ fragment that was then analyzed in the grid search. The low impact of the $\chi^2$ torsion angle on the number of stereospecific assignments confirms that the homonuclear and heteronuclear scalar couplings with the atoms N, $H^\alpha$ and C′ of the polypeptide backbone are the most important data determining the conformation of the β-methylene group. The inclusion of $\chi^2$ in the grid search can nevertheless be beneficial, since the resulting $\chi^2$ torsion angle restraints may improve convergence of the subsequent structure calculation.

In double helix DNA and RNA the majority of NMR-derived structural restraints are intraresidual or sequential, and hence amenable to systematic analysis of dinucleotide fragments by FOUND. A very high percentage of all experimental restraints may thus be tested for inconsistencies and translated into ranges of allowed torsion angles prior to the actual structure calculation. A simulated NMR data set for the RNA dinucleotide ApA in the A-type conformation ($\alpha = -68°$, $\beta = 174°$, $\gamma = 52°$, $\delta = 82°$, $\nu_1 = -23.5°$, $\nu_2 = 38°$, $\chi = -164°$, $\varepsilon = -156°$, $\zeta = -72°$) was generated by extracting all $^1$H–$^1$H distances shorter than 4.5 Å and imposing corresponding distance restraints, with the lower limits 0.75 Å smaller and the upper limits 0.75 Å larger than the actual distance. It was further assumed that individual assignments were available for $NH_2$ protons in the adenine bases, but not for the 5′ sugar protons. In addition, the simulated data set included $^1$H–$^{31}$P vicinal scalar coupling constants of $^3J_{H3'P} = 6.5 \pm 2.5$ Hz, $^3J_{PH5'} = 3.1 \pm 2.5$ Hz and $^3J_{PH5''} = 1.5 \pm 2.5$ Hz. The molecular fragment that was analyzed by FOUND is shown in Figure 4, where the nine degrees of freedom for the grid search are identified. In order to speed up the calculation, smaller grid searches for the following three parts of the dinucleotide fragment were performed before starting the nine-dimensional grid search: each sugar ring with the attached base, and the backbone part from $\varepsilon_1$ to $\gamma_2$ connecting the two sugar rings. The results of these grid searches were added to the input for the final nine-dimensional grid search. Figure 5 shows that torsion angle ranges smaller than 60° are obtained for all 13 angles. Note that due to the absence of hydrogen atoms in the phosphate group, no direct experimental restraints were available for $\zeta_1$ and $\alpha_2$ (Figure 4), but these angles where nonetheless confined to limited ranges by the steric crowding encountered in the search for possible relative orientations of the two nucleoside moieties. All allowed torsion angle ranges include the standard A-RNA value, which was used to extract the input restraints. In the nine-dimensional conformation space with grid steps of 10 degrees, 829 out of the theoretical total of $36^9 \approx 10^{14}$ conformations were found to be compatible with the experimental and steric conformational restraints. The input was also sufficient to obtain stereospecific assignments for the 5′ sugar protons. All grid searches together required about 100 seconds on a Silicon Graphics Indigo2 computer with an R10000 processor.

FOUND has been applied to experimental NMR data obtained for a 19 nucleotide $^{13}$C,$^{15}$N-labelled RNA hairpin loop (Sich et al., 1997). A total of 280

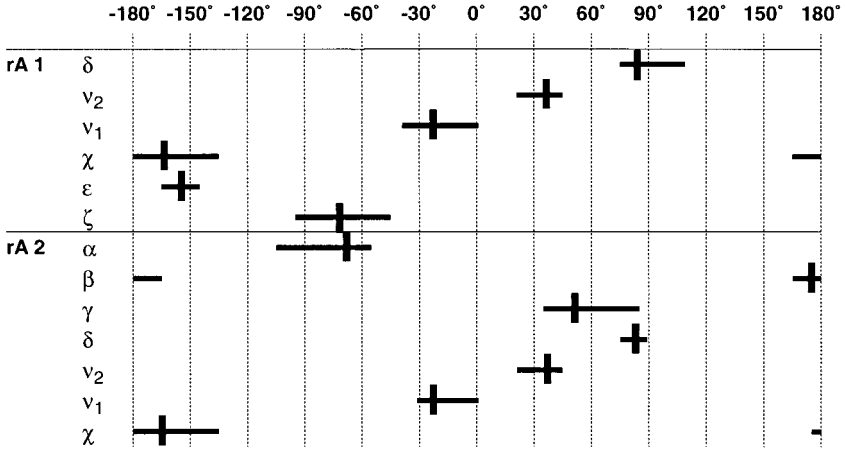


*Figure 5.* Allowed torsion angle ranges (horizontal bars) obtained in a 9-dimensional grid search for the dinucleotide fragment of Figure 4. The torsion angle values of standard A-RNA, from which the NMR data were simulated, are indicated by vertical bars.

distance restraints and 132 scalar coupling constants ($17\,^3J_{H1'H2'}$, $11\,^3J_{H2'H3'}$, $16\,^3J_{H3'H4'}$, $22\,^3J_{H4'H5'/H5''}$, $17\,^3J_{PC4'}$, $26\,^3J_{PH5'/H5''}$, $14\,^3J_{PH3'}$, $9\,^3J_{PC2'}$) resulted in 129 torsion angle restraints. The grid search identified stereospecific assignments of the C5′ hydrogens for four of the 12 nucleotides in the A-form stem of this molecule, and the large number of torsion angle restraints obtained with FOUND greatly improved the convergence of subsequent distance geometry calculations with the program DIANA (Güntert et al., 1991).

Overall, when compared with our own precursor grid search program HABAS (Güntert et al., 1989) and those used by others (e.g., Nilges et al., 1990; Polshakov et al., 1995) the new FOUND algorithm is more versatile due to the higher accessible dimensionality of the search, which makes FOUND attractive also for use with nucleic acids. The increase, when compared with HABAS, of 70% in the total number of stereospecific assignments determined from the experimental NMR data set for the protein P14a (Fernández et al., 1997) is also most encouraging, since the availability of stereospecific assignments at the outset of a structure calculation not only improves convergence, but can also improve the overall quality of the structure determination (Güntert et al., 1989).

The FOUND algorithm is available from the authors at http://www.mol.biol.ethz.ch/dyana/.

## Acknowledgements

Financial support by the Swiss National Science Foundation (project 49047.96) and by the Deutsche Forschungsgemeinschaft (project Br 1487/2) is gratefully acknowledged. We thank Mrs. M. Geier for the careful processing of the manuscript.

## References

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz Jr., K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A. (1995) *J. Am. Chem. Soc.*, **117**, 5179–5197.

Fernández, C., Szyperski, T., Bruyère, T., Ramage, P., Mösinger, E. and Wüthrich, K. (1997) *J. Mol. Biol.*, **266**, 576–593.

Güntert, P., Braun, W., Billeter, M. and Wüthrich, K. (1989) *J. Am. Chem. Soc.*, **111**, 3997–4004.

Güntert, P., Braun, W. and Wüthrich, K. (1991) *J. Mol. Biol.*, **217**, 517–530.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273,** 283–298.

Karplus, M. (1959) *J. Chem. Phys.*, **30**, 11–15.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51–55.

Momany, F. A., McGuire, R. F., Burgess, A. W. and Scheraga, H. A. (1975) *J. Phys. Chem.*, **79**, 2361–2381.

Nilges, M., Clore, G. M. and Gronenborn, A. M. (1990) *Biopolymers*, **29**, 813–822.

Polshakov, V. I., Frenkiel, T. A., Birdsall, B., Soteriu, A. and Feeney, J. (1995) *J. Magn. Reson. B*, **108**, 31–43.

Saenger, W. (1984) Principles of Nucleic Acid Structure, Springer, New York, NY.

Sich, C., Ohlenschläger, O., Ramachandran, R., Görlach, M. and Brown, L. R. (1997) *Biochemistry*, **36**, 13989–14002.

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