Journal of Computer-Aided Molecular Design, 3 (1989) 195–210
ESCOM

J-CAMD 069

GEOM: A new tool for molecular modelling based on distance geometry calculations with NMR data

Michel Sanner[a], Armin Widmer[a], Hans Senn[a,*] and Werner Braun[b]

[a] *Sandoz AG, Preclinical Research, CH-4002 Basel, Switzerland*
[b] *Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule Zürich, CH-8093 Zürich, Switzerland*

Received 16 March 1989
Accepted 17 July 1989

Key words: Computer graphics; Distance geometry; Nuclear magnetic resonance; Cyclic structures; H-bonds; Cyclosporin A

SUMMARY

GEOM is a new graphics tool which allows the use of distance geometry to compute linear and cyclic structures typically arising in drug design situations. Modified amino acids or monomeric organic entities can be easily constructed in an interactive way and deposited in the library of the distance geometry program together with geometric information required for structure calculation in dihedral angle space. In addition, GEOM is able to produce all files needed to calculate a structure based on NMR data (NOE and J-coupling constraints) and it permits the graphic analysis and comparison of computed structures. The application of GEOM is demonstrated in three examples: modelling of cyclosporin A structures with and without a limited set of H-bond constraints and modelling of a cyclic hexapeptide with a full NMR data set.

INTRODUCTION

Nuclear magnetic resonance (NMR) methods, especially since the advent of new two dimensional experiments [1], have proven to be a powerful experimental technique to determine the three-dimensional structure of biological molecules in atomic detail in solution [2–5]. The experimentally measured data are nuclear Overhauser enhancements [6] and spin–spin coupling constants [7] which are semiempirically calibrated as proton-proton distance and dihedral angle constraints. These constraints are then used in distance geometry calculations [8–13] and/or restrained molecular dynamics calculations [14–19] to determine the three-dimensional structure in solution, without any a priori knowledge or hypothesis concerning the three-dimensional structure. This procedure does not rely on a proposed three-dimensional model to be examined for consistency with the experimental data. Computational tools such as the distance geometry program DISMAN [4, 20] produce an ensemble of possible solution structures which are consistent with the experimental data.

*To whom correspondence should be addressed.

0920-654X/89/$3.50

In using this computational tool for purposes of drug design, special problems often arise, e.g., occurrence of cyclic molecules [21–23] or chemically modified amino acid residues. We describe here a software package called GEOM which provides the NMR experimentalist with the necessary tools to perform distance geometry calculations with any covalent structure typically arising in drug design situations, and report on initial practical applications.

By use of the program GEOM, any covalent structure can be sketched by the user on a graphics screen and regularized by the program package to a standard geometry. This molecule can then be deposited as a building block into a library which is maintained and updated by the program. The format of the library is the same as the format used by the distance geometry program DISMAN. After the distance geometry calculations, the structures obtained can be analyzed by the user of the program on the basis of residual distance and dihedral angle violations. In addition, structural similarities can be examined. The user has several options to analyze these properties on a quantitative level in terms of tables, as well as on a qualitative graphical level where, e.g., several distance geometry structures may be superposed in a best-fit orientation and residual violations can be graphically indicated on the calculated structures. The program performs the necessary bookkeeping of the distance and dihedral angle constraints, so that the NMR data list can be easily updated.

The program package has been designed so that it can be used on several low cost graphics devices. Practical experience with GEOM is described in the creation of cyclic molecular structures of cyclosporin A, an immunosuppressive drug [24] and for a synthetic cyclic hexapeptide, where an experimental NMR data set was available. Details of the NMR data collection and structural analysis are described elsewhere.

SYSTEM DESIGN REQUIREMENTS

The layout of the system is shown in Fig. 1. GEOM can be considered as a graphics layer or interface which has to meet two specific requirements of DISMAN for structural calculation with experimental constraints.

The first design requirement for GEOM is that it should allow DISMAN the conformational computation of a large variety of molecular structures with distance and dihedral angle constraints as obtained from NMR spectroscopy [1–4]. So far DISMAN has exclusively been used with regular peptides and proteins containing natural amino acids. For this requirement it is essential that GEOM can provide the library of DISMAN (Fig. 1) molecular structures as monomers or as building blocks for polymers with all information needed for conformational analysis in dihedral angle space. The possibility of computing structures such as cyclic peptides or peptides with irregular side chains and backbone is required. Some restrictions in DISMAN concerning the backbone representation were overcome by treating a highly complex covalent structure as a side chain of a pseudo residue with dummy backbone atoms.

The second design requirement for GEOM is that the program can analyze structures, computed with DISMAN, with respect to residual distance and dihedral angle violations. GEOM has therefore to read the structures generated by DISMAN as well as all distance and angle constraints used in the calculation. Then, GEOM should graphically show where the experimental constraints are not fulfilled, how big the violations are, or, by fitting a set of structures together, where and to what extent they are similar or different.

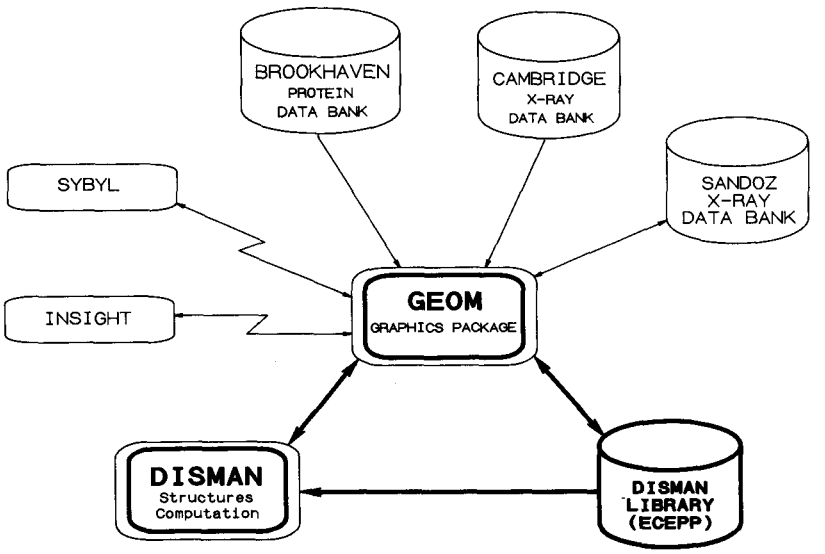

Fig. 1. The relationship of GEOM with the distance geometry program DISMAN, its library and various molecular graphics programs and data bases.

Finally, we wanted our system to be able to communicate with several in-house and external data bases such as the Brookhaven Protein Data Bank [25], the Cambridge X-ray Data Bank [26] and the SANDOZ in-house crystallographic data base.

To be independent of graphics work stations such as the Evans & Sutherland PS 350 or PS 390, GEOM is written for low cost terminals of the type TEKTRONIX. However, we required that GEOM should be able to communicate with our existing molecular modelling software, so that the structures could also be analyzed on graphics work stations with, e.g., SYBYL (Tripos Associates, St. Louis, MO 63144 [27]) or INSIGHT (Biosym Technologies Inc., San Diego, CA 92121).

DESCRIPTION OF GEOM

GEOM is a menu-driven program for the construction of three-dimensional molecular structures and for efficient communication with DISMAN (Fig. 1) and the ECEPP library [28].

The characteristic feature of GEOM is that graphic input of the structure is done in three dimensions. Simple editing functions such as addition or removal of an atom or a bond allow facile sketching of the molecule, which is then regularized in a second step with the help of molecular mechanics in order to satisfy the chemical constraints, e.g., bond length and bond angles. The molecule can be displayed in different orientations in line drawing, ball-and-stick and space-filling mode. Computed structures from experimental distance and dihedral constraints by means of DISMAN calculation can be analyzed and compared graphically by various GEOM functions. Of particular interest is GEOM's ability to communicate with the ECEPP library of DISMAN: structural elements can be graphically designed with all the information needed for distance geometry calculation in dihedral angle space.

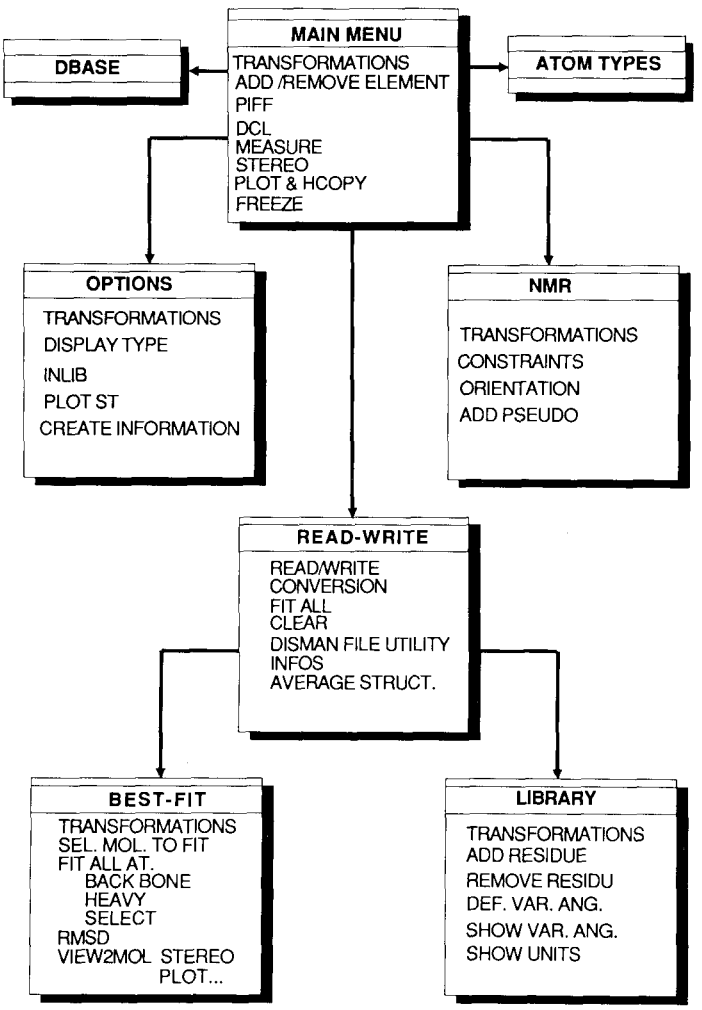

Fig. 2. Organization and hierarchy of the eight GEOM menus with their main functional features. The different menus are accessed from the main menu according to the arrows.

GEOM is written in FORTRAN and uses a graphics tablet or a mouse as graphics input device. The program can store 10 molecules simultaneously, each having up to 1000 atoms and up to 100 residues in the case of polypeptide structures.

The data structure of an atom is described by a record containing the following fields: name, type, atomic coordinates in Angström and display coordinates, number of neighbors, neighbors list, and number and name of residue to which it belongs. The display coordinates are related to the atomic coordinates by a scaling factor, a rotation and a translation.

The hierarchical menu structure of GEOM is shown in Fig. 2. The items in the menus refer either to specific functions or a group of functions.

The main menu

The main menu (Fig. 3) appears on screen when the program is started. Most of the functions required for editing a molecular structure are available in the main menu except for the choice of atom types which are selected in the ATYPES menu.

Graphic input of atoms is always done in a plane $Z = c$ where c can take any real value. The default value of c is 0. Through rotation or regularization, the atom's x, y and also z coordinates are changed. By selecting an atom α with coordinates (X_α, Y_α, Z_α) the plane $Z = Z_\alpha$ is selected and an atom β, neigbor of α, can be added in position (X_β, Y_β, Z_α). If the atom β is created before selecting the neighbor atom, a default value of $Z = 0$ will be assumed. The user can check the chirality of the built structure by using the ball-and-stick representation.

The functions called TRANSFORMATIONS (Fig. 2) are the scaling functions: **bigger** and **smaller**, the rotation functions around the three axes: **X-rot**, **Y-rot**, **Z-rot** and the translation function **move**. These functions are available in all menus except in read-write. The transformation always applies to the currently displayed molecule. For the rotations, the user can either input graphically an approximate value of the rotation angle or enter the value from the keyboard. By activating the **move** function twice, another molecule can be selected for display.

Atom used together with the move function allows movement of one atom at a time in a plane of constant z. All bonds connecting this atom will be stretched. This feature is useful for the specification of local stereochemistry or to pre-form the conformation before running the regularization program. **Piff** starts a molecular mechanics program to regularize the three-dimensional structure of the currently displayed molecule. This program is a modified version of PIMM [29]. **Anatom** accesses the plot program ANATOM which draws a structure in ball-and-stick or space-filling representation (Fig. 3). **Space** and **type** are toggles used to choose the representation type and the atom colors for ANATOM. **Show** provides various information about the selected atom, as atom number, atom type, atom name, number of neighbors, neighbors, and residue information. **Measure** calculates distances, bond angles and dihedral angles within the molecule. **Freeze** applies the current translation and rotation matrix to the atomic coordinates. **Twist** rotates a substructure of the molecule around a selected bond. **DCL** starts a VMS shell.

The ATYPES and DBASE menu

GEOM uses 33 different atom types which define the valencies and van der Waals radii. The same atom types are used by SYBYL. When an atom is added to a structure, its default type is Csp3. This type can be changed in the ATYPES menu. The Dbase menu performs substructure searches in the in-house crystallographic data base.

The READ-WRITE menu

This menu provides all functions for input/output operations. GEOM reads and writes files in the following formats: SYBYL's '**.mol**' format containing the molecule's atomic coordinates, DISMAN's '**.cor**' files providing the coordinates and the residue information and the '**.pdb**' format containing molecules stored in the Brookhaven Data Bank format. With the **res lib.** function a residue can be loaded from the library. This structure may be used as a starting point to build a new residue to add to the existing library.

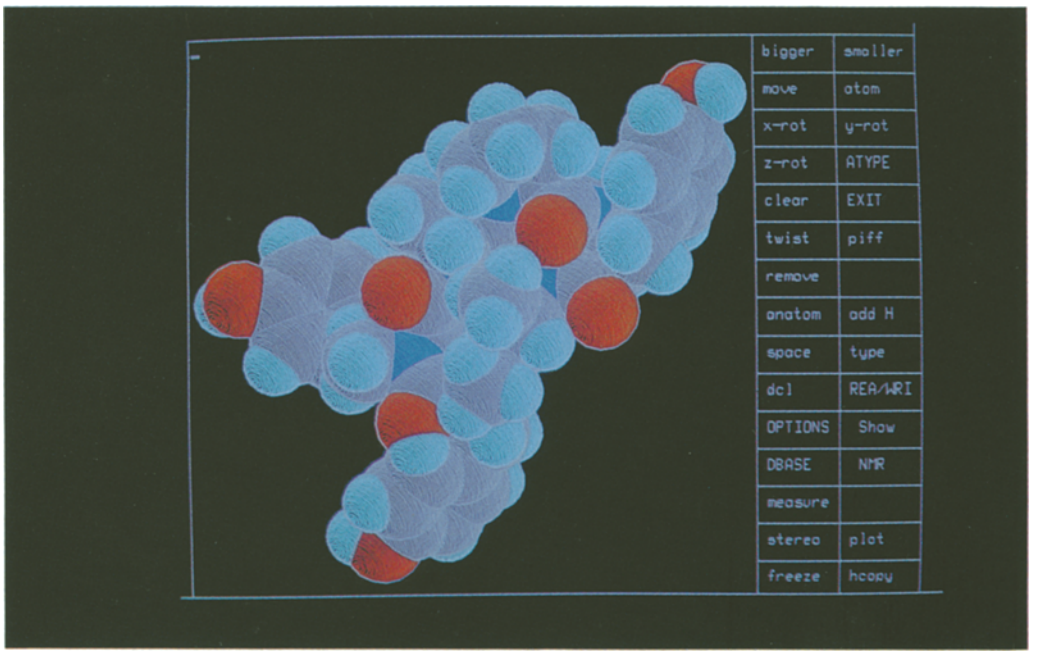

Fig. 3. Screen view when GEOM is in the main menu. The commands are listed on the right where all menu keys appear in capital letters. A cyclic hexapeptide structure, cyclo(–Pro–MeTyr–Ala–MeTyr–MeTyr–D-Ala–) which was built by GEOM and DISMAN, is shown in space-filling mode. O = red, N = blue, C = white, H = cyan.

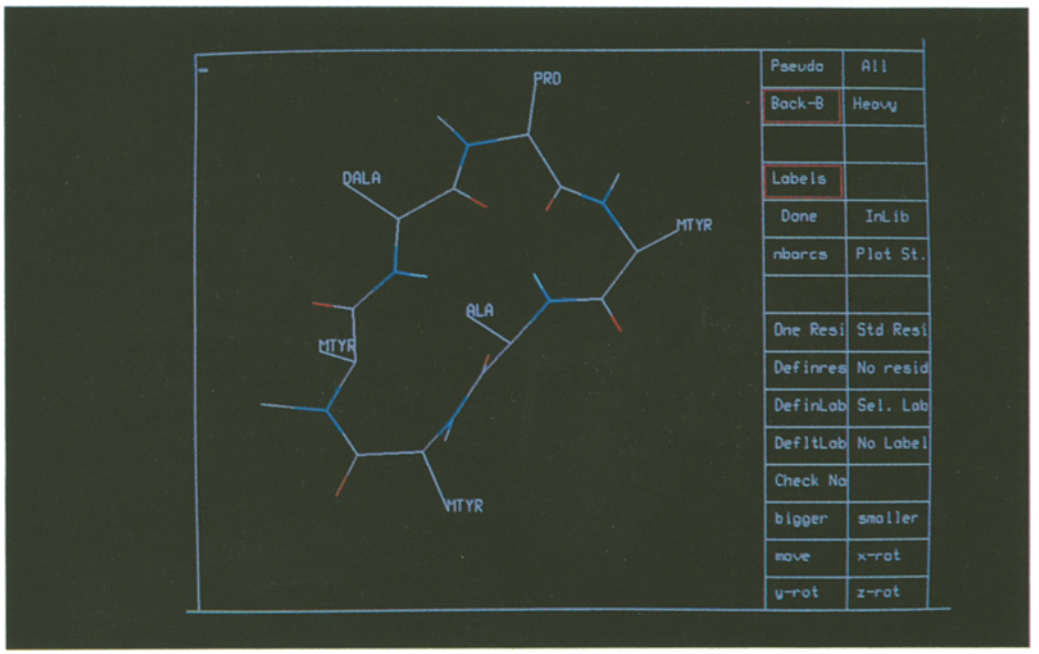

Fig. 4. Screen view of the GEOM menu OPTION which allows the recognition and orientation of structural details in complex molecules. The cyclic hexapeptide of Fig. 3 is shown in stick representation with the residue names at the C_β-side chain position. O = red, N = blue, C = white.

The communication between GEOM and DISMAN is handled by the user with the following DISMAN file utilities: **dist** reads or writes upper and lower distance constraints files and **jco** reads dihedral angles constraints files. **Sel. cor** does the bookkeeping of the structure files generated during a distance geometry calculation. It sorts the structures according to the value of the target function and creates a list of the best structures ordered according to the value of the target function [20].

Info provides some information about the molecules present in memory. Also in this menu the user can compute the average coordinates of optimally superposed structures.

The OPTION menu

There are two main function sets in this menu. The first set is related to easy recognition and orientation of structural details. Figure 4 shows an example of a representation of a hexapeptide with all side chains replaced by their residue names. The structure can also be displayed with selected side chains only, or in the heavy atoms mode.

Different types of file do not provide the same information. With the second set of functions, missing information can be defined for file format conversion. For instance, the '**.mol**' format used by SYBYL contains no information on residue type. The function **one resi** defines the whole molecule as one residue. With the **Def Res** function, the user can define a subset of atoms in the molecule as a residue. **Std Resi** searches, on the base of chemical connectivity, for an α-amino acid structure. The user can add or remove atoms or pseudo atoms from this residue.

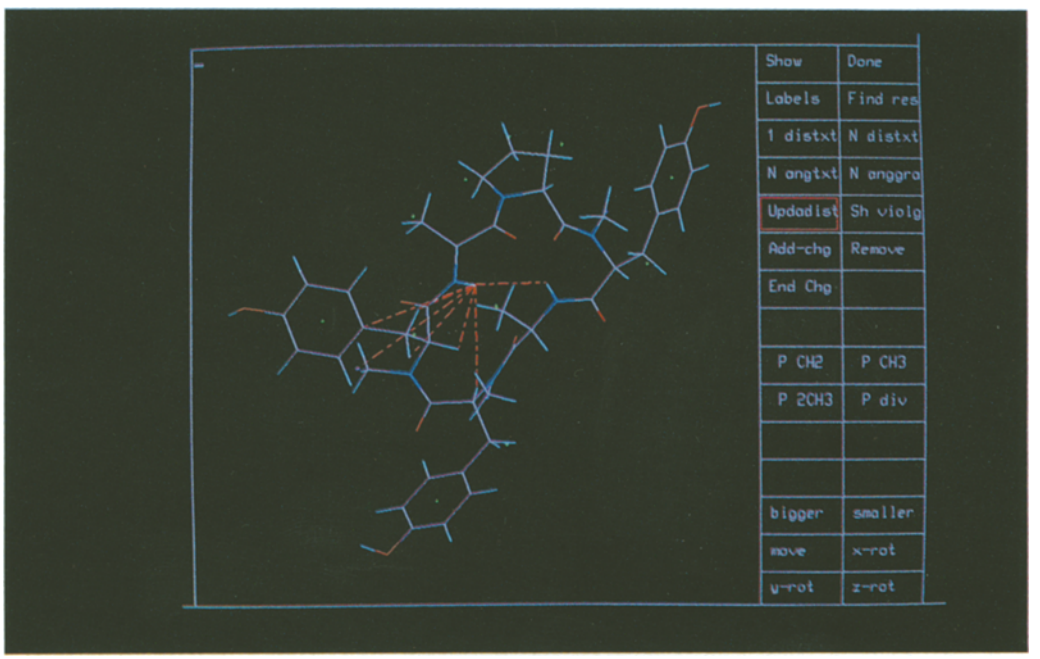

Fig. 5. Screen view of the NMR menu. The cyclic hexapeptide of Fig. 3 is shown. Red dotted lines indicate all distance constraints to the selected atom, here the HN-proton of D-Ala^6, which were experimentally observed. O = red, N = blue, C = white, H = cyan, pseudo atoms = green.

The NMR menu

The function of this menu (Fig. 5) is twofold. First, to add pseudo atoms [30] to a molecule. **P CH2, P CH3** and **P 2CH3** add a pseudo atom for the methylene, methyl and the two methyls of the isopropyl group, respectively. With the **P div** function the user can define any set of atoms which should be represented by a pseudo atom. The pseudo atom is always added at the centroid of the selected atoms.

The second set of commands is used to analyze the structure in relation to the NMR dihedral angle and distance constraints. The constraints set can be updated graphically as shown in Fig. 5. The **Updadist** function allows addition, removal or change of the value of a constraint (**Add-chg, Remove, End chg**). When an atom is selected, red dotted lines show all atoms linked to that atom by a distance constraint. Distance constraints can also be displayed in text mode with the value of the violation. With **N distxt** and **1 distxt** all or a specific distance constraint(s) can be displayed in text mode. Using **N angtxt** and **N anggra** dihedral angle constraints can be viewed in text and graphics mode. The **Sh violg** function gives a graphic overview of all violations. The molecule is drawn in white, and red arrows show the vectorial sum of all violations on each atom (see Fig. 9).

For the orientation of the user **Labels** displays the residue names and **Find res** retrieves a residue by its name and number and translates the molecule, so that this residue is in the center of the screen.

The BEST-FIT menu

This menu allows superposition of molecules to get a best fit [31]. Standard subsets of atoms like **heavy** or **backbone** atoms, can be used for the best-fit operation. With the **Select** function the user can manually define corresponding pairs of atoms within the structures to be fitted. The two superposed molecules can be displayed and viewed in different orientations.

The LIBRARY menu

Programs for distance geometry calculation in dihedral angle space (e.g., DISMAN) depend on a library to build the molecular structure. To use the method for rapid calculation of derivatives of the target function with respect to dihedral angles [32], this library has to provide information about units of atoms. A unit is a set of atoms of which the relative three-dimensional position is not affected by changing any dihedral angle within the molecule (Fig. 6). GEOM searches automatically for the unit structures of the constructed molecule and deposits this information, together with geometric data, in ECEPP format into the library. All variable dihedral angles have to be defined by the user.

APPLICATIONS

(1) Modelling of cyclic peptide and peptolide structures with nonstandard amino acid residues

Cyclosporin A is an example of a cyclic undecapeptide containing nonstandard amino acids. Its structure is the following: cyclo(–MeBmt[1]–L-Abu–Sar–L-MeLeu–L-Val–L-MeLeu–L-Ala–D-Ala–

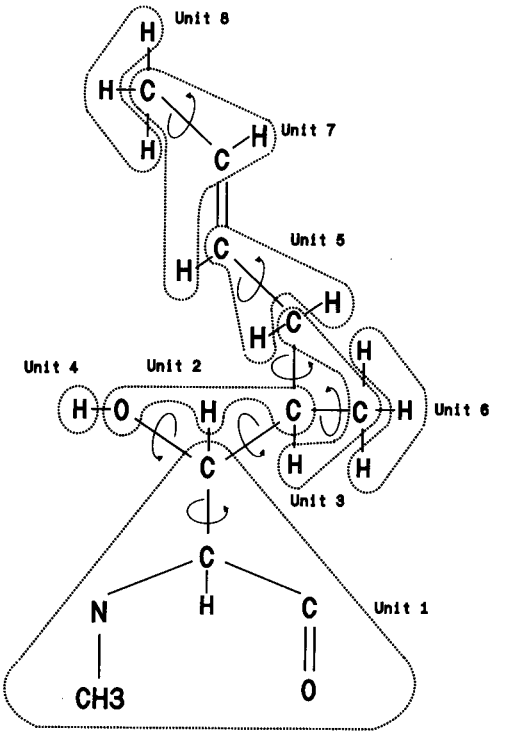

Fig. 6. Unit structure of MeBmt ((4*R*)-4[*E*)-2-butenyl]-4,-*N*-dimethyl-L-threonine) as found by GEOM for the variable side chain dihedral angles. The units for the backbone dihedral angles φ and ψ are defined internally in DISMAN.

L-MeLeu–L-MeLeu–L-MeVal^{11}–) where Abu, MeBmt and Sar represent α-aminobutyric acid, (4*R*)-4[(*E*)-2-butenyl]-4,-*N*-dimethyl-L-threonine and *N*-methyl glycine, respectively. Me denotes the substitution of the amide proton in the amino acids by a methyl group. In iso-cyclosporin the peptide bond between L-MeVal^{11} and L-MeBmt^{1} is replaced by an ester bond to the β-hydroxy group of MeBmt^{1}.

GEOM's potential was fully utilized to provide all the monomeric building blocks to the library of DISMAN. A linear analogue of the cyclic structure is formed from the monomers with MeBmt^{1} at the N-terminal and L-MeVal^{11} at the C-terminal end. Geometric cyclization between CO of L-MeVal^{11} and NH-α or O-β of MeBmt, respectively, is then achieved by DISMAN calculation where nine distance constraints are used as upper and lower limits to define the peptide bond (Table 1), whereby the backbone torsion angle ω is constrained to 180° by eight fixed distances across the peptide bond.

Table 1 shows a statistic from a DISMAN cyclization with 100 randomly generated linear starting conformations. Thirty percent of all computed (pseudo-)cyclic structures (peptide bond MeVal^{11}-MeBmt^{1} is covalently not closed) have a torsion angle within the range $160° < |\omega| < 180°$ and, at the same time, residual distance violation of the C'^{11}-N^{1} bond length less than 0.1 Å. Nine of these structures even show bond length violations of only +/− 0.01 Å. At the end of the DISMAN cyclization, the van der Waals contact terms in the target function [20] were equally

TABLE 1
DISTANCE CONSTRAINTS, AVERAGE DISTANCE AND RESIDUAL VIOLATION AFTER DISMAN CYCLIZATION, STARTING FROM 100 CONFORMATIONS OF CYCLOSPORIN A[a]

Atoms of		Constraints[b]	Average Violation[d]	Maximum Violation	Average distance	Violation in best structure
MeBMT[1]	L-Val[11]					
N	C	1.36	0.12	0.31	1.42	0.00
N	O	2.27	0.04	0.21	2.23	0.03
N	CA	2.48	0.14	0.52	2.54	0.02
CA	C	2.41	0.09	0.40	2.49	0.01
CA	O	2.72	0.07	0.54	2.75	0.05
CA	CA	3.80	0.03	0.31	3.78	0.03
QCN[d]	C	2.77	0.05	0.16	2.74	0.04
QCN[d]	O	3.93	0.01	0.03	3.80	0.02
QCN[d]	CA	3.03	0.15	0.55	3.15	0.14
ω (deg)		180				1

[a]The violations in the best structure are also tabulated. The used constraints define a trans peptide bond ($\omega = 180°$) between MeBmt[1] and Val[11]. All values are given in Å.
[b]Distances obtained from X-ray structure [22].
[c]Average over all 100 calculated structures.
[d]Pseudo atoms for *N*-methyl group.

weighted as the cyclization constraints so that only stereochemically correct cyclic structures were obtained. A superposition of the backbone of the 10 best cyclic conformations with each other showed a quite large RMS deviation ranging from 1.8 Å to 3.1 Å. This means that many different but stereochemically correct conformations of cyclosporin A are obtained.

Figure 7 illustrates the backbone of a structure before and after cyclization. The starting conformation (red) was generated by DISMAN with randomly chosen variable dihedral angles. Distance geometry calculation was performed with the backbone and all side-chain atoms of cyclosporin A (232 atoms). The boxed atoms (C′ of MeVal[11] and NH of MeBmt[1]) show the peptide bond closed by distance constraints. The amide bond has a C′-N bond length of 1.36 Å and the backbone torsion angle ω is 178.9°. The geometrically correct but covalently open bond between MeVal[11] and MeBmt[1] can now easily be closed by GEOM.

For iso-cyclosporin A a statistic of more than one hundred computed structures gives similar results as observed above: 37% of all structures have an ω torsion angle in the range of $160° < |\omega| < 180°$ and, at the same time, a residual distance violation of the C′[11]-Oβ ester bond of less than 0.1 Å. Twelve of these structures violate this bond length with +/− 0.01 Å.

The average CPU time to compute one of these structures is 7.5 min on a VAX 8530. By decreasing the number of refinement steps in computing van der Waals interactions this value can be reduced approximately by a factor of 2.

(2) Model studies of cyclosporin A conformations with given H-bond network

Distance constraints to model the four H-bonds between Abu[2](NH)-Val[5](CO), Val[5](NH)-

TABLE 2
DIFFERENT CYCLOSPORIN CONFORMERS GENERATED BY DISTANCE GEOMETRY CALCULATION WITH CONSTRAINTS FOR CYCLIZATION AND FOUR H-BONDS (SEE TEXT)[a]

Structure No.	1A	2A	3A	4A	5A	1B	2B	3B	4B	5B
H-bond										
Average violation (Å)	0.14	0.18	0.17	0.18	0.18	0.29	0.23	0.21	0.23	0.29
Maximum violation (Å)	0.47	0.34	0.27	0.26	0.48	0.72	0.57	0.52	0.54	0.74
Cyclization[a] Val[11]-MeBmt[1]										
C′–N	1.39	1.36	1.34	1.40	1.33	1.37	1.39	1.39	1.37	1.39
ω (deg)	−146	146	−155	138	141	−172	163	174	166	170
Best fit X-ray										
RMSD[c]	1.53	2.29	3.17	1.79	1.56	2.54	1.65	1.75	1.25[d]	3.02

[a]Structures with lowest residual distance violations for the H-bond (xA) and with best cyclization and lowest target function error (xB) are selected from 90 DISMAN generated structures.
[b]Constraints C′–O = 1.36 Å, $\omega = 180°$ defined by eight distance constraints.
[c]Superposition with backbone of crystal structure.
[d]Lowest RMSD value of all 90 structures.

Abu[2](CO), Ala[7](NH)-MeVal[11](CO) and D-Ala[8](NH)-MeLeu[6](CO) are taken from the X-ray structure [24] and used together with the cyclization constraints (Table 1) as upper and lower boundaries for DISMAN calculation. Each H-bond is defined by two distances, O-N and O-NH. The constraints for cyclization are weighted with respect to the H-bond by a factor of 15 to ensure proper ring closure. The DISMAN calculation was done in four levels. The first three levels included distances between residues i,i+2; i,i+4 and i,i+10, respectively. At the end of the distance geometry calculation the weight of the van der Waals interactions was increased from 0.2 to 1.0.

The results in Table 2 highlight the usefulness of the distance geometry generated structures in terms of efficient generation of conformers which closely fulfil the constraints. From 90 calculated conformations about 20 are in good agreement with the constraint data and at the same time show correct van der Waals interactions. However, the ring conformations obtained are remarkably different, e.g., the RMS deviations from a one-to-one superposition (backbone) of the ten selected structures shown in Table 2 range from 1.37 Å to 3.69 Å. The backbone of all 90 calculated cyclosporin conformations was compared to the X-ray backbone structure. The best fit was found with conformation 4B of Table 2 (RMSD 1.25). Figure 8 shows a superposition of the backbone of this structure with that of the crystal structure. The two backbones differ mainly at three positions: the peptide bonds Sar[3]–MeLeu[4], Ala[7]–D-Ala[8] and MeLeu[10]–MeVal[11] of structure 4B are flipped by about 60°, 90° and 180°, respectively.

The average CPU time to compute one of these structures is 11.1 min on a VAX 8530. By decreasing the number of refinement steps in computing van der Waals interactions, this value can be reduced by a factor of about 2.

(3) Modelling with experimental constraints obtained from NMR

The ^1H and ^{13}C NMR spectra of the cyclic hexapeptide cyclo(–L-Pro–MeTyr–L-Ala–MeTyr–

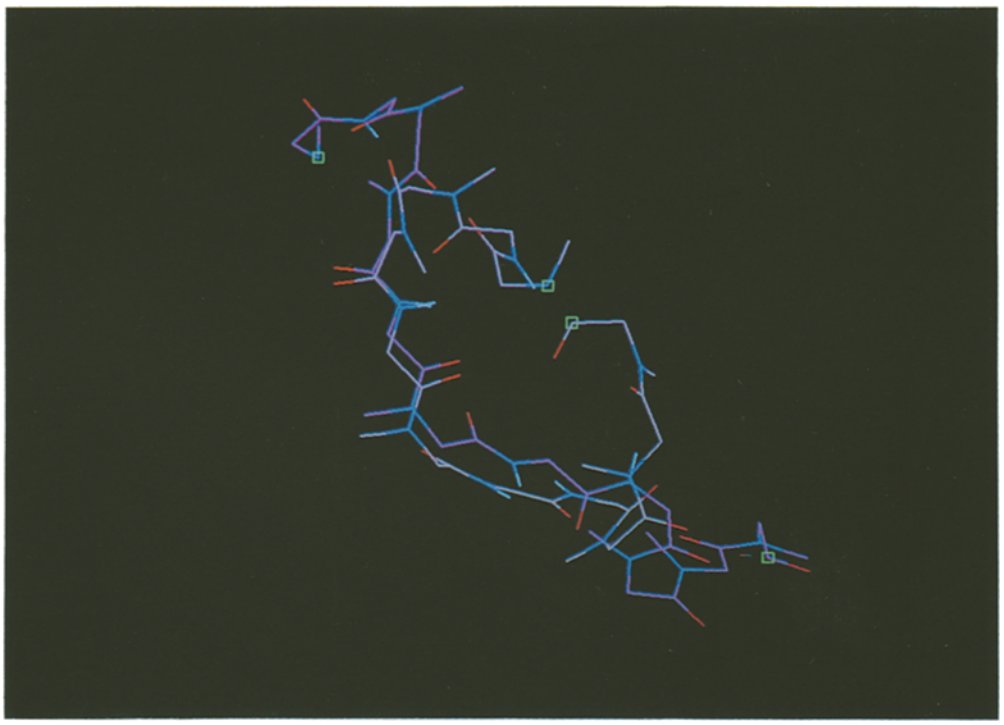

Fig. 7. Backbone of cyclosporin A before (red) and after (white) DISMAN cyclization. The boxed atoms show the C′ of L-Val^{11} and N-α of MeBmt^{1} where a peptide bond is geometrically defined with distance constraints (Table 1). The cyclic structure has a C′–N bond length of 1.362 Å (constraint 1.36 Å) and a backbone torsion angle ω of 178.9° (confined to 180° by eight distance constraints).

MeTyr–D-Ala–), have been assigned and 37 distance and 4 dihedral angle constraints extracted from the NOE and vicinal coupling data [unpublished]. The nonstandard residue MeTyr was constructed and added to the library of DISMAN by GEOM. The linear analog Pro^{1}–MeTyr–Ala–MeTyr–MeTyr–D-Ala^{6} was used as the starting structure for distance geometry calculation with a total of 41 constraints (distance and torsion angle). As in the first application, ring closure was ensured by nine cyclization constraints. The residual distance violations of the structures obtained after DISMAN minimization were analyzed by GEOM.

A graphic representation of one of these structures is shown in Fig. 9. The red arrows attached to individual atoms point in the direction in which the atom should be moved to further minimize the residual distance constraints involving this atom. The length of the arrows is a measure of the sum of residual violations for a particular atom. The conformation shown in Fig. 9 fulfils the 37 distance and 9 cyclization constraints with an average violation of 0.052 Å and 0.017 Å for the cyclization and NMR constraints, respectively.

DISCUSSION

Distance geometry in combination with restrained molecular mechanics or dynamics calculations is the method of choice to determine the conformations of polypeptides and proteins from

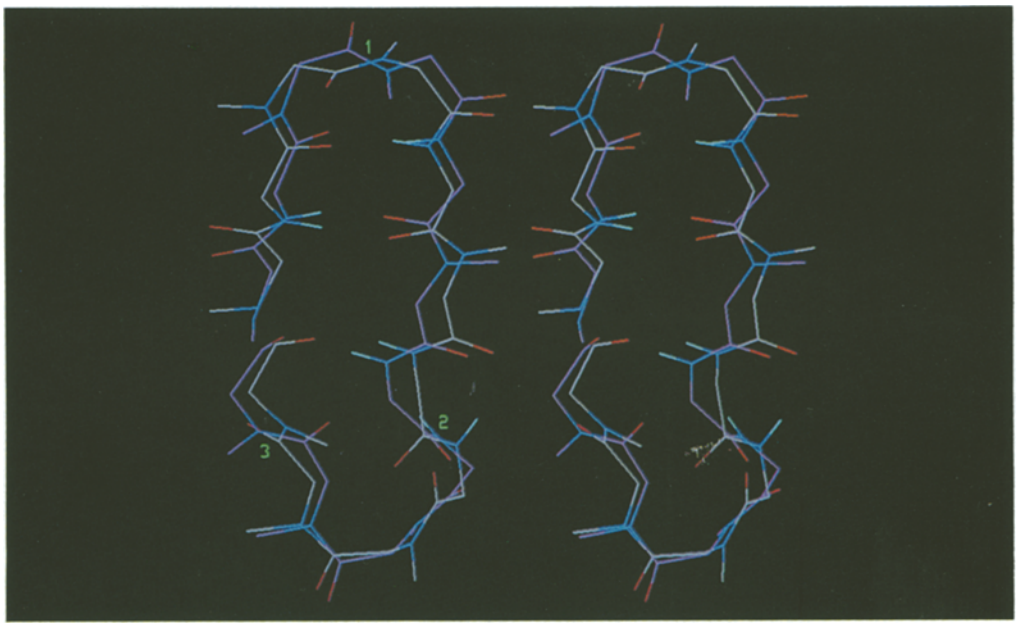

Fig. 8. Superposition of the DISMAN-generated ring conformation 4B (see Table 2) with the X-ray structure of cyclosporin A backbone (RMSD = 1.25 Å). Structure 4B (red) shows the best fit from 90 generated conformations.

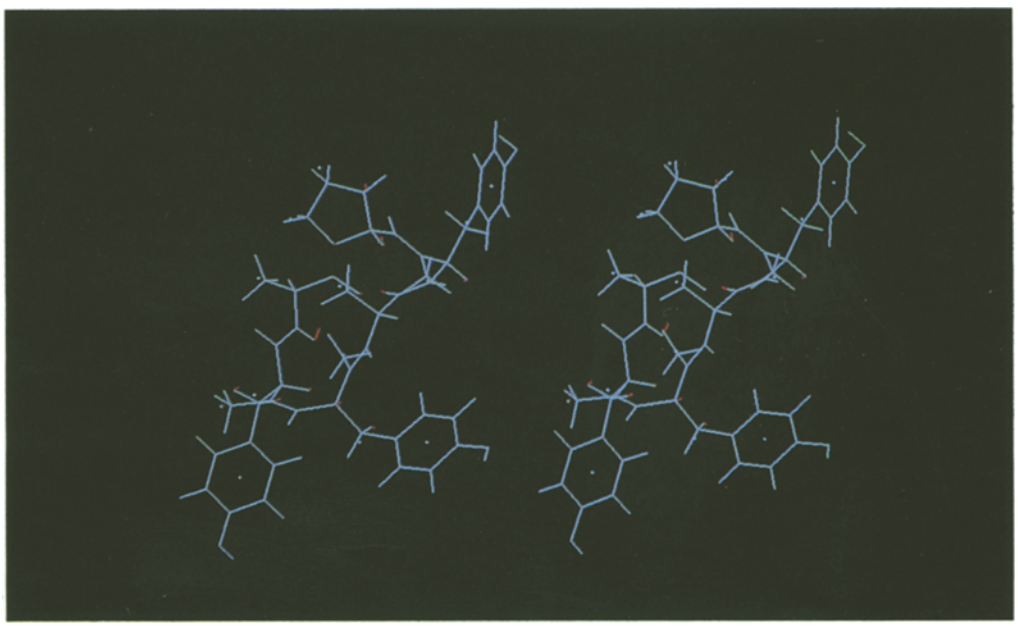

Fig. 9. Conformation of the cyclic hexapeptide of Fig. 3 generated with experimentally measured NMR constraints and cyclization conditions between D-Ala^6 and Pro^1 for a trans peptide bond. The red arrows show residual distance violations and point in the direction where the atom has to be moved to fulfil the boundary conditions.

NMR data [4, 5, 33]. Using GEOM, calculations with the distance geometry program DISMAN [20] can be easily extended to molecules with highly complex covalent structures. The user of GEOM does not need to know the internal structure of DISMAN: he can use his chemical intuition to sketch a molecule on the screen and deposit this structure in the library used by DISMAN. In addition, computed structures can be graphically compared and analyzed in close connection with NMR data. A tool providing nearly the same analytical features, to examine NMR structures on a graphics workstation (PS 3xx), has previously been described [34]. The advantage offered by GEOM is that it operates on low cost terminals such as VT100, NEC with TNT software (TRIPOS NEC TERMINAL SOFTWARE) or TEKTRONIX 41xx. The uniqueness of GEOM is that it expands the application of distance geometry calculation to nonpeptidic molecules of high interest in pharmacology. For Allinger's molecular mechanics program MM2 [35], a similar tool has been described [36].

The aim of distance geometry calculations is to characterize the ensemble of all conformations consistent with experimental data derived from NOE measurements. The absolute configuration of the amino acids is usually known, e.g., in regular polypeptide or protein structures. However, in some applications to drug design problems, the configuration might not be known a priori and, in this case, could be potentially determined by the experimental NMR data. Some approaches have been described so far [37–39] to deal with such situations. GEOM's potential to generate different types of building blocks with all the information needed for DISMAN calculations makes it a promising new tool to explore configurational aspects in organic and bioorganic diastereomers with NMR data.

Application of GEOM combined with DISMAN has been used to model cyclic structures. These need special consideration, because only particular combinations of torsion angles can form cyclic structures. Exhaustive methods for grid search can only be applied in practice to small rings, because of the exponentially increasing number of conformations which have to be considered [40]. By carefully checking the van der Waals radii and using experimental data early in the tree search, the number of conformations to be looked at can be dramatically decreased [21]; but it has still not been shown whether the reduction in the number of conformations is high enough for practical application with typical NMR data sets.

Gō and Scheraga [41, 42] gave an exact analytical solution of the conditions to close a ring. In order to form a six-membered ring, they showed how the conditions for the torsion angles could be reduced to one equation which could be solved numerically. This meant that in a cyclic structure with n variable torsion angles, an exhaustive search had to be performed only in a conformation space with $n-6$ degrees of freedom. However, this reduction was not enough for medium-sized rings.

The method we have applied is also not exhaustive, but it definitely explores a much greater variety of initial structures than has been previously possible [22]. The procedure is similar to the approach followed by Tonge et al. [23], where sequential and short range information are carefully examined to discard forbidden conformations, but has the advantage that most of the selection is done automatically by the DISMAN program through the variable target function procedure.

ACKNOWLEDGEMENTS

We would like to thank Dr. Martin Heller for providing the molecular graphics program ANA-

TOM, Dr. Paul Taylor at the University of Edinburgh for writing the interface to the in-house crystallographic data base and Dr. T. Petcher and H.R. Loosli for helpful discussions.

REFERENCES

1 Ernst, R.R., Bodenhausen, G. and Wokaun, A., Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon Press, Oxford, 1986.
2 Wemmer, D.E. and Reid, B.R., Ann. Rev. Phys. Chem., 36 (1985) 105–137.
3 Wüthrich, K., NMR of Proteins and Nucleic Acids, Wiley, New York, 1986.
4 Braun, W., Quart. Rev. Biophys., 19 (1987) 115–157.
5 Clore, G.M. and Gronenborn, A.M., Protein Eng., 1 (1987) 275–288.
6 Noggle, J.H. and Schirmer, R.E., The Nuclear Overhauser Effect, Academic Press, New York, 1971.
7 Karplus, M., J. Am. Chem. Soc., 85 (1963) 2870–2871.
8 Braun, W., Bösch, C., Brown, L.R., Gō, N. and Wüthrich, K., Biochim. Biophys. Acta, 667 (1981) 377–396.
9 Braun, W., Wider, G., Lee, K.H. and Wüthrich, K., J. Mol. Biol., 169 (1983) 921–948.
10 Arseniev, S.A., Kondakov, V.I., Maiorov, V.N. and Bystov, V.F., FEBS Lett., 165 (1984) 57–62.
11 Williamson, M.P., Havel, T.F. and Wüthrich, K., J. Mol. Biol., 182 (1985) 295–315.
12 Schultze, P., Wörgötter, E., Braun, W., Wagner, G., Vasák, M., Kägi, J.H.R. and Wüthrich, K., J. Mol. Biol., 203 (1988) 251–268.
13 Kline, A.D., Braun, W. and Wüthrich, K., J. Mol. Biol., 204 (1988) 675–724.
14 Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and Van Gunsteren, W.F., J. Mol. Biol., 182 (1985) 179–182.
15 Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M., Proc. Natl. Acad. Sci. U.S.A., 83 (1986) 3801–3805.
16 Clore, G.M., Sukumaran, D.K., Nigles, M., Zarbock, J. and Gronenborn, A.M., EMBO J., 6 (1987) 529–537.
17 Clore, G.M., Gronenborn, A.M., Kjear, M. and Poulsen, F.M., Protein Eng., 1 (1987) 305–311.
18 Kessler, H., Bats, J.W., Griesinger, Ch., Koll, S., Will, M. and Wagner, K., J. Am. Chem. Soc., 110 (1988) 1033–1049.
19 Kessler, H., Griesinger, Ch., Lautz, J., Müller, A., Van Gunsteren, W.F. and Berendsen, H.J.C., J. Am. Chem. Soc., 110 (1988) 3393–3396.
20 Braun, W. and Gō, N., J. Mol. Biol., 186 (1985) 611–626.
21 Smith, G.M. and Veber, D.F., Biochem. Biophys. Res. Commun., 134 (1986) 907–914.
22 Marion, D., Genest, M., Caille, A., Peypoux, F., Michel, G. and Ptak, M., Biopolymers, 25 (1986) 153–170.
23 Tonge, A.P., Murray-Rust, P., Gibbons, W.A. and Mclachlan, L.K., J. Comput. Chem., 9 (1988) 522–538.
24 Loosli, H.R., Kessler, H., Oschkinat, H., Weber, H.P., Petcher, T.J. and Widmer, A., Helv. Chim. Acta, 68 (1985) 682–704.
25 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535–542.
26 Allen, F.H., Bellard, S., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.J., Kennard, O., Motherwell, W.D.S., Rodgers, J.R., Watson, D.G., Acta Crystallogr., B35 (1979) 2331.
27 Naruto, S., Motoc, I., Marshall, G.R., Daniels, S.B., Sofia, M.J. and Katzenellenbogen J.A., J. Am. Chem. Soc., 107 (1985) 5262–5270.
28 Nemethy, G., Pottle, M.S. and Scheraga, H.A., J. Phys. Chem., 87 (1983) 1883–1887.
29 Lindner, H.P., Tetrahedron, 30 (1974) 1127.
30 Wüthrich, K., Billeter, M. and Braun, W., J. Mol. Biol., 169 (1983) 949–961.
31 Kabsch, W., Acta Crystallogr., A34 (1978) 827–828.
32 Abe, H., Braun, W., Noguti, T. and Gō, N., Comput. Chem., 8 (1984) 239–247.
33 Kaptein, R., Boelens, R., Scheek, R.M. and Van Gunsteren, W.F., Biochemistry, 27 (1988) 5389–5395.
34 Billeter, M., Engeli, M. and Wüthrich, K., J. Mol. Graph., 3 (1985) 79–83.
35 Berkert, U. and Allinger, N.L., Molecular Mechanics, ACS Monograph No. 177, American Chemical Society, Washington, D.C., 1982.
36 Saunders, M. and Jarret, R.M., J. Comput. Chem., 7 (1986) 578–588.
37 Jeffs, P.W., Müller, L., DeBrosse, Ch., Heald, S.L. and Fisher, R., J. Am. Chem. Soc., 108 (1986) 3063–3075.

38 Haruyama, H. and Kondo, M., Chem. Pharm. Bull., 35 (1987) 170–181.
39 Senn, H., Billeter, M. and Wüthrich, K., Eur. Biophys. J., 11 (1984) 3–15.
40 Taylor, J.S., Garret, D.S. and Wang, M.J., Biopolymers, 27 (1988) 1571–1593.
41 Gō, N. and Scheraga, H.A., Macromolecules, 3 (1970) 178–187.
42 Dygert, M., Gō, N. and Scheraga, H.A., Macromolecules, 8 (1975) 750–761.