Title	DETERMINATION OF THREE-DIMENSIONAL STRUCTURES OF POLYPEPTIDE IN SOLUTION BY 1H-NMR AND DISTANCE GEOMETRY
Author(s)	大久保, 忠恭
Citation	大阪大学, 1988, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/1204
rights	
Note	

DETERMINATION OF THREE-DIMENSIONAL STRUCTURES OF POLYPEPTIDE IN SOLUTION BY ^{1}H-NMR AND DISTANCE GEOMETRY

DETERMINATION OF THREE-DIMENSIONAL STRUCTURES OF POLYPEPTIDE IN SOLUTION BY 1H-NMR AND DISTANCE GEOMETRY

A Dissertion Submitted to the Faculty
of Science, Osaka University
in Fulfillment of the Degree of
Doctor of Science

Tadayasu Ohkubo

Division of Molecular Biophysics
Institute for Protein Research
Osaka University

CONTENTS

General Introduction

This work provides an approach to determine the three dimensional structure of protein and polypeptide in solution by a combined use of nuclear magnetic resonance spectroscopy (NMR) and distance geometry.

There has been a great demand to establish the method for obtaining information concerning the solution structures of macromolecules of biological interest. The knowledge of the three dimensional structure is sometimes decisive or inevitable to understand their biological functions. X-ray crystallography has long been used as the one and only technique to elucidate the structures of macromolecules represented by the atomic coordinate[1]. Although it is needless to mention that X-ray analysis has made a great contribution to advances in molecular biology[2,3], this technique involves essentially a limit that it treats single crystals of macromolecules. There has not been any universal technique to crystallize macromolecules. A condition to obtain a large enough single crystal of a good quality is found by trial and error in each case. Especially it is so difficult to crystallize polypeptide with around 20 amino acids residues that there has been almost no report of the work of X-ray analysis about polypeptide with a degree of polymerization of this range. There is another disadvantage that, even an X-ray analysis has been done on some macromolecule, there remains an inherent question if the structures in solid

state are preserved in solution state where biological molecules show their activities.

Recently, by the development of high resolution NMR with high magnetic field of super conductive magnets and new pulse techniques including various modes of two dimensional NMR experiments[4,5], the amounts of conformational information about biomolecules in solution have been increased drastically[6]. It is remarkable that information of interproton distances in macromolecules have been available through nuclear Overhauser effect (NOE) measurements[7].

Table 1

Approaches to tertiary structure in solution

Manual method	
1. Model building	: CPK model[13]
2. Interactive computer graphics	: CONFOR[14]
Computer calculation	
3. Metric matrix	: DISGEO[15-18]
4. Least-squares minimization (Variable target function)	: DADAS[19-23]
5. Restrained molecular dynamics	: CHARMM[24-27]

There have been several approaches including ours in these three years to construct three dimensional structures of macromolecules with the informations of interproton distances[6,8,9]. The strategies used to realize these approaches are listed in Table 1. They are originated from a basic method named as distance geometry algorithm[10,11] which is a mathematical way to elucidate the atomic coordinates of the molecule from the interatomic distances. Crippen related the interatomic distances to the atomic coordinates by an elegant concept of metric matrix, G, as follows.

The metric matrix for an actual molecule with N atoms is defined by

$$G_{ij}=r_i \cdot r_j \quad (i,j=1, \cdots, N)$$

where r_i represents the vector from the geometric center of mass to atom i and $\cdot$ the dot product. Thus G_{ij} is an N x N symmetric matrix. The matrix G_{ij} can be calculated from the distances d_{ij} as

$$G_{ii} = \frac{1}{N}\sum_{j}^{N} d_{ij}^2 - \frac{1}{2N^2}\sum_{j}^{N}\sum_{k}^{N} d_{jk}^2$$

$$G_{ij} = (G_{ii} + G_{jj} - d^2_{ij})$$

A spectral representation of metric matrix by diagonalization is,

$$G_{ij} = \sum_{\alpha}^{N} \lambda_\alpha E_{i\alpha} E_{j\alpha}$$

where λ_α is the eigenvalue and E_α is the corresponding

eigenvector.

A comparison of two different representations of G_{ij} leads to

$$r^{\alpha}_{i} = \sqrt{\lambda_{\alpha}}\, E_{i\alpha}$$

If all distances between all pairs of atoms are exactly known, the metric matrix should have properties of rank 3. This means that number of non-zero eigenvalues is three. Then coordinate of atom (x_i, y_i, z_i) is embedded as

$$x_i = \sqrt{\lambda_1}\, E_{i1},\ y_i = \sqrt{\lambda_2}\, E_{i2},\ z_i = \sqrt{\lambda_3}\, E_{i3}$$

This procedure is usually named as embedding.

Therefore the atomic coordinates can be calculated from the interatomic distances. It is noteworthy to mention here that there is a serious problem to determine the solution structure of protein and polypeptide from experimentally obtained distance information by ^{1}H-NMR. That is, elucidated solution structures are not expected to be unique ones[12]. Mainly three reasons might be considered for this problem. First, enough number of the NOE data with good qualities, are not available because of limited resolution of NMR spectrum such as overlapping and broadening of peaks and noise, even in two-dimensional NMR. Second, strict treatments in the interpretation of NOE data to their corresponding distances where the fluctuation of distances is taken into account should lead a loss of some information; i.e. the upper distance constraints might overestimate. Third, since the conformation of macromolecule should fluctuate in

solution, parameters are averaged in the time scale of NMR. Therefore one should expect that NMR structure in solution might be a group of the somewhat different conformations which satisfy the imposed distance constraints to some extent.

Among the strategy listed in Table 1, the model building with suitable molecular models such as CPK model[13] is the simple and conventional way to survey structures which satisfy the distance constraints. It is obvious that there is a limit in manual handling of the enormous data. Interactive computer graphics improves the situation to some extent. For example, CONFOR[14] is able to display the sum of the violations of the distance constraints and has facilities to accept a manual adjustment to minimize the violation. Even so, these techniques have a limit also that they can not survey systematically the all possibilities nor give the quantitative evaluation of the resulting conformation. Thus the other three kinds of approaches listed in Table 1 have been considered to be much reliable. The approach of DISGEO[15-18] follows the idea of the distance geometry mentioned above. The other two methods[19-23] are the minimization of the difference between interatomic distances in calculated structures and the corresponding distance constraints.

We have completed the method of the minimization of the variable target function; DADAS[19-23], which uses the dihedral angles in peptide as variables to minimize the difference. In this thesis, I will describe in sequence the method which was developed by us to determine the solution structure of polypeptide with 10 - 20 amino acid residues and the evaluation of the methods listed in Table 1.

In Chapter 1, the application of DADAS on the structural analysis of polypeptide in solution is reported. The sample was a shorter peptide of a heat-stable enterotoxin (STh) produced by a human strain of enterotoxigenic Escherichia coli (SK-1) designated as STh[6-19]. We investigate there the validity of the interatomic distances provided by interpreting of NOE's which were observed in the two-dimensional 1H-NMR spectra. Considering the intramolecular motion, we used the three methods to interpret the intensity of NOE to corresponding interatomic distance, i.e. use of rigid model, use of uniform averaging model and setting the maximum upper distance constraint. Various criteria such as averaged root mean square distance between conformers and violation of distance constraints were provided to evaluate the obtained structures. The result revealed that the segment between 9th to 19th residues of STh takes a well defined cylindrical structure. This is the first determination of structure of polypeptide in solution.

In Chapter 2, an investigation of the structure and biological activity relationships of polypeptide is discussed using a neurotoxic peptide, conotoxin GI as a sample. Two newly developed two-dimensional NMR techniques; i.e. pure phase detection and homonuclear Hartmann-Hahn (HOHAHA) spectroscopy were also used to improve the spectral resolution and eliminate the ambiguity of assignment. The pseudo atom treatment was introduced in DADAS to decrease the loss of distance information for the prochiral protons. The calculated structures of conotoxin GI are converged into two groups in space which differ at the C-terminus part of polypeptide. One group of conformers

explains its activity with regard to the binding mechanism to the acetylcholine receptor protein in the neuromuscular junction. This demonstrates that our strategy to determine the solution structure of biological molecule is able to provide a powerful tool in molecular biology.

In Chapter 3, the reliability and limitation on the determination of solution structure are discussed through the comparison of three representative methods of which computer programs; DISGEO[15-18], DADAS[19-23] and CHARMM[24-27] are available. DISGEO is provided for the calculation of the metric matrix method, DADAS for the minimization of the variable target function and CHARMM for the restrained molecular dynamics. There is no report of such a comparison using identical input data so far. Here we carried out calculations to elucidate the three dimensional structure using two sets of input data; one is obtained from our experiment for conotoxin GI and the other is provided through the simulation from the structure of Tendamistat determined by the X-ray analysis. From the aspect of the application in the case of structure determination by NMR, merits and demerits of each method are discussed to achieve further developments of the new born method of solution structure determination of macromolecules.

References

1. Blundell, T.L. & Johnson, L.N. (1976) "Protein Crystallography", Academic Press, New York.
2. Schulz, G.E. & Schirmer, R.H. (1979) "Principles of Protein Structure", Springer-Verlag, New York.
3. Cantor, G. & Schimmel, P.R. (1980) "Biophysical Chemistry", W.H.Freeman, San Francisco.
4. Bax, A. (1982) "Two-Dimensional Nuclear Magnetic Resonance in Liquids", D. Reidel Publishing Comp., Dordrecht.
5. Ernst, R.R., Bodenhausen, G. & Wokaun, A. (1986) "Principles of Nuclear Magnetic Resonances in One and Two Dimensions", Oxford University Press, Oxford.
6. Wüthrich, K. (1986) "NMR of Proteins and Nucleic Acids", Wiley, New York.
7. Noggle, J.H. & Schirmer, R.E. (1971) "The Nuclear Overhauser Effect", Academic Press, New York.
8. Braun, W. (1987) Quart. Rev. Biophys., 19, 115-157.
9. Clore, G.M. & Gronenborn, A.M. (1987) Protein Engineering, 1, 275-288.
10. Blumenthal, L.M. (1970) "Theory and Applications of Distance Geometry", Chelsea, New York.
11. Crippen, G.M. (1981) "Distance Geometry and Conformational Calculation", Research Studies Press, New York.
12. Braun, W., Bösch, C., Brown, L.R., Gō, N., & Wüthrich, K., (1981) Biochim. Biophys. Acta., 667, 377-396.
13. CPK-model, The Ealing Corporation, South Natick,

Massachusetts 01760.

14. Billeter, M., Engeli, M. & Wüthrich, K. (1985) Mol. Graphics, 3. 79-83
15. Havel, T.F. & Wüthrich, K., (1985) J. Mol. Biol., 182, 281-294.
16. Havel, T.F. (1986) DISGEO, Quantum Chemistry Program Exchange, Program No. 507, Indiana University.
17. Braun, W., Wider, G., Lee, K.H. & Wüthrich, K (1983) J. Mol. Biol., 169, 921-948.
18. Williamson, M.P., Havel, T.F. & Wüthrich, K. (1985) J. Mol. Biol., 182, 295-315
19. Braun, W. & Gō, N. (1985) J. Mol. Biol., 186, 611-626.
20. Braun, W., Wagner, G., Wörgotter, E., Vasak, M., Kägi,J.H.R. & Wüthrich, K. (1986) J. Mol. Biol., 187, 125-129.
21. Kobayashi, Y., Ohkubo, T., Kyogoku, Y., Nishiuchi, Y., Sakakibara, S., Braun, W. & Gō, N. (1985) Proc. 9th Am. Peptide Symp. (ed. K.D.Kopple & C.M.Deber), 101-104.
22. Ohkubo, T., Kobayashi, Y., Shimonishi, Y., Kyogoku, Y., Braun, W. & Gō, N. (1986) Biopolymers, 25, 123-134.
23. Kline, A.D., Braun, W. & Wüthrich, K. (1986) J. Mol. Biol., 189, 377-382.
24. Brünger, A.T., Clore, G.M., Gronenborn, A.M. & Karplus, M. (1986) Proc. Natl. Acad. Sci. USA., 83, 3801-3805.
25. Clore, G.M., Brünger, A.T., Karplus, M. & Gronenborn, A.M. (1986) J. Mol. Biol., 191, 523-551.
26. Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, A.T., Karplus, M. and Gronenborn, A.M. (1986) EMBO J., 5, 2729-2735.

27. Brook, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) J. Comput. Chem., 4, 187-217.

CHAPTER I

A Conformational Study of Polypeptides in Solution by ^{1}H-NMR and Distance Geometry

[Summary]

We are developing a strategy to elucidate the conformation of polypeptide and protein in solution. Recently an efficient algorithm of distance geometry has been developed to calculate a conformation with proton-proton distance constraints. The combined use of the distance geometry algorithm and ^{1}H-NMR measurement to determine a conformation is reported here as applied to a shorter peptide of a heat-stable enterotoxin produced by a human strain of enterotoxigenic Escherichia coli (SK-1). Considering the intramolecular motion, we used the three methods to interpret the intensity of NOE to corresponding interatomic distance, i.e. use of rigid model, use of uniform averaging model and setting the maximum upper distance constraint. Various criteria such as averaged root mean square distance between conformers and violation of distance constraints were provided to evaluate the obtained structures. The result revealed that the segment between 9th to 19th residues of STh takes a well defined cylindrical structure.

[Introduction]

Determination of the solution conformation of a polypeptide is indispensable in the study of its structure-activity relationship. Various kinds of spectroscopic methods have been applied to the investigation of polypeptide conformation. However, it has been difficult to delineate the solution conformations at the atomic level, even though the X-ray crystallographic analysis is available in the solid state, there are a large number of polypeptides of which single crystals are difficult to be obtained.

The developments of high resolution proton nuclear magnetic resonance (^{1}H-NMR),[1,2] especially two-dimensional NMR techniques,[3-5] have made it possible to make assignments of ^{1}H-NMR peaks of small proteins consisting of up to about 60 amino acid residues.[6] A large number of data of nuclear Overhauser effects (NOE)[7] which contain information about interatomic distances can be collected and identified on the basis of these assignments. These pieces of distance information restrict possible tertiary structures of a polypeptide molecule. Recently computerized distance geometry algorithms have been applied to several proteins to construct molecular structure in the atomic level from a limited number of distance constraints.[8-14] However these algorithms have a rather narrow applicability because of limitation of computer memory size and speed. We have developed a new efficient distance geometry algorithm which significantly expands the applicability.[15,16] In this paper, we

will describe the application of this algorithm to the analysis of a conformation of a shorter peptide of a heat-stable enterotoxin (STh).[17-20]

Asn-Ser-Ser-Asn-Tyr-Cys-Cys-Glu-Leu-Cys-Cys-Asn-Pro-Ala-Cys-Thr-Gly-Cys-Tyr

Fig. 1. Amino acids sequence of the heat-stable enterotoxin (STh) produced by a human strain SK-1 of enterotoxigenic Escherichia coli.

STh is produced by a human strain of enterotoxigenic Escherichia coli (SK-1) and causes acute diarrhea in human. The amino acid sequence of STh has been determined as shown in Figure 1.[17] It is remarkable that STh consists of six cysteinyl residues which occupies one-third amino acid residues of the molecule and are bound by intramolecular disulfide bonds. The mode of disulfide linkages is still unknown and is under study by chemical analysis.

NMR studies were carried out on an analogue of STh, designated as STh[6-19] which corresponds to the sequence of STh from 6th to 19th residues (Fig. 1).[18-20] This analogue has the biological activity almost the same as the native STh and is more heat-stable. Various homologues of STh have been found to have similar amino acid sequence in this segment. The six cysteinyl residues occur at the same locations.

[Materials and Methods]

STh[6-19] was synthesized using the same procedure as described previously.[18] The chemical synthesis provided an enough amount of sample for NMR measurements, especially for two dimensional NMR. STh[6-19] was dissolved in mixed solvent of 85% dimethylsulfoxide and 15% acetonitrile to be ca 10 mM.

The ^{1}H-NMR measurements were performed on a JEOL GX-500s spectrometer at 500 MHz. Two dimensional correlated spectroscopy (COSY) and NOE correlated spectroscopy (NOESY) were recorded with the standard pulse sequences[3,4]. Size of the data in time domain were 2K points for the t_2 direction and 256 points for the t_1 direction. The data size was expanded to 512 in t_1 direction by zero-filling. The spectra have the digital resolutions of 9.8 Hz/point for the ω_1 direction and 4.9 Hz/point for the ω_2 direction. The spectra were obtained in the absolute value mode. The NOESY spectra were recorded with various mixing times for evaluation of the effects of spin diffusion and coherent magnetization transfer. All NMR measurements were performed at 10.0 °C.

Interatomic distances are evaluated by interpreting the NOE data. The way of interpretation of these data will be described later. For each pair of atoms, i and j, for which the NOE is observed, upper and lower constraints, U_{ij} and L_{ij}, of the interatomic distance r_{ij} are determined as detailed later. For a given set of constraints, the following error function is defined as a target function to be minimized.[15]

$$T = \Sigma' \frac{(U_{ij}^2 - r_{ij}^2)^2}{U_{ij}^2} + \Sigma' \frac{(L_{ij}^2 - r_{ij}^2)^2}{L_{ij}^2} + \Sigma' w[(s_{io} + s_{jo})^2 - r_{ij}^2]^2$$

where Σ' means summation only over the terms which violate distance constraints, and s_i, s_j are repulsive core radii[15] of i and j atoms (they were determined for each type of atoms such that the Lennard-Jones energy for a pair of atoms at the contact radii is 3.0 kcal/mol, i.e., five times RT.), and w is a relative weight of the third term to first and second terms.

The value of the target function T should vanish for conformations which satisfy the constraints. At least it should correspond to one of minima of the target function. In the minimization of the target function, only the dihedral angles of polypeptide[30]; ϕ,ψ and χ's are treated as independent variables by fixing the bond angles and bond lengths of peptide residues to the values listed in Empirical Conformational Energy Program for Peptide (ECEPP).[21,22] With GENER sub-program in ECEPP, a structure is generated for a set of given values of dihedral angles and r_{ij} are calculated. Reductions of the calculation time and computer memory space were achieved by treating only dihedral angles as independent variable.[15,23,24] All of the atoms including hydrogen atoms were taken into account in the calculation.

In order to avoid the multiple minimum problem and to search for the global minimum of the target function T, the distance constraints were applied stepwise following the idea of Ooi et al as follows.[25] At first, the distance constraints were

classified according to the distance k between the residues to which each of paired atoms under consideration belongs, where k is defined as the difference of the residue numbers. Next, minimization was carried out for the target function with the distance constraints only of smaller values of k. Then, by starting from the conformation obtained in this minimization, calculation was continued to minimize the target function with the constraints of up to slightly larger values of k. Starting from the previously minimized conformation, the target function was then minimized with the constraints of the second range between the residues aparted by four residues. A series of the minimization was continued by increasing the range of constraints. Furthermore such a series of the minimization was repeated according to the various values of the weighting factor w which was a ratio between the term of the core radii to that of the distance constraints imposed by NOE. The value of w was changed gradually from 0.1 to 5.0. In earlier stages of the minimization where a small value of w is taken, the distance constraints by NOE's played a relatively important role. During these stages, the peptide folds into grossly correct structures. In later stages with a larger value, the term of the core radii predominates and leads stereochemically satisfactory conformers.

Ten trials with a set of distance constraints were performed with ten different initial conformations respectively. Initial conformations were generated by giving random number to the all rotatable dihedral angles. In the course of the calculations all of the dihedral angles ω for the peptide bonds were set to 180°.

The computers used in the calculations were an ACOS 1000 and

a FACOM M-230 at the computer centers of Osaka University and Kyushu University, respectively. Computer drawings and some of the structure analysis were carried out with aids of a special computer graphics program, VENUS[26] at Crystallographic Research Center, Institute for Protein Research of Osaka University.

[Results and discussion]

Sequential assignments of the protons in the backbone

In order to make assignments of ^{1}H-nmr peaks of STh[6-19], a sequential resonance assignments procedure by Wuthrich and co-workers was employed. J-coupling connectivities with the assist of value of chemical shift & intensity, revealed from the analysis of the COSY spectrum, established almost all of the spin systems.

The individual protons belonging to Glu-8, Leu-9, Pro-13, Ala-14, Thr-16 and Gly-17 were easily assigned by judging from their coupling profiles. Because the spin systems of these six residues are unique respectively in the molecule, their coupling patterns are distinguishable from the others. The protons of aromatic ring of Tyr-19 were identified by the through-space connectivities to each C^{β} protons in the NOESY spectra. Two C^{α} protons of Cys-15 and Cys-18 had incidentally a same chemical shift value. In this case, intra-residual coupling connectivities including amide and C^{β} protons were established respectively via each intraresidual NOE among these protons. Cross peaks between $C^{\alpha}H$ and $C^{\beta}H_2$ in COSY spectrum which are expected two peaks for magnetically non-equivalent two C^{β} protons, were missed sometimes because of the values of dihedral angles. In such cases, the intraresidual NOE's were used to establish each coupling connectivity as well. Thus, all the spin systems of each amino acid residue were identified.

Then sequential resonance assignment was carried out using

NOE connectivities between adjacent residues through amide protons, i.e. NOE connectivities between $C^{\alpha}H$-NH (d_1-connectivity) and between NH-NH (d_2-connectivity). Stereochemical study showed that at least one of the through-space distances corresponding to d_1 and d_2, is smaller than 3.0$\mathring{A}$ which might cause a strong NOE[27]. From the J-coupling cross peaks in COSY spectrum the amide proton and C^{α} proton in the same residue were connected, and then they were connected to the C^{α} proton of the preceding residue by NOEs of d_1- and d_2-connectivities in NOESY spectrum. By repeating this process, sequential assignments along the primary structure were performed. These connectivities were interrupted by Pro13 position for the reason of the absence of the amide proton. The amide protons of Cys6 and Cys7 were not observed because of fast exchanges of the amide protons due to the trace of water contaminated in the solvent. The combined COSY-NOESY connectivity diagram in Figure 2 shows the results of the assignments except Cys6 and Cys7 which could not be identified.

The NOESY spectrum with mixing time of 120 msec was used for the semi-quantitative analysis of the intensity of NOE cross peak because this spectrum could avoid the troubles of spin diffusion and coherent magnetization transfer judging from the comparison of a series of NOESY experiments with different mixing times. For the discrimination of NOE from noise, only those NOEs that appeared commonly in the other NOESY spectra with other mixing times (160, 220 m sec) were adopted. Thirty-eight NOEs between interresidue protons and 29 NOEs between intraresidue protons were collected and identified based on the previously described

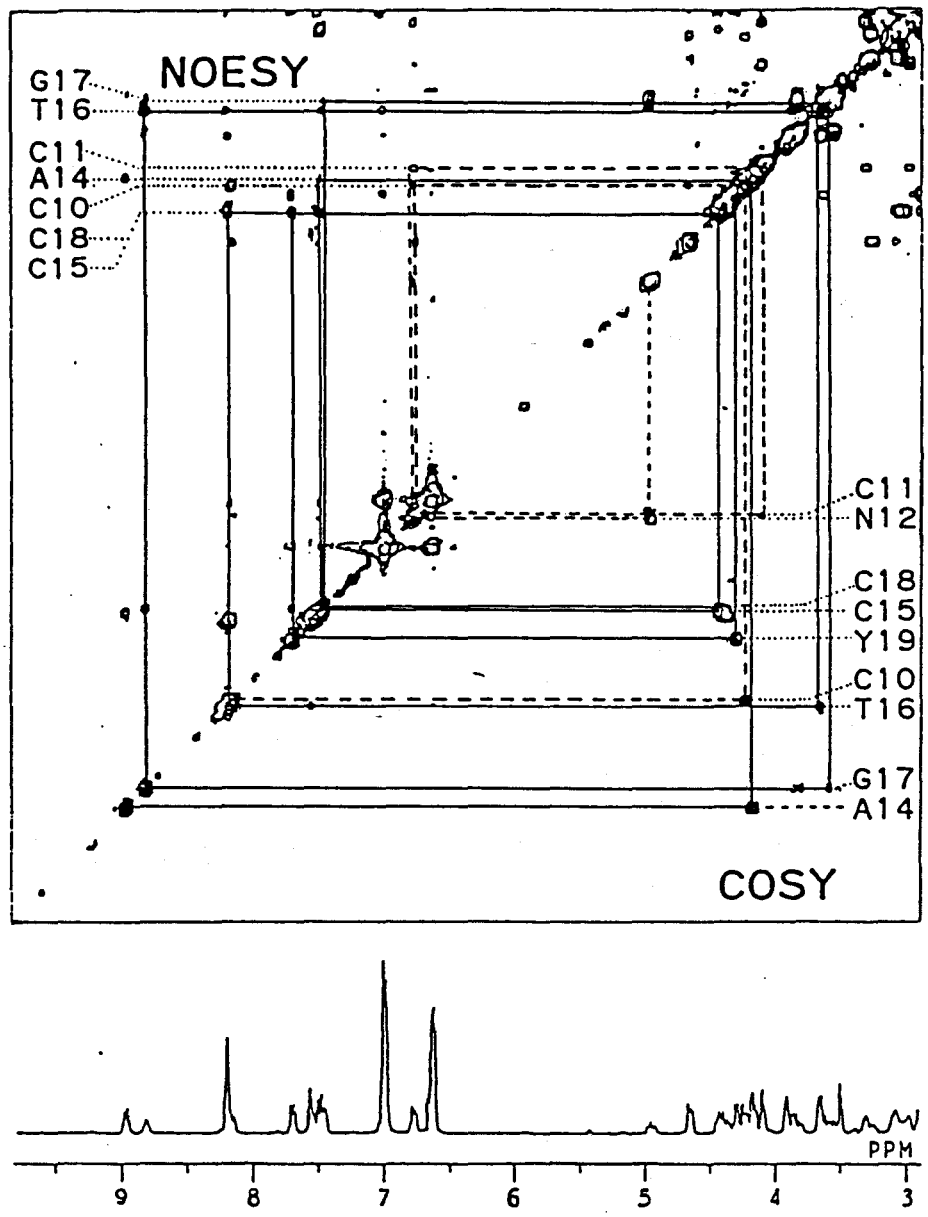


Fig. 2. Contour plot of a combined COSY-NOESY connectivity diagram of the region (ω_1 = 3.9-9.8 ppm, ω_2 = 3.9-9.8 ppm) for sequential resonance assignments via NOE's between amide protons and C^{α} protons of the preceding residue. The bottom of the figure is the conventional one dimensional spectrum of corresponding region (3.9-9.8 ppm). In this contour plot the upper left triangle was taken from the NOESY spectrum at 500 MHz of the STh[6-19] of a 10 mM solution in mixed solvent of 85% dimethylsulfoxide (DMSO) and 15% acetonitrile, 10.0°C with mixing time of 120msec. The lower right triangle was taken from the COSY spectrum of the same sample and condition. The spectra have the digital resolutions of 9.8 Hz/point for the ω_1 direction and 4.9 Hz/point for the ω_2 direction. The solid lines in the lower right triangle indicate J-connectivities between intraresidual amide protons and C^{α} protons, and the solid lines in the upper left triangle indicate d_1-connectivities between amide protons and C^{α} protons of the preceding residue.

assignments. Among the interresidual NOE's, eleven NOE's were those between protons of non-neighboring residues which play important roles in constructing overall structures of the polypeptide.

Relation between NOE and interatomic distances

In the next step the intensities of NOE peaks in the NOESY spectra were translated to corresponding atomic distances. The intensity of NOE is not only a function of inter-atomic distances but also a function of the internal motions of a molecule which are represented by effective rotational correlation times. The molecular motions must be considered carefully to deduce interatomic distance from an NOE. Almost all of the residues of STh[6-19] were detected as incorporated into a network of the interresidual through-space connectivities by negative NOE's. This indicates that all parts of the molecule are tightly folded with three disulfide linkages. There, we assume that all the protons have a common correlation time. Since local intramolecular motions reduce the intensity of NOE, only upper distance constraints were derived from the NOE data. The intensity of NOE between two C^{β} protons of a Cys residue which showed the strongest intensity of NOE among all pairs of C^{β} protons was used for a standard to calibrate relative intensities of NOEs.

Collected NOEs were classified into two groups, i.e.,short range NOEs which mean the NOEs between intraresidual protons and NOE's among backbone hydrogen atoms and C^{β} protons of the adjacent residues, and long range NOEs which mean NOE's among

the remaining pairs of protons. In the case of the interpretation of short range NOE's, a rigid model[8] which treats atomic distances as fixed ones was applied. Since proton pairs which show a short range NOE are separated each other by a small numbers of covalent bonds in the polypeptide structure, fluctuation of distances between these protons can be regarded as small. Therefore, it is reasonable to use the rigid model to short range NOE's.

To interpret long range NOE's, fluctuation of proton-proton distances should be taken into account. Two methods of interpretation were applied as follows. First, we adopted an idea of a fixed upper distance constraint and assumed it to be a threshold value of distance where NOE first was detected. In this interpretation, the distances between protons which showed a long range NOE were assumed to be $4\overset{\circ}{A}$ or $5\overset{\circ}{A}$. The value of $4\overset{\circ}{A}$ was longer than the distances which was obtained from the rigid model. Second, we adopted a so called uniform averaging model which was used ,for example, in the determination of the structure of glucagon in solution by Braun et al.[8,11] The uniform averaging model treats atomic distances as uniformly distributed ones between the minimum and maximum values. The minimum distance between two protons was assumed to be the sum of the core radii of the two protons, i. e., 2.0 $\overset{\circ}{A}$ and maximum distance was derived from the intensity of NOE.

For the sake of convenience, these ways of interpretation were called UNI (uniform averaging model), Rm 5 (setting maximum upper distance constraint as 5.0 $\overset{\circ}{A}$) and Rm 4 (setting maximum upper distance constraints as 4.0 $\overset{\circ}{A}$), respectively (Table

1). The correlations between intensities of NOE peaks in the NOESY spectra and constraints on interatomic distances were summarized in Table 1.

It should be noted that we cannot make stereospecific assignments of two individual C^{β} proton resonances. Therefore when an NOE is observed between any proton and C^{β} proton, the distance is replaced by that measured from the C^{β} carbon atom. The upper limit from the proton to the replaced C^{β} carbon is the sum of the upper distance constraint derived from the NOE and the maximum distance between the C^{β} carbon and the $C^{\beta}H$ hydrogen atoms corresponding to the length of the C-H bond. The other protons, such as protons of methylene and methyl groups, which are also impossible to be stereospecifically differentiated was also treated in the same way. In the similar case of aromatic ring protons of tyrosyl residue, C2 and C6 protons, and C3 and C5 protons are referred to the C^{γ} and C^{ζ} carbon atoms respectively. For a leucyl residue, two C^{δ} methyl protons were referred to the C^{γ} carbon atoms. After these corrections, input data for the distance geometry calculation were prepared respectively for the three sets of interpretations. For example the data between non adjacent residues were shown in Table 2.

Comparison of the calculated structures

By the repetition of the calculations starting from the randomly chosen initial conformations, an assembly of the somewhat different conformers that conform to the imposed distance constraints could be obtained. Starting from ten initial conformations, ten conformers were calculated respectively for

Table 1

Correlation between cross peak intensities in NOESY spectra of STh[6-19] and constraints on ^1H-^1H distances [$\overset{\circ}{A}$]

short range NOEs*		long range NOEs*			
intensity**	rigid model	intensity**	UNI	Rm5	Rm4
> 24	2.5	> 12	5.0	5.0	4.0
23 - 16	2.7	11 - 8	6.5	5.0	4.0
15 - 8	3.0	7 - 6	8.0	5.0	4.0

* Short range NOEs were those between intraresidual protons and those between backbone hydrogen atoms and C^β protons of adjacent residues. Long range NOEs are those between the remaining pairs of protons. Rigid model was applied to the short range NOEs. Three different ways of interpretations were applied to the long range NOEs. These are called as UNI (uniform averaging model), Rm 5 (setting maximum upper distance constraint as 5.0 $\overset{\circ}{A}$) and Rm 4 (setting maximum upper distance constraints as 4.0$\overset{\circ}{A}$).

** Intensities of the cross peaks in the NOESY spectra were given as the values of contour levels.

Table 2

Upper distance constraints of non adjacent residues in STh[6-19] which were input in the calculation of the distance geometry algorit

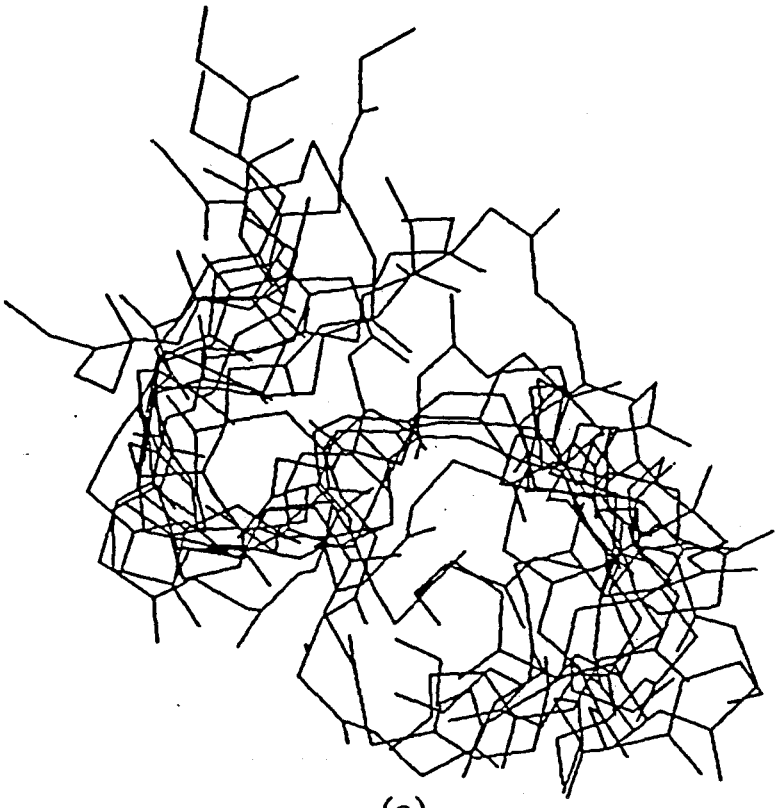
Observed NOEs		Corresponding upper distance constraints* [Å]			
			UNI	Rm5	Rm4
Leu 9 $C^{\delta}H3$	Thr 16 NH	C^{γ} - NH	10.2	7.2	6.2
Leu 9 $C^{\delta}H3$	Thr 16 $C^{\beta}H$	C^{γ} - $C^{\beta}H$	10.2	7.2	6.2
Cys 11 $C^{\delta}H2$	Cys 15 $C^{\beta}H2$	C^{β} - C^{β}	10.2	7.2	6.2
Asn 12 $N^{\delta}H2$	Ala 14 $C^{\beta}H3$	N^{δ} - C^{β}	10.1	7.1	6.1
Asn 12 $C^{\beta}H2$	Cys 15 NH	C^{β} - NH	9.1	6.1	5.1
Ala 14 $C^{\beta}H3$	Tyr 19 $C^{\alpha}H$	C^{β} - $C^{\alpha}H$	9.1	6.1	5.1
Ala 14 $C^{\alpha}H$	Tyr 19 C2,6H	$C^{\alpha}H$ - C^{γ}	10.1	7.1	6.1
Thr 16 $C^{\alpha}H$	Tyr 19 C2,6H	$C^{\alpha}H$ - C^{γ}	7.1	7.1	6.1
Thr 16 $C^{\alpha}H$	Cys 18 NH	$C^{\alpha}H$ - NH	8.0	5.0	4.0
Thr 16 NH	Tyr 19 NH	NH - NH	8.0	5.0	4.0
Thr 16 NH	Tyr 19 C2,6H	NH - C^{γ}	8.6	7.1	6.1

* In the cases of proton groups which cannot to be stereospecifically assigned by NMR, reference atoms of distance constraints are replaced to suitable atoms. (See text.)

each of the three sets of the constraints. Totally 30 conformers were obtained as shown in Figure 3. The computers spent about 15 minutes for each trial in the calculation of distance geometry.

To investigate the consistency between the distance constraints used for calculation and distances in the obtained conformer, the averaged values of the sum of violations of the upper distance constraints imposed by NOE data (V_{upper}) and the averaged values of the sum of violation with allowance of 0.1 $\overset{\circ}{A}$ of the core radius (V_{lower}) were calculated for each conformer (Table 3). The violation within 0.1 $\overset{\circ}{A}$ corresponds to the range of the possible contact between two atoms. All of the values of V_{lower} were nearly zero and the values of V_{upper} were considerably larger than those of V_{lower}. This fact reflects that we made the relative weight of core radius w gradually increase and to reach a value of 5.0. All of the values of Vupper were about $2\sim6\overset{\circ}{A}$, but most of the violations for the upper distance constraints fell into the range of 0.1 $\sim$0.5 $\overset{\circ}{A}$. These values of violation seem reasonable from the point of semi-quantitative interpretation of NOE.

To select the practical interpretation of the NOE data, the criteria which we chose are whether resulting conformers were well converged to each other, and how the mutual distances between the atoms in a conformer are consistent to the input constraints. The consistency of the resulting conformers were examined with respect to V_{upper} and the number of unacceptable conformers where the distance constraints of core radius with allowance of 0.1 $\overset{\circ}{A}$ were violated to cause an extremely strong repulsive force. The interpretation of Rm4 gave only four



(a)

(b)

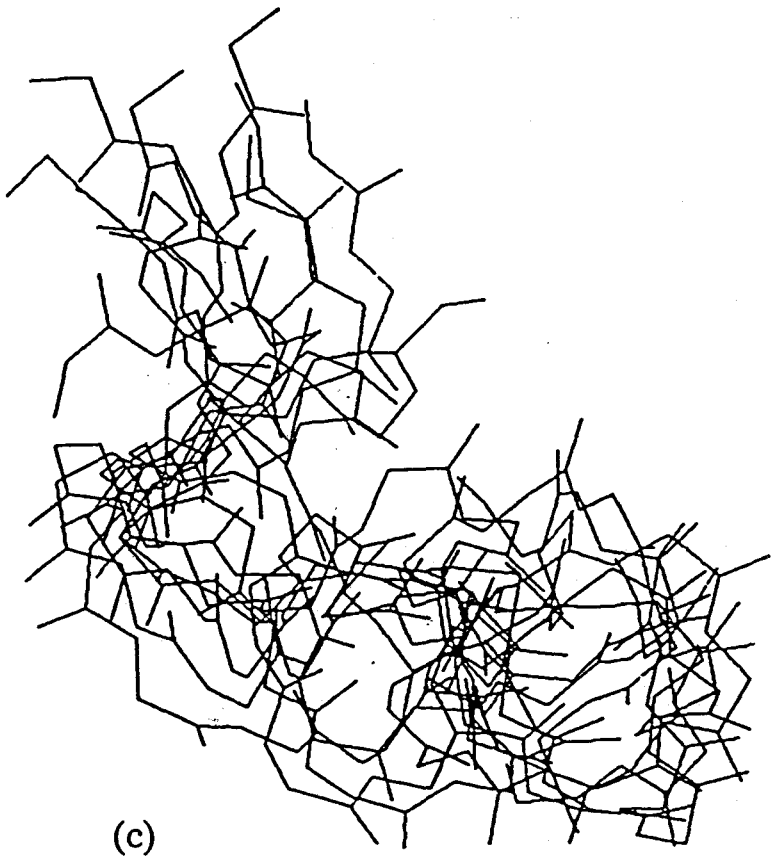


Fig. 3.

Computer drawings of the 10 conformers of STh[6-19] for the three different interpretations (UNI, Rm5, Rm4) of NOE data. Ten conformers were superimposed so as to fit each conformers best in space for three interpretations. Drawings of all main-chain backbone atoms (C^{α}, C', C'O, N) are shown. (a) 10 conformers calculated with the interpretation of NOE data, UNI (treating long range NOEs by uniform averaging model); (b) Rm5 (setting maximum upper distance constraint of long range NOEs as 5.0 $\mathring{A}$); (c) Rm4 (setting maximum upper distance constraints of long range NOEs as 4.0$\mathring{A}$).

Table 3

Violations of distance constraints in ten calculated conformers for three different upper bounds

	UNI	Rm5	Rm4
Vupper[a] [Å]	2.69	3.62	5.48
Vlower[b] [Å]	0.0	0.12	0.24
Number of sterically unacceptable conformer[c]	0	2	6

a : The averaged values of the sum of the violations of the upper distance constraints imposed by NOE data

b : The averaged values of the sum of the violations of the core radii with allowance of 0.1 Å.

c : The number of the conformers that violate the distance constraints of core radii with allowance of 0.1 Å.

stereochemically acceptable structures. V_{upper} of Rm4 was the largest value among the three different interpretations. Thus, interpretation of Rm4 was the worst in terms of consistency and not suitable for the practical interpretation. The convergence of the resulting conformers were examined with respect to averaged root-mean-square distances (RMSD) among each ten conformers. Table 4 lists the RMSD values calculated for two sets of atoms; the backbone and hydrogen atoms bonded directly to them, and for all the atoms. The order of the RMSD values for the sets are parallel to each other. Although results obtained for UNI satisfies the imposed constraints, they did not achieve as good a convergence between the conformers as in the other two cases. This implies that the upper distance constraints in UNI might contain some overestimations in the interpretation of UNI, and the constraints of the core radii might be easily satisfied by the conformation obtained with overestimated upper distance constraints. The convergence of Rm 4 was worse than that of Rm 5. This might be explained by underestimation of the upper distance constraints, i.e. setting 4 $\mathring{A}$ to the long range NOEs made some upper distance constraints shorter than actual distances. Therefore the result of Rm5 was concluded to be best judging from the criteria and will be used for conformational analysis of STh[6-19].

Conformation of STh[6-19] in solution

Two conformers with the violation of core radius more than 0.1 $\mathring{A}$ were discarded. The remaining eight conformers were then regarded to be stereochemically acceptable. In these conformers,

Table 4

Averaged root-mean-square distances (RMSD) among ten calculated conformers for three different ways of interpreting upper bounds [$\overset{\circ}{A}$].*

	UNI	Rm5	Rm4
Backbone atoms**	4.18	3.21	3.62
All atoms	5.39	4.26	4.51

** : Atoms in peptide backbone and atoms bonded directly to them.

* To investigate the convergence of the resulting conformers, root-mean-square distances (RMSD) between any two conformers were calculated using the algorithm described by McLaclan.[29] The RMSD in the N atomic coordinates between two conformers is defined by

$$RMSD = [\frac{1}{N} \Sigma' (r_i - Rr'_i)]^{1/2}$$

where r_i and r'_i are atomic coordinates of the two conformers with the same centers of mass and R is the best fit rotation matrix calculated to fit the two structures best in space.

the segments 9-19 were well converged and the segments 6-8 were randomly distributed (Fig. 4). Averaged r.m.s.d values between these eight conformers in the segment 9-19 were 2.33 Å for the backbone atoms including hydrogen atoms and 3.28 Å for all the atoms. None of interresidual NOE were interpreted in the segment 6-8 because of the failing in the assignments of Cys 6 and Cys 7 and of the missing of the amide proton peaks of Glu 8. Thus in this segment distance constraints imposed by NOE data were very poor. Therefore it is difficult to tell the structure of the segment 6-9. However a common feature could be extracted in segment 9-19 from the comparison of obtained conformers (Fig. 4, Fig. 5).

The backbone structure in the segment 9-19 of STh[6-19] takes a cylindrical structure. This structure consists of three recurring loops with almost all of the residues in the segment 9-19. These turns take place at Leu9-Cys11, Asn12-Cys15 and Gly17-Tyr19, where some of the previously described d_2-connectivity were detected to give a support of these turn formations.

In the present structure analysis, we could not decide the mode of the formation of disulfide bonds of STh[6-19], mainly because of the failure to assign of Cys6 and Cys7 and also because of the scarcity of the NOE data concerning side-chain atoms. To obtain reliable overall structure, the assignment of Cys6 and Cys7 are underway.

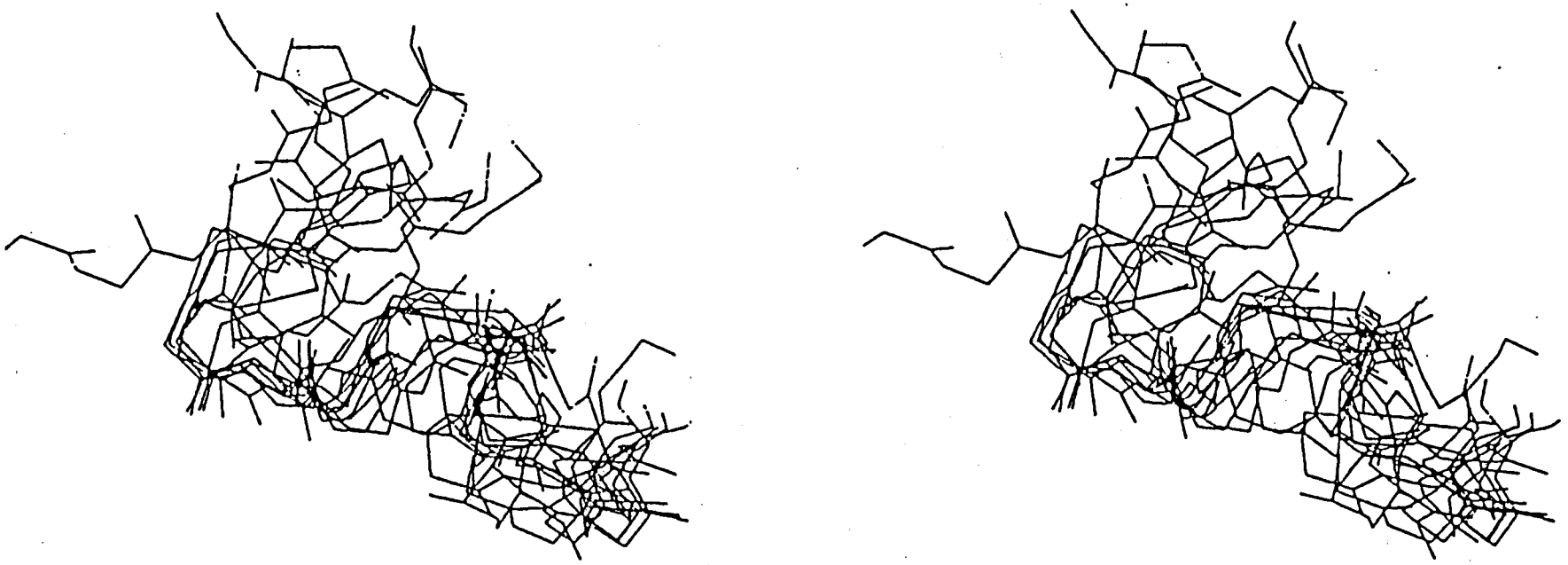


Fig. 4.

Computer drawings of the 8 resulting conformers of STh[6-19] calculated with the interpretation of Rm5 of NOE data, (setting maximum upper distance constraint of long range NOEs as 5.0 $\overset{\circ}{A}$) which were stereochemicaly acceptable. They were superimposed so as to fit segments 9-19 of each conformers best in space. Drawings of all main-chain backbone atoms (C^{α}, C', $C'O$, N) are shown.

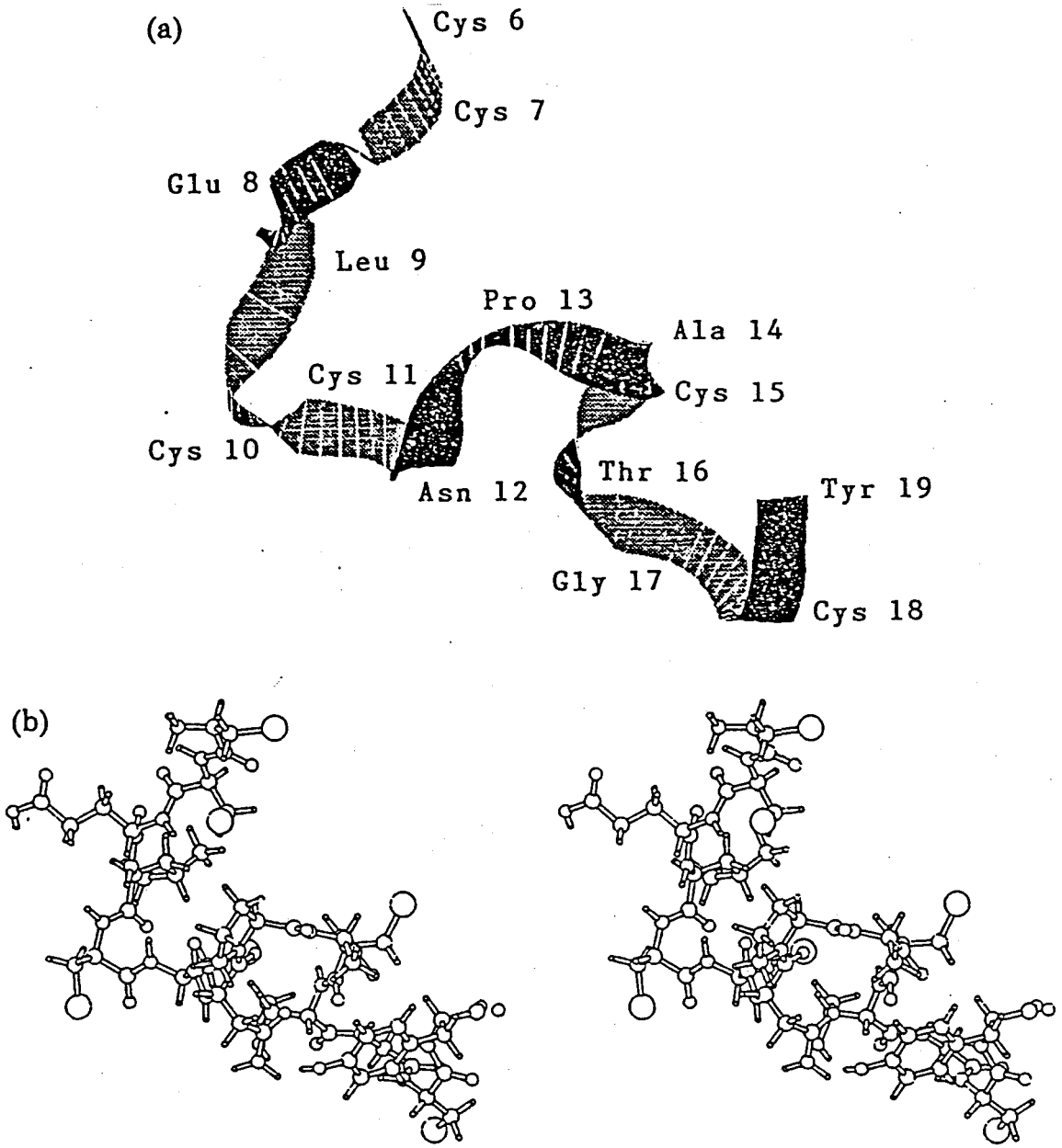


Fig. 5.

Computer drawing of the molecular conformation of STh[6-19] with the lowest value of the target function calculated with the interpretation of Rm5 of NOE data. (a) Schematic drawing of the polypeptide backbone is shown. (b) Stereo drawing of the all atoms including the hydrogen atoms is reprsented.

References

1. Markley, J.L. & Ulrich, E.L.
   (1984) Ann. Rev. Biophys. Bioeng. 13, 493-521
2. Jardetzkey, O. & Roberts, G.C.K.
   (1981) "NMR in Molecular Biology", Academic Press, New York
3. Aue, W.P., Bartholdi, E. & Ernst, R.R.
   (1976) J. Chem. Phys. 64, 2229-2246
4. Jeener, J., Meier, B., Bachman, P. & Ernst, R.R.
   (1979) J. Chem. Phys. 71, 4546-4553
5. Bax, A
   (1982) "Two-Dimensional Nuclear Magnetic Resonance in Liquids", D. Reidel Publishing Comp., Dordrecht
6. Wüthrich, K., Wider, G., Wagner, G. & Braun, W.
   (1982) J. Mol. Biol. 155, 311-319
7. Noggle, J.H. & Schirmer, R.E.
   (1971) "The Nuclear Overhauser Effect", Academic Press, New York
8. Braun, W., Bösch, C., Brown, L.R., Gō, N., & Wüthrich, K.
   (1981) Biochim. Biophys. Acta. 667, 377-396
9. Crippen, G.M., Oppenheimer, N.J. & Connolly, M.L.
   (1981) Int. J. Peptide Protein Res. 17, 156-169
10. Brown, L.R., Braun, W., Anil Kumar & Wüthrich, K.
    (1982) Biophys. J. 37, 319-328
11. Braun, W., Wider, G., Lee, K.H. & Wüthrich, K
    (1983) J. Mol. Biol. 169, 921-948
12. Arseniev, A.S., Kondakov, V.I., Maiorov, V.N. & Bystrov, V.F.

(1984) FEBS Letters 165, 57-62

13. Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. & van Gunsteren, W.F.
(1985) J. Mol. Biol. 182, 179-182

14. Williamson, M.P., Havel, T.F. & Wüthrich, K.
(1985) J. Mol. Biol. 182, 295-315

15. Braun, W. & Gō, N.
(1985) J. Mol. Biol. 186, 611-626.

16. Kobayashi, Y., Ohkubo, T., Kyogoku, Y., Braun, W. & Gō, N.
(1985) Peptides. Proc. American Peptide Symp. (Kopple, K.D. & Deber, C.M. ed.) Pierce Chem. Comp. Rockford, 101-104.

17. Aimoto, S., Takao, T., Shimonishi, Y., Hara, S., Takeda, T., Takeda, Y. & Miwatani, T.
(1982) Eur. J. Biochem. 129, 257-263

18. Ikemura, H., Yoshimura, S., Aimoto, S., Shimonishi, Y., Hara, S., Takeda, T., Takeda, Y., & Miwatani, T.
(1984) Bull. Chem. Soc. Jap. 57, 2543-2549

19. Ikemura, H., Watanabe, H., Aimoto, S., Shimonishi, Y., Hara, S., Takeda, T., Takeda, Y., & Miwatani, T.
(1984) Bull. Chem. Soc. Jap. 57, 2543-2549

20. Yoshimura, S., Ikemura, H., Watanabe, H., Aimoto, S., Shimonishi, Y., Hara, S., Takede, T., Miwatani, T., & Takeda, Y.
(1985) FEBS Letters 181, 138-142

21. Momany, F.A., McGuire, R.F., Burgess, A.W. & Scheraga, H.A.
(1975) J. Phys. Chem. 79, 2361-2381

22. Nemethy, G., Pottle, M.S. & Scheraga, H.A.
(1983) J. Phys. Chem. 87, 1883-1887

23. Abe, H., Braun. W., Noguti T. & Gō. N.

(1984) Comp. Chem., 8, 239-247.

24. Noguti, T. & Gō, N.
(1983) J. Phys. Soc. Jpn. 52, 3685-3690

25. Ooi, T., Nishikawa, K., Oobatake, M. & Scheraga, H.A.
(1978) Biochim. Biophys. Acta. 536, 390-405

26. Iga, Y. & Yasuoka, N.
(1984) J. Mol. Graph. 2, 79-82

27. Billeter, M., Braun, W. & Wüthrich, K.
(1982) J. Mol. Biol. 155, 321-346.

28. Wüthrich, K., Billeter, M. & Braun, W.
(1984) J. Mol. Biol. 180, 715-740

29. McLachlan, A.D.
(1979) J. Mol. Biol. 106, 983-994

30. IUPAC-IUB Commission on Biochemical Nomenclature
(1983) J. Mol. Biol. 52, 1-17

CHAPTER II

Solution Conformation of Conotoxin GI by ^{1}H-Nuclear Magnetic Resonance and Distance Geometry

[Summary]

Conformational analysis of conotoxin GI which is one of the neurotoxic peptides produced by a kind of marine snail, genus Conus, is performed by the combined use of nuclear magnetic resonance spectroscopy (NMR) and distance geometry calculations. The resulting conformers by the minimization of the target function are classified into two groups. The difference of structures of the conformers comes mainly from the difference of the orientation of the side chain of the tyrosyl residue. The results show that the solution structure of conotoxin GI satisfies the conformational requirements for the biological activities of the antagonist to nicotinic cholinergic receptors elucidated from a series of studies on alkaloids. The structure is discussed from the aspect of the comparison to the atomic arrangements of the active site of snake venom peptides and to the molecular models based on the secondary structure prediction.

[Introduction]

Fish-eating marine snails of genus Conus produce several classes of toxic peptides which are specifically bound to key elements in nerve and muscle cell membranes resulting in successive blocks of a series of the neuromuscular systems of fish[1-5]. They are named as conotoxins and classified into three large groups: α-conotoxins are presynaptic[6,7], ω-conotoxins are postsynaptic[8,9] and μ-conotoxins are muscle channel inhibitors[10]. Gray et al. have made a systematic study of these peptides and shown their common features that they are relatively small peptides of 13 to 29 amino acid residues and include 2 to 3 disulfide bonds. As α-conotoxins, three related peptide toxins consisting of 13 to 15 amino acid residues were isolated from the venom of Conus geographus and shown to block synaptic transmission by binding to the acetyl choline receptor to cause paralysis[11,7]. These three peptides, designated conotoxin GI, GIa and GII were chemically synthesized and the modes of their disulfide formations were revealed[12,6].

The molecular conformation of conotoxin GI was proposed by model buildings through the secondary structure analysis with circular dichroism (CD) measurements and Chou & Fasman's way of prediction[13-15].

Its toxic activity is similar to those of snake α-neurotoxins with 60 to 70 amino acid residues[16,17]. These toxins have been well investigated in relation to curare-like alkaloids which show a potent activity to block the nicotinic

acetylcholine receptor[18,19]. Comparing to their structures it has been suggested that there is a strong resemblance between the conformation of conotoxin GI and those of the reactive centers, so called active tips, of snake α-neurotoxins[20-23].

In order to testify the structure of the model, we have determined the three-dimensional structure of conotoxin GI in solution. The strategy to elucidate the structure is the combined use of ^{1}H-NMR and a distance geometry algorithm[24-26]. The result of this small peptide with such a remarkable activity should provide a good explanation for the structure-activity relationship of conotoxin GI.

[Materials and Methods]

Sample: Conotoxin GI of which amino acid sequence is shown in Fig.1 was synthesized and demonstrated its activity as described previously. [11,12]. The chemical synthesis provided an enough amount of sample for NMR measurements, especially for two dimensional NMR. Deuterated dimethyl-sulfoxide (DMSO-d_6) was used as the solvent and the concentration of the solution was 8 mM.

NMR spectroscopy: Proton NMR spectra were recorded at 500MHz on a JEOL GX-500 spectrometer. Two-dimensional spectra of Nuclear Overhauser Effect Spectroscopy (NOESY)[27], Double Quantum Filtered Chemical-Shift Correlated Spectroscopy (DQF-COSY)[28] and Homonuclear Hartmann-Hahn Spectroscopy (HOHAHA)[29] were recorded in the phase sensitive detection mode by the method in four quadrants proposed by States et al[30]. The NOESY spectra were recorded with various mixing times, 100, 120 and 200 ms, for evaluation of the effects of spin diffusion and coherent magnetization transfer. The HOHAHA spectra were also recorded with various mixing times, 25.1, 43.0, 65.5 and 95.5 ms, to follow successively direct, single and multiple relayed through-bond connectivities. The size of the data in time domains were 2K points for the t_2 direction and 256 points for the t_1 direction. The digital resolutions in NOESY and HOHAHA spectra were 10.74 Hz/point in the ω_1 direction and 5.37 Hz/point in the ω_2 direction. These were obtained by zero filling in the 1_1 direction. All NMR measurements were carried out at 20°C.

Calculations: In order to calculate the conformation which

satisfies the distance constraints between paired protons in the molecule, the minimization of the variable target function was carried out with the program, DADAS, as described by Braun & Gō and Ohkubo et al[25,26]. The mainframe computer ACOS 2000 at the computer center of Osaka University was used. Analysis of the structures and molecular drawings were carried out with the program, VENUS, for ACOS 900 computer at the Crystallographic Research Center, Institute for Protein Research of Osaka University[31].

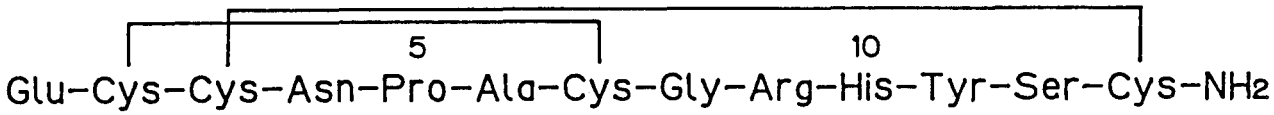


Fig. 1 Amino acid sequence of conotoxin GI from Conus geographus[11,12].

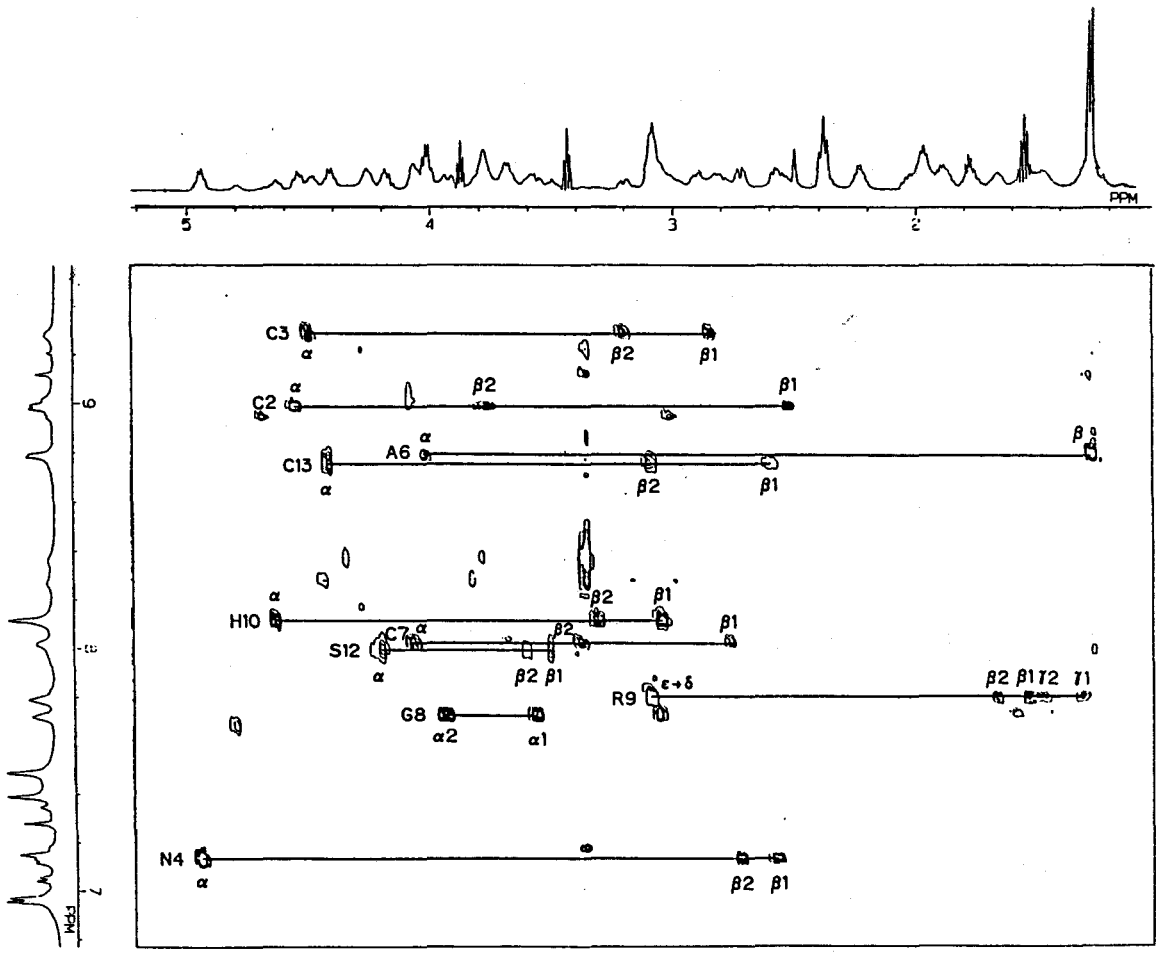


Fig. 2 The NH-aliphatic region of the HOHAHA spectrum of conotoxin GI in DMSO-d_6.
Horizontal lines show the intraresidual spin-coupling connectivities. The backbone NH's of Glu^1, Arg^9, and His^{11} are missing. The connectivity among Arg^9 is demonstrated with NεH-CδH.

[Results and Discussion]

Assignment of the ^{1}H nmr spectrum.

The sequence specific resonance assignment procedure which was introduced by Wüthrich et al. was used to make whole peak assignments as follows(32).

First, amino acid spin systems were identified by surveying direct and relayed through-bond connectivities in the DQF-COSY and the HOHAHA spectra. Fig.2 shows the NH-aliphatic region of the HOHAHA spectrum. The series of HOHAHA spectra contributed generally to delineate the spin-coupling connectivities more than DQF-COSY spectra. Thus, the latter were used rather complimentarily.

The protons of aromatic rings of His^{10} and Tyr^{11} were identified by the through-space connectivities to each C_β protons in the NOESY spectra. The amide protons of the side chain of Asn^4 and the C-terminus of the peptide were also connected to the corresponding C_β protons of Asn^4 and Cys^{13} respectively in the NOESY spectra. The OH proton of Ser^{12} and the z-NH_2 proton of Arg^9 were missed in all spectra. These signals might be broadened because of a fast exchange of such a labile protons with those of water which existed in DMSO in an extremely small quantity. The backbone NH's of Arg^9 and Tyr^{11} were detected only in the sequential Nuclear Overhauser Effect (NOE) connectivities with the corresponding C_α protons of Gly^8 and His^{10}, respectively. The individual protons belonging to Glu^1, Pro^5, Ala^6, Gly^8, Arg^9 and Ser^{12} were easily assigned by judging from their coupling

profiles. Because, the spin systems of these six residues are unique respectively in the molecule, so that their coupling patterns are distinguishable from the others.

Next, the assignments of the resonances of remaining residues including four cysteinyl residues were carried out by the mean of the sequential assignment using NH-C_αH, NH-NH and NH-C_βH protons connectivities. Fig.3 shows the NOESY-COSY connectivity diagrams where the assignment pathway is demonstrated by indicating alternatively COSY connectivties, NH_i-$C_\alpha H_i$, and NOESY connectivities, $C_\alpha H_i$-NH_{i+1}. Although the network is interrupted at Asn^4 because of lack of NH in Pro^5, there is a through-space connectivity between the C_α proton of the former and the C_δ protons of the latter. Table I shows the list of resonance assignments.

Some satellite signals were detected in the spectra as indicated in Figs.4 and 5. Fig.4 which is a part of the HOHAHA spectra shows that Pro^5 and Ala^6 exist in two states. Each minor peak was assigned respectively by their unique spin systems. Differences of chemical shifts of each satellite peak to those of the corresponding major peaks are so small except for those of C_α protons that the existence of isomer was only detectable in the two dimensional spectra.

There were no NOE connectivities detected between major and minor peaks in NOESY spectra even with long mixing time, 200 msec. Thus the conformational exchange between these isomers should be slow. Although any satellite signals at Asn^4 was not detectable, such a kind of interconversion is generally attributable to the cis-trans isomerization at X-Pro peptide

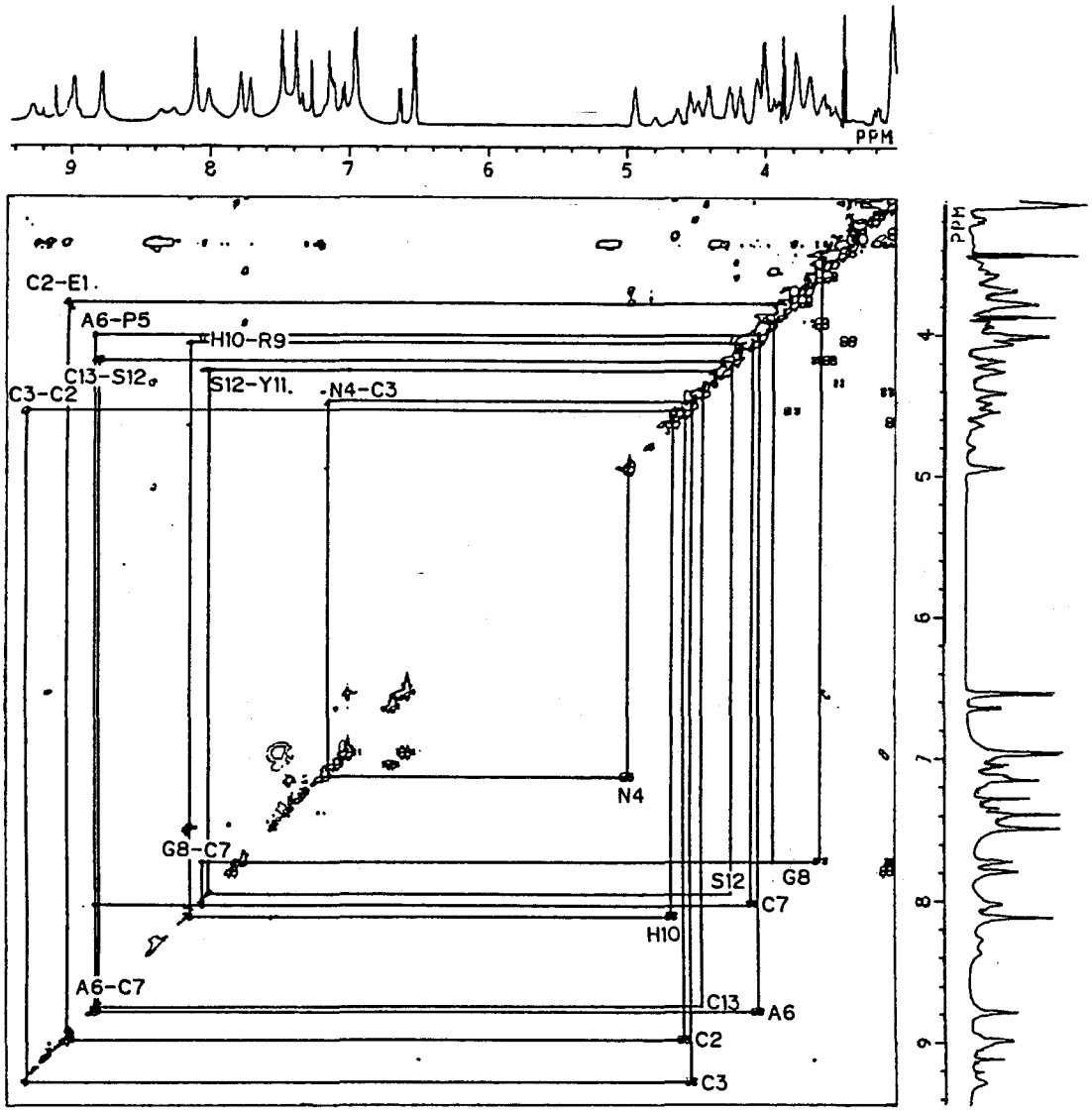


Fig. 3 NOESY-COSY connectivity diagram for sequential assignments. The lower right triangle was taken from the COSY spectrum of conotoxin GI in DMSO-d_6 solution. The cross peaks corresponding to intraresidual spin couplings between $C_\alpha H$ and NH are labeled. Such cross peaks of E1, P5, R9, and Y11 are missing. The upper left triangle was taken from the NOESY spectrum of same sample with a mixing time of 120 ms. The cross peaks corresponding to NOE connectivities between $C_\alpha H_i$ and NH_{i+1} are labeled. The assignment pathway is demonstrated by indicating alternative connectivities of COSY and NOESY. These are interrupted at N4-P5, G8-R9, and H10-Y11 because of the lack of NH's. The NOE connectivities between NH_i-NH_{i+1} were used in the junctions at A6-C7 and C7-G8.

	NH	$C_\alpha H$	$C_\beta H$	$C_\gamma H$	$C_\delta H$	$C_\varepsilon H$	OH
Glu 1	8,27	3.78	1.96	2.38			
	8.36		2.01				
Cys 2	8.95	4.54	2.52				
			3.77				
Cys 3	9.28	4.49	2.82				
			3.19				
Asn 4	7.13	4.94	2.56		7.49		
			2.71		8.11		
Pro 5		4.00	1.78	1.88	3.68		
			2.23	1.97	3.77		
Ala 6	8.79	4.01	1.28				
Cys 7	8.02	4.05	2.75				
			3.36				
Gly 8	7.71	3.54					
		3.92					
Arg 9		4.07	1.52	1.31	3.08	7.79	
			1.66	1.46			
His 10	8.10	4.62	3.03		7.26	8.95	
			3.30				
Tyr 11		4.25	2.79		6.96	6.54	9.12
					(7.05)	(6.64)	(9.21)
			2.89				
Ser 12		4.18	3.61				
			3.50				
Cys 13	8.74	4.41	2.56		7.15		
			3.07		7.39		

Table I The list of the resonance assignment of conotoxin GI in DMSO-d_6 at 20°C. All values of chemical shift are indicated by ppm from tetramethylsilane. Numbers in parentheses indicate the chemical shifts of satellite signals.

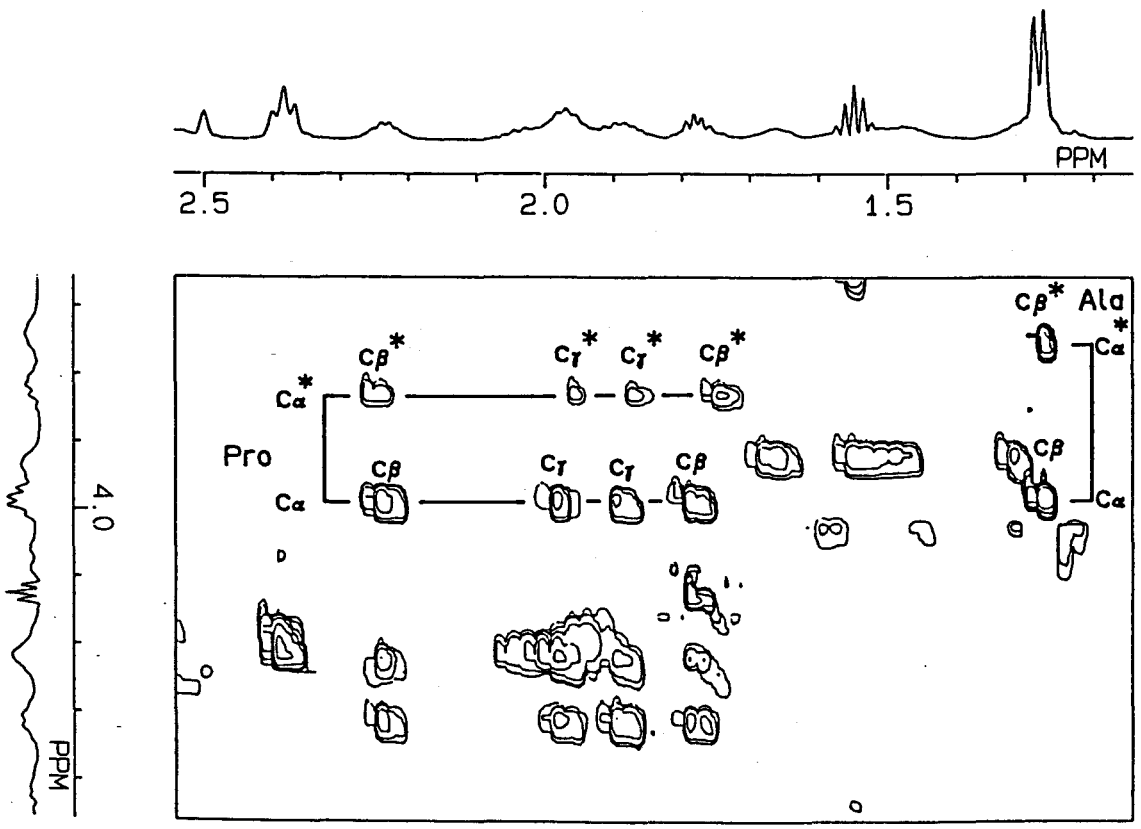


Fig. 4 Part of aliphatic region (ω_2 1.25 - 2.55 ppm, ω_1 3.65 - 4.45 ppm) of the HOHAHA spectrum of conotoxin GI in DMSO-d_6. Horizontal lines at 1.7 - 2.4 ppm connects the two spin coupling networks of the pyrrolidine ring corresponding to the major and minor peaks of $C_\alpha H$ of Pro^5. The peaks of the minor component are indicated by (*). The peaks around 1.3 ppm are $C_\beta H$ corresponding to the major and minor peaks of C_α of Ala^6.

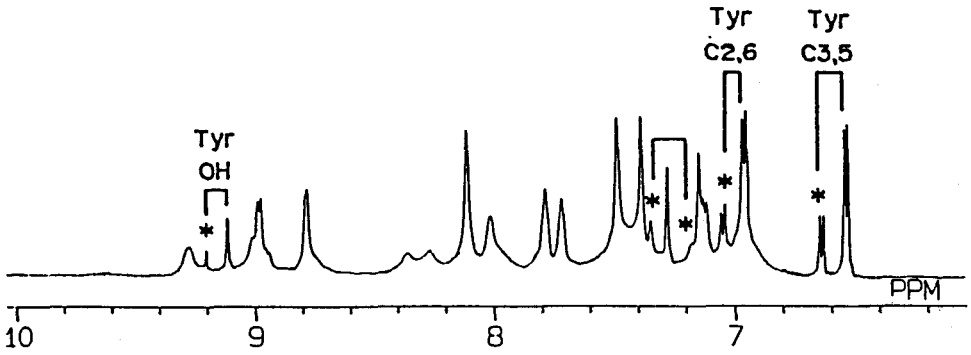


Fig. 5 Part of the one-dimensional spectrum of conotoxin GI in DMSO-d_6. The signals of aromatic CH and OH protons of Tyr^{11} are shown with their satellite signals indicated with (*). A paired signals at 7.18 and 7.35 ppm could not been assigned.

bond[33,34].

Other satellite signals were found at Tyr^{11}. As shown in Fig.5, the signals of aromatic and OH protons carried satellite signals beside the major peaks. The isomerization around Tyr^{11} residue is not clear so far. This could not be attributed to the prolyl cis-trans isomerization because the ratios of the peak intensities are quite different from the cases of Pro^{5} and Ala^{6}. Such an isomerization around the Tyr residue was also found in the NMR spectra of conotoxin MI which is a homologue of conotoxin GI produced by Conus magus. The ^{1}H-NMR spectrum of conotoxin MI showed such satellites of which intensities depend largely on solvent and temperature and no NOE connectivities was detected among major and minor peaks of Tyr^{12} residues of conotoxin MI[35]. These phenomena which should be attributed to some isomerization had been detected in the HPLC profiles of conotoxin MI as well[36,37]. The detail of the isomerization of conotoxin GI relating to conotoxin MI is under investigation.

Another paired peaks at 7.18 ppm and 7.35 ppm as indicated in Fig. 5 were observed. However, the assignment of these peaks was not possible, because there were no connectivities detected with these peaks as if they are isolated from all of others.

Judging from these findings, there should be two kinds of conformational isomerizations about conotoxin GI in DMSO. However, the rates of interconversions were so slow that any correlation peak between the isomers was not detected in the NOESY spectra, and the intensities of the satellite peaks were quite weak. Furthermore the conformational isomerizations are localized. Thus, all these isomerizations were neglected in this

structure determination.

Secondary structure

Because conotoxin GI is so small and tightened by two disulfide bonds, any common secondary structure could not be expected so much. However, Hider pointed out by curve fitting of the CD spectrum of conotoxin GI that the spectrum shows a high α-helical content up to 50%. He carried out the secondary structure prediction with the modified Chou & Fasman's method and postulated the structure placing α helical region between Cys^6 to His^{10} without interruption by two disulfide bonds and Pro^5 residues[13].

The regular secondary structure elements of polypeptides are able to be elucidated by analyzing the sequential and medium range NOE's[38]. By investigating such NOE's which were found in the NOESY spectra of conotoxin GI and are summarized in Fig.6, its secondary structure in solution was surveyed to test such a postulated chain folding. In the NOESY spectra, strong NOE's between $C_\alpha H(i)$ and NH(i+1), $d_{\alpha N}$, were detected at the residues of i=1,2,3,9,11 and 12. These NOE's are indicative that these segments of polypeptides are almost in fully extended or in β-sheet structures. Weak $d_{\alpha N}$ NOE's were detected at the residues of i=5,7 as well. Strong NOE's between NH(i) and NH(i+1), d_{NN}, were found at residues of i=6 and 7 and weak NOE's were at i=3. The d_{NN} NOE's indicate the local arrangements of conformation in α-helix or in a turn I structure. Weak NOE between $C_\alpha H(i)$ and NH(i+2) was detected at the residues of i=6. These profiles of sequential and medium range NOE connectivities in the range

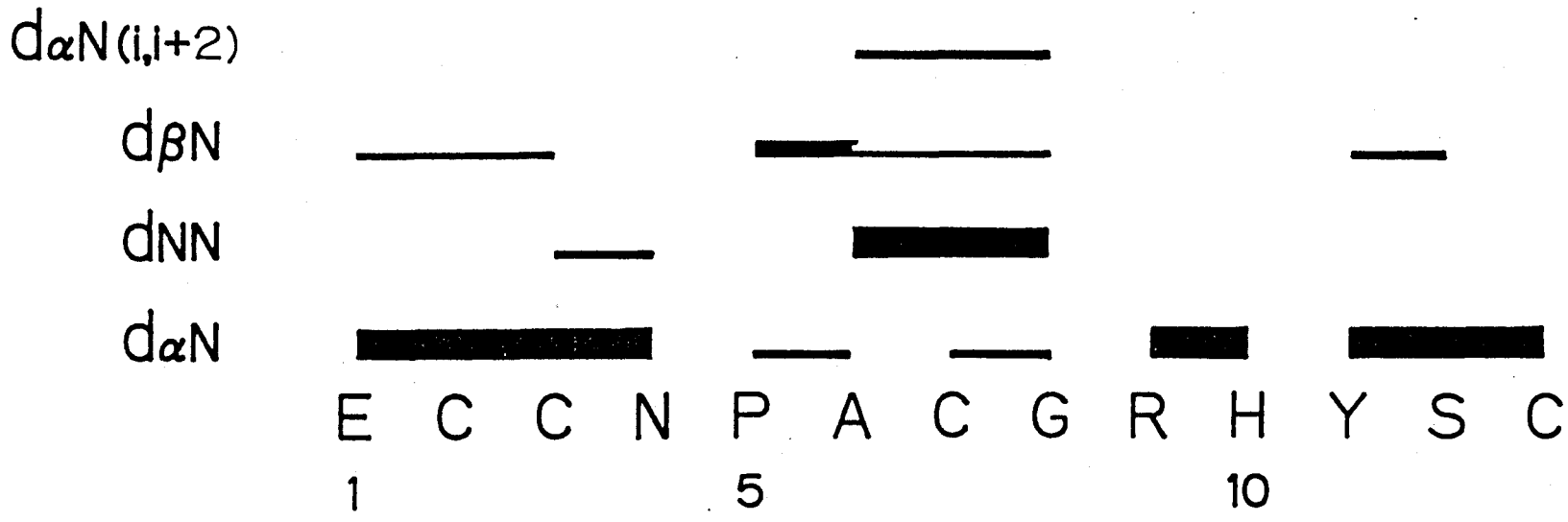


Fig. 6 Sequential and medium range NOE's detected in the NOESY spectrum of conotoxin GI. $d_{\alpha N}$ means NOE between $C_{\alpha}H(i)$ and NH(i+1) where i is the residual number. d_{NN} means NOE between NH(i) and NH(i+1) and $d_{\beta N}$ between $C_{\beta}H(i)$ and NH(i+1). The thickness of lines represents the relative intensities of cross-peaks in the NOESY spectrum.

between 5 to 8 suggest that a turn I structure should exist in this region. This region corresponds to the segment of which secondary structure was predicted to be an α-helical conformation by Hider[13].

The other combination of weak $d_{\alpha N}$ and d_{NN} suggest another turn I structure in the sequence around 3 to 4 as well. The existence of turn I structures was predicted by Gray et al. in the regions at Asn^4 to Cys^7 and Gly^8 to Tyr^{11} by Chou and Fasman's way[14,15,39].

In this stage, it could be said that the places where ordered structures, α-helix or β-turn, had been predicted correspond roughly to the two regions where turn I structures were found through the NOE connectivities. predicted.

Distance constraints

The inter-residual NOE connectivities which were detected in the NOESY spectra and used for the calculation are presented in a diagonal plot in Fig.7.

The intensities of the NOE peaks were translated to corresponding atomic distances in the following way[24].

The observed NOE's were classified into two groups, i.e. short range NOE's --- which are intra-residual NOE's and those between backbone hydrogen atoms and C_β protons of the adjacent residues --- and long range NOE's --- which are those between the remaining pairs of protons. The intensities of short range NOE's were interpreted by a rigid model. In this model, the fluctuation of the distance was assumed to be so small that the intensity of an NOE should be related to the atomic distance

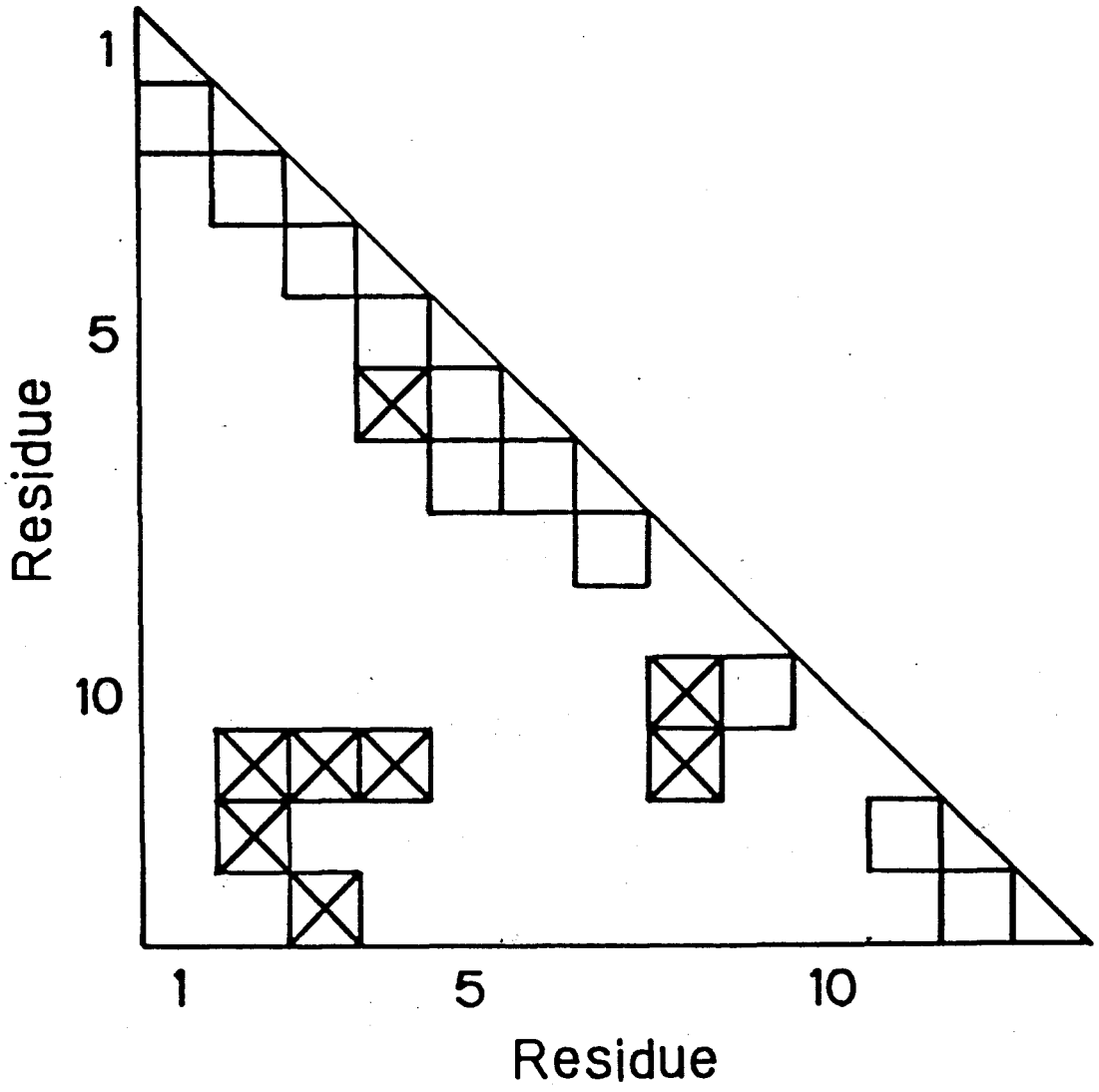


Fig. 7 Diagonal plot of the NOE connectivities used as distance constraints in the conformation determination of conotoxin GI in DMSO-d_6. Both axes represent the amino acid sequence. Open squares indicate that NOE's between backbone protons of two residues corresponding to each sequential location were observed. Squares with cross indicate NOE's between a proton of a residue and a side chain proton of another residue. Inter-residual side chain - side chain NOE was not detected.

according to an equation of the general form, NOE $\propto 1/r^6$.

By counting the contour lines in contour plots of the NOESY spectra, the short-range NOE's were classified as weak, medium and strong, and the corresponding upper distance constraints were calibrated relative to the standard distance so as to be 3.5, 3.0 and 2.5Å respectively[40]. The standard to calibrate relative intensities was provided by the intensity of the NOE between two C_β protons of a Cys residue that showed the strongest NOE intensity among all pairs of C_β protons. The distances between these protons were estimated to be 1.75A. The upper distance constraints between the paired atoms which show long range NOE's were assumed to be 4A adopting the idea of the threshold value of distance where NOE first was detected. Such a treatment of the long range NOE's was discussed previously[26].

The lower distance constraints were assumed throughout to be the sum of the core radii of the two protons, i.e., 2.0Å. The additional distance constraints were provided by the two disulfide bonds. When some of NOE peaks for side chain protons are not able to be assigned stereospecifically, the interpretation of these NOE's were carried out by using so-called pseudoatoms as reference points for distance constraints following the way suggested by Wüthrich et al[32].

Distance geometry calculation

In order to elucidate the three dimensional structure of conotoxin GI in solution on the basis of experimental interproton distances determined by NOE measurement, Braun and Gō's method was used[25]. It consists of the minimization of the square sum

of the differences between atomic distances in the calculated structure and the corresponding distances obtained from NOE information was used as follows. The details of the procedure are available in the previous paper with the example of applications on a toxic peptide, STh[26].

First, the following error function T for a given set of constraints is determined as target function to be minimized.

$$T = \sum' \frac{\left(U_{ij}^2 - r_{ij}^2\right)^2}{U_{ij}^2} + \sum' \frac{\left(L_{ij}^2 - r_{ij}^2\right)^2}{L_{ij}^2} + \sum' w \left[(s_i + s_j)^2 - r_{ij}^2\right]^2$$

where r_{ij}, U_{ij} and L_{ij} are interatomic distance between atoms i and j, upper and lower constraints of corresponding r_{ij}, respectively. s_i and s_j are repulsive core radii of i and j atoms. Σ' means summation only over the terms that violate distance constraints and w is the relative weight of the third term to the first and the second terms.

Next, an initial conformation is generated using random values of dihedral angles and r_{ij} are calculated. Then the target function is minimized by changing the values of the dihedral angles, and with fixed values of bond angles and bond lengths of peptide residues. In order to avoid the multiple minimum problem during the minimization, the distance constraints were applied stepwise according to the difference between residue numbers to which each of the paired protons under consideration belongs. By starting from the conformation obtained at the previous stage, a series of the minimizations are carried out with constraints between further residues along the primary structure.

These calculations were carried out with the computer program, "DADAS", written by Braun and Gō[25].

Results of the calculations

Starting from a hundred of initial conformations which were chosen individually and randomly, an assembly of a hundred conformers were obtained by the repetition of the calculations. Although the value of the target function T should vanish if the resulting conformation satisfies the all constraints, each resulting conformer remains actually some values for the target function at the final stage of each series of minimization. These values indicate how each conformer satisfies the constraints. The resulting conformers were numbered as I, II, III,... according to the residual value of the target function from the smallest value. Ten conformers from I to X which satisfies the constraints well were picked up and compared to judge how these structures converge to each other. The convergence was examined in the term of root-mean square distances (rmsd) among each of the ten conformers.

The analysis of these rmsd values reveals a following interest relationship among the conformers. As shown in Table 2, they are classified into two groups, A and B. Group A consists of the conformers I, IV, V and IX and group B consists of the conformers II, III, VI, VII, VIII and X. Although the average value of the rmsd calculated for back bone atoms among all ten conformers is 2.32Å, that among the conformers of group A is 1.58Å and 1.45Å for group B. These values calculated for all atoms are 3.54Å, 2.52Å and 2.61Å, respectively. This fact shows that the minimization lead two kinds of conformations. The difference of conformations between the two groups look to come

I	0.24	*	2.46	2.21	2.73	4.69	4.91	4.80	4.86	4.74	4.57
IV	0.39	1.77	*	1.67	3.30	4.30	4.55	4.04	4.09	4.25	3.82
V	0.47	1.48	1.05	*	2.93	4.15	4.41	3.90	4.14	3.99	3.73
IX	0.57	1.77	2.13	1.67	*	3.90	3.70	4.80	4.35	4.67	4.40
II	0.29	3.90	3.16	3.18	2.78	*	1.74	2.98	2.61	3.22	2.93
III	0.29	3.62	2.92	2.94	2.36	1.06	*	3.26	2.30	3.38	3.36
VI	0.50	3.56	2.61	2.66	2.93	1.62	1.70	*	2.34	2.13	1.63
VII	0.56	3.69	2.61	2.82	2.91	1.64	1.46	1.13	*	3.13	2.42
VIII	0.56	3.63	2.97	2.94	2.91	1.89	1.65	1.43	1.73	*	2.34
X	0.59	3.53	2.60	2.60	2.88	1.67	1.77	0.35	1.21	1.43	*
		I	IV	V	IX	II	III	VI	VII	VIII	X

Table II Root-mean square distances (rmsd) among ten conformers with the lowest ten residual values of the target function. First column shows the residual values of the target function. The lower-left triangle shows the rmsd calculated for the backbone atoms and the upper-right triangle shows the rmsd for all atoms. [Å]

mainly from the different orientation of the side chain of Tyr^{11}.

The reason why the minimization led to converge into two conformers are not clear so far. However the main reason should come from the lack of NOE connectivities found around Tyr^{11} residue. The work to investigate the dynamics of this molecule is under the course. In this stage both conformers should be candidates of the conformation of conotoxin GI.

The resulting conformers of each group are illustrated respectively in Figs.8 and 9 by drawing the bonds connecting the backbone atoms, N, C_α and C' atoms superimposed to get the best fit in the space.

Conformation of conotoxin GI in solution

The aim of this structural study is to provide informations for structure-activity relationships of conotoxin GI. Although there have been discussions of the structure resemblance between conotoxin GI and other toxins in literatures, they are based on hypothetical structures of conotoxin GI derived by Chou and Fasman's way of the prediction. Using the structure presented in Figs.8 and 9, even with the limitation that DMSO was used as solvent, we can now testify these hypothetical structures and cultivate a better understanding the structure-activity relationships.

A series of alkaloids, which block the neurotransmission at the nueromuscular junction, for example, d-tubocurarine and C-alkaloid E, have been investigated extensively by X-ray analysis and the structural requirements of such antagonists to the nicotinic cholinergic receptors for the physiological activity

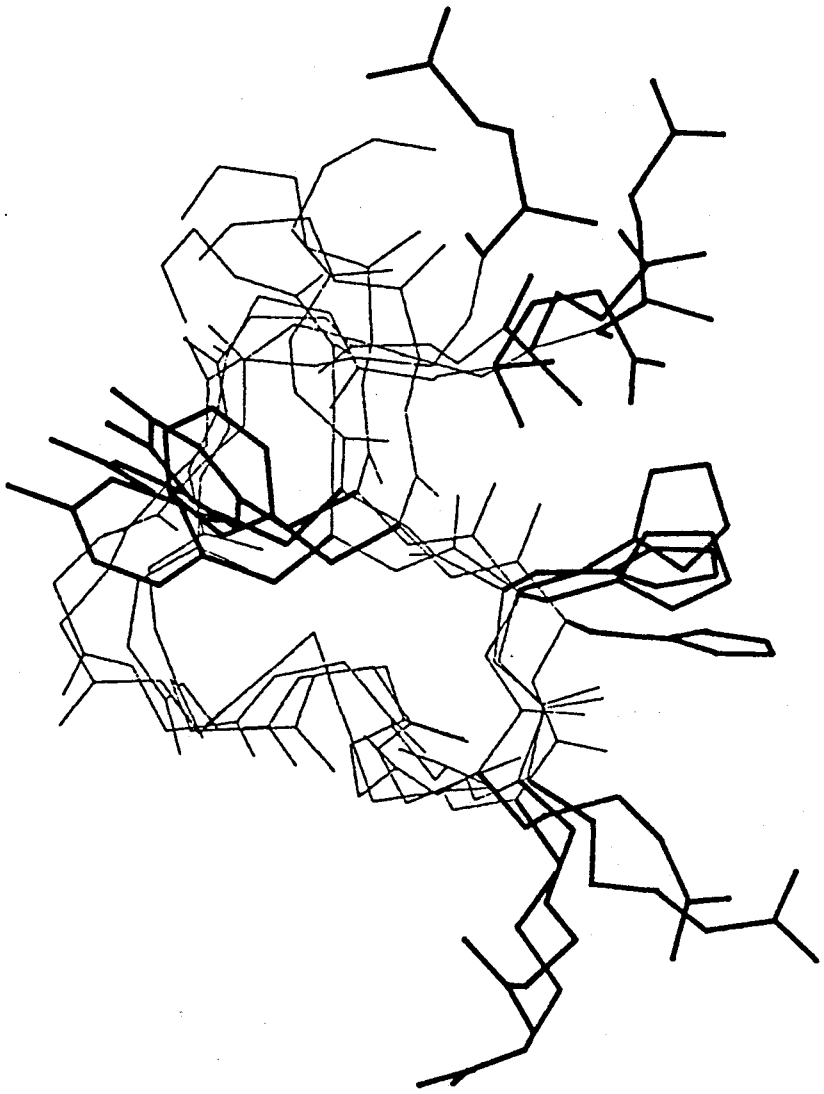


Fig. 8 The backbone conformations of the four conformers; I,IV, V and IX of group A, of conotoxin GI calculated by DADAS plotting only backbone atoms N, Cα, C' and the sidechain of Glu^{1},Arg^{9},His^{10} and Tyr^{11} were drawn in bold lines. The structures were superimposed so as to get the best fit in the space.

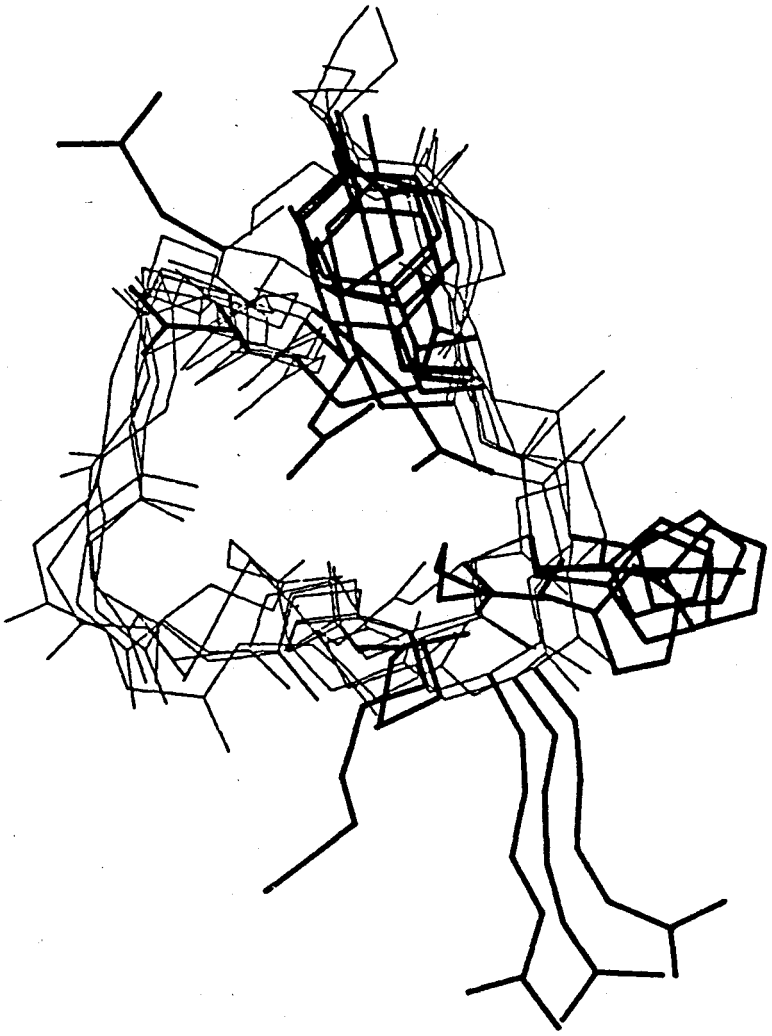


Fig. 9 The backbone conformations of the six conformers; II, III, VI, VII, VIII and X of group B, of conotoxin GI calculated by DADAS. Plotted as the same way in Fig. 8.

are elucidated by Pauling and Petcher[18] as follows, (i) rigid conformation (ii) two quaternary nitrogen atoms separated by 10.8 ± 0.3Å (iii) a sufficient degree of lipophilicity in the gap of the two cationic centers (iv) oxygen atoms disposed to aid the molecular orientation.

It was demonstrated that these requirements could be satisfied in the case of neurotoxic polypeptides of snake venoms by Dufton and Hider. They pointed out that the active tip of erabutoxin b, one of snake venomes, could have a disposition of amino acid side chains to fit these requirements as that Arg^{33} and Lys^{47} play the cationic centers and Phe^{32} and Tyr^{29} provide the lipophilicity[20,19]. Such a disposition itself is not found in the structure determined by X-ray analysis of the crystal of erabutoxin B[41]. They postulated the existence of some flip-flop movements around the active tip in solution to modify the structure so as to explain the resemblance of their activities. The modification of the structure standing on the conformational equilibrium in solution was supported by a certain degree of flexibility of the conformation of the neurotoxic peptide detected by NMR and CD studies[19,21,22].

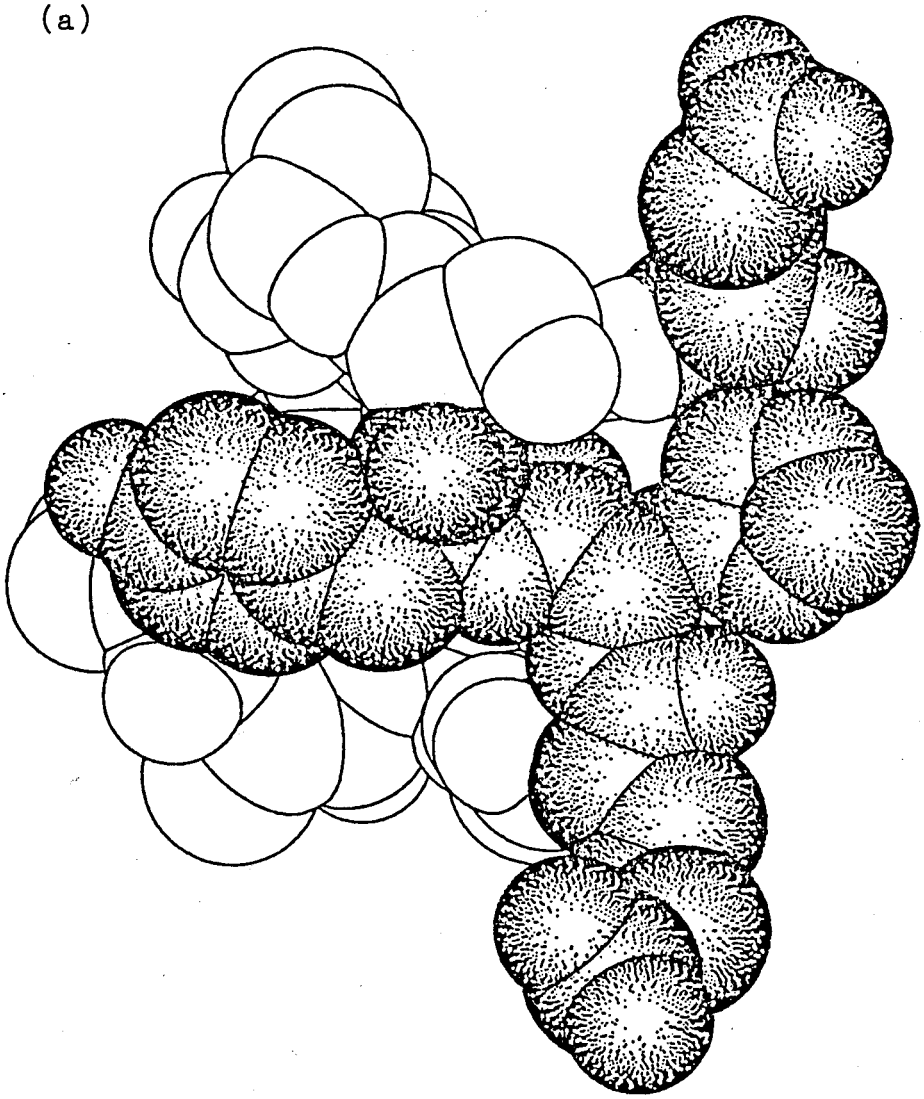
Concerning invariant residues and the predicted secondary structure of α-conotoxin, and the resemblance of the neurotoxic activities of conotoxin GI to those of snake venoms, it was postulated that the N-terminus of Glu^{1} plays one of the cationic center, the side chain of Arg^{9} plays another cationic center and His^{10} and Tyr^{11} stand between these two centers[13].

The spatial arrangements of these four amino acid residues of conotoxin GI in the calculated structures are shown in Figs.8

and 9. It is obvious that the arrangements of these residues in the structures of group A in Fig.8 fit surprisingly well to the postulated arrangement. In Fig.10, the space filling drawing of the conformer I is given as a representative conformation of group A.

The structures of B in Fig.9 gives better agreement with the model of conotoxin GI built by Gray et al. according to the prediction of secondary structure which shows a strong resemblance to that of the active tips of erabutoxin B[14]. Although the backbone structures look like quite similar each other, there are, of course, disagreements found in details which must come from the overestimations during the prediction. A salt bridge between the side chains of Glu^1 and Arg^9 postulated in the model was not found in the calculated structures so that its formation should have rather low probability. There might be an explanation of the difference that the solvent used in this structure determination was DMSO while the prediction was carried out in the expectation of an aqueous solution. In fact, the pH titration experiment on the aqueous solution of conotoxin GI did not give any indication of such a salt bridge[35]. Although the presence of regularly tight turn conformation in the calculated structure is not so clear, the manner of the chain fold of the backbone looks essentially consistent to that of the model which was determined by Gray et al. so as to form two β-turns initiating at Asn^4 and Gly^8. The profile of NOE connectivities in the NMR spectra of conotoxin GI and the dihedral angles of the backbone chains of the calculated structure indicate that some highly constrained structures exist around the segments where β-

(a)



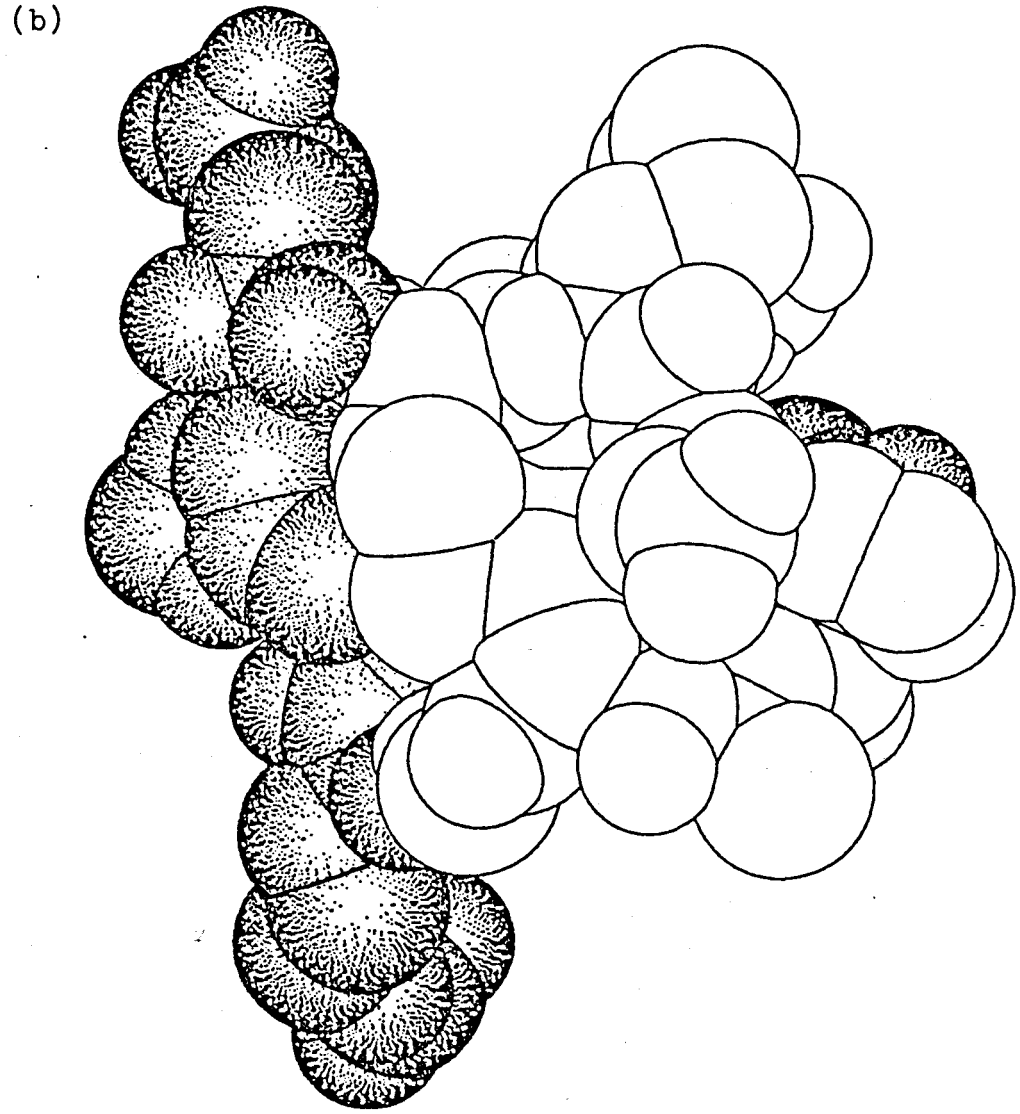


Fig. 10 The space filling drawing of conformer I of with lowest value of the target function of DADAS as a representative structure of conotoxin GI in solution. The four residues related to conformational requirement for biological activities, Glu^{1}, Arg^{9}, His^{10} and Tyr^{11}, are shadowed. (a) is the front view of the molecule where these four residues were localized and (b) is the reverse view.

turns were postulated. They suggest also that the possibility of the presence of α-helical conformation in the region between Ala^6 to His^{10} which was predicted by Hider to explain the CD spectrum of conotoxin GI indicating the high α-helical content up to 50% should be low. The CD pattern could reflect some partial arrangements of dipole moments of peptides in the constrained segment.

While there remain some ambiguities in the detailed structure because of rather poor convergence, we have thus demonstrated that the conotoxin GI takes a spatial atomic arrangement in solution essentially similar to those of a series of alkaloids of antagonists to the nicotinic cholinergic receptors. These similarities are of the same sort which were postulated by Dufson and Hider in the cases of neurotoxin peptides of snake venoms.

In order to obtain further information of structure-activity relationship of neurotoxin we are now in the course of structure determination of chemical derivatives of conotoxins GI and MI of which biological activities have been investigated by comparing with those of native ones. These results should provide some information of nicotinic receptors as well.

References

1. Kohn, A.J., Sanders, P.R. & Wiener, S. (1960) Ann. NY Acad. Sci., 90, 706-725.
2. Endean, R. & Rudkin, C. (1963) Toxicon, 1, 46-64.
3. Spence, I., Gillessen, D., Gregson, R.P. & Quinn, R.J. (1977) Life Sci, 21, 1759-1770.
4. Cruz, L.J., Gray, W.R. & Olivera, B.M. (1978) Arch.. Biochem. Biophys., 190, 539-548.
5. Olivera, B.M., Gray, W.R., Zeikus, R., McIntosh, J.M., Varga, J., Rivier, J., de Santos, V. & Cruz, L.J. (1985) Science, 230, 1338-1343.
6. Gray, W.R., Luque, F.A., Galyean, R., Atherton, E.,Sheppard, R.C., Stone, A., Reyes, A., Alford, J., McIntosh, M., Olivera, B.M., Cruz, L.J. & Rivier, J. (1984) Biochemistry, 23, 2796-2802.
7. McIntosh, M., Cruz, L.Z., Hunkapiller, M.W., Gray, W.R. & Olivera, B.M. (1982) Arch, Biochem. Biophys. 218, 329-334.
8. Olivera, B.M., McInroah, J.M., Cruz, L.J., Luque, F.A. & Gray, W.R. (1984) Biochemistry, 23, 2087-2090.
9. Olivera, B.M., Cruz, L.J., de Santos, V., LeCheminant, G.W., Griffin, D., Zeikus, R., McIntosh, J.M., Galyean, R., Varga, J., Gray, W.R. & Rivier, J. (1987) Biochemistry, 26, 2086-2090
10. Cruz, L.J., Gray, W.R., Olivera, B.M., Zeikus, R.Z., Kerr, L., Yoshikami, D. & Moczydlowski, E. (1985) J. Biol. Chem., 260, 9280-9288.

11. Gray, W.R., Luque, A., Olivera, B.M., Barrett, J. & Cruz, L.J. (1981) J. Biol. Chem., 256, 4734-4740.
12. Nishiuchi, Y. & Sakakibara, S. (1982) FEBS Lett., 148, 260-262.
13. Hider, R.C. (1985) FEBS Lett., 184, 181-184.
14. Gray, W.R., Middlemas, D.M., Zeikus, R., Olivera, B.M. & Cruz, L.J. in "Peptides: Structure and Function", eds.Deber, C.M., Hruby,V.J. and Kopple,K.D. (1985) Pierce Chem. Co., pp. 823-832.
15. Chou, P.Y. & Fasman, G.D. (1974) Biochemistry, 13, 211-222.
16. Yang, C.C. (1974) Toxicon,12, 1-43.
17. McManus, O.B., Musick, J.R. & Gonzalex, C. (1981) Neurosci. Sci., 5, 53-56.
18. Pauling, P. & Petcher, T.J. (1973) Chem. Biol. Interact., 6, 351-365.
19. Hider, R.C. & Dufton, M.J., in "Natural Toxins", eds. Eakere, D. & Wadstrom, T. (1979) Pergamon, pp. 515-522.
20. Dufton, M.J. & Hider, R.C. (1977) J. Mol. Biol., 115, 177-193.
21. Dufton, M.J. & Hider, R.C. (1980) Trend Biochem. Sci., 5, 53-56.
22. Dufton, M.J. & Hider, R.C. (1983) CRC Crit. Rev. Biochem. 14, 113-171.
23. Low, B.W., Preston, H.S., Sata, A., Rosen, L.S. Searl, J.E., Rudko, A.D. & Richardson, J.S. (1976) Proc. Nat. Acvad. Sci., U.S.A. 73, 2991-2994.
24. Braun, W., Bösch, C., Brown, L.R., Gō, N. & Wüthrich,K. (1981) Biochim. Biophys. Acta, 667, 337-396.

25. Braun, W. & Gō, N. (1985) J. Mol. Biol., 186, 611-626.
26. Ohkubo, T., Kobayashi, Y., Shimonishi, Y., Kyogoku, Y., Braun, W. & Gō, N. (1986) Biopolymers, 25, s123-s134.
27. Kumar, A., Ernst, R.R. & Wüthrich, K. (1980) Biochem. Biophys. Res. Commun., 95, 1-6.
28. Rance, M., Sorenson, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. & Wüthrich, K. (1983) Biochem. Biophys. Res. Commun., 117, 479-485.
29. Davis, D.G. & Bax, A. (1985) J. Am. Chem.Soc., 107, 2820-2821.
30. States, D.J., Haberkorn, R.A. & Ruben, D.J. (1982) J. Magn. Reson., 48, 286-292.
31. Iga, Y. & Yasuoka, N. (1984) J. Mol. Graph., 2, 79-82.
32. Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids. J. Wiley, New York
33. Torchia, D.A. (1972) Biochemistry, 11, 1462-1468.
34. Cheng, H.N. & Bovey, F.A. (1977) Biopolymers, 16, 1465-1472.
35. Kobayashi, Y., Yamazaki, K., Ohkubo, T., Nishimura, S., Kyogoku, Y., Nishiuchi, Y. and Sakakibara. S. unpublished data
36. Gray, W.R., Rivier, J.E., Galyean, R., Cruz,L.J. & Olivera, B.M. (1983) J. Biol Chem., 258, 12247-12251.
37. Nishiuchi, Y. & Sakakibara, S., in "Peptides 1984", ed. Rangnarsson, U. (1984) de Gruyter, pp. 537-540.
38. Wüthrich, K., Billeter, M. & Braun, W. (1984) J. Mol. Biol., 180, 715-740.
39. Wemmer, K. & Kallenbach, N.R. (1977) Biochemistry,22, 1901-1906.

40. Williamson, M.P., Havel,T.F. & Wüthrich, K. (1985) J. Mol. Biol., 182, 295-315.

41. Kimball, M.R., Sato, A., Richardson, J.S., Rosen, L.S. & Low, B.W. (1979) Biochem. Biophys. Res. Comm., 88, 950-959.

CHAPTER III

A Comparison of the Polypeptide and Protein Structures Determined by Three Different Methods; DISGEO, DADAS and CHARMM

[Summary]

The reliability and limitation on the determination of solution structure are discussed through the comparison of three algorithms summarized representatively in the corresponding computer programs ; DISGEO, DADAS and CHARMM. Comparisons among these three computer programs with the same input data which have not been reported so far, are carried out here with the input data derived from experimental NOE data for a peptide ; conotoxin GI, and from a set of simulated distance constraints for a protein ; Tendamistat. From the aspect of the application merits and demerits of each method are discussed in the case of structure determination by NMR.

[Introduction]

Various spectroscopic measurements offer informations related to mutual distances between atoms and dihedral angles. These data have made it possible to construct the solution structure of protein or polypeptide with model building or computer calculation. In order to realize it several approaches are available up to date[1-3]. Distance constraints between pairs of protons obtained from nuclear Overhauser effect (NOE) measurements[4] play an inevitable role in these approaches. In contrast to X-ray diffraction analysis which usually offer unique structure, solution structures are represented as an assembly of the various conformers. They should be consistent with the boundary of the conformational space of the allowed structure in solution defined by NMR data[5]. In this study we attempted to discuss about the reliability and limitation on solution structure determination with these methods through the comparison of three representative ones summarized in the corresponding computer programs ; DISGEO[6-8], DADAS[9,10] and CHARMM[11-13].

DISGEO was written by Havel et al[8]. The algorithm of this program is based on the Crippen's metric matrix method[14,15] in which distances between all pairs of atoms are converted into atomic coordinates and expanded in DISGEO to treat the imprecise distance data which are, for example, obtained from NMR experiments. DADAS was written by Braun and Gō[9]. In this program, least-square minimization of a suitable target function

which is a measure how good the distance constraints are satisfied, is performed in the dihedral angle space. In order to avoid local minima, the target function itself is treated as variables. CHARMM is a set of programs written by Karplus et al[16]. It is equipped with the facility of restrained molecular dynamics which is used in this study. Restrained molecular dynamics[11,12,17] dissolve the classical equations of motion in the molecule and to trace the trajectory of the molecular motion under the condition where distance constraints are incorporated into the total energy function as pseudo potentials. The application of CHARMM program to analyze the NMR data to elucidate the conformation of macromolecules was done by Clore et al[13,18].

Comparison between DISGEO and CHARMM with a simulated NMR data from crambin crystal structure has been reported by Clore et al[19]. One between DISGEO and DADAS with experimental NMR data of bovine pancreatic trypsin inhibitor (BPTI) has been reported by Wagner et al[20]. Comparisons among these three computer programs with the same input data which has not been reported so far, are carried out here with the input data derived from experimental NOE data of a peptide; conotoxin GI[21], and from a set of simulated distance constraints of a protein ; Tendamistat[22].

[Materials and Methods]

Conotoxin GI and Tendamistat

Conotoxin GI was isolated from the venom of Conus geographus which is a fish-eating marine snails of genus Conus and shown to block synaptic transmission by binding to the acetyl choline receptor to cause paralysis[21]. It was chemically synthesized and the mode of their disulfide formations were revealed (Cys2-Cys7 and Cys3-Cys13)[23].

The α-Amylase inhibitor, Tendamistat from Streptomyces tendae binds tightly to mammalian α-amylase[22]. Tendamistat of 74 amino acid residue has two disulfide bridges (Cys11-Cys27, Cys45-Cys73) and the proposed active-site residues, Trp18 - Arg19 - Tyr20. This is also well known by the fact that the conformational studies on this sample has been carried out independently and in parallel by the two techniques ; X-ray diffraction[24] and NMR[10], and showed a good agreement (Fig. 1), especially a close correspondence in the chain fold, both results (Fig 1). The profile of the chain fold of Tendamistat[24] is that the two twisted β-sheets of three anti-parallel strands each are gathered to form a β-barrel . A sheet is consisted of the strands 11 to 17 (I), 19 to 26 (II) and 51 to 57 (V), and another in the opposite side is consists of the strands 64 to 73 (VI), 30 to 37 (III) and 40 to 49 (IV). There are three b-turns at 17-20, 37-40 and 49-52.

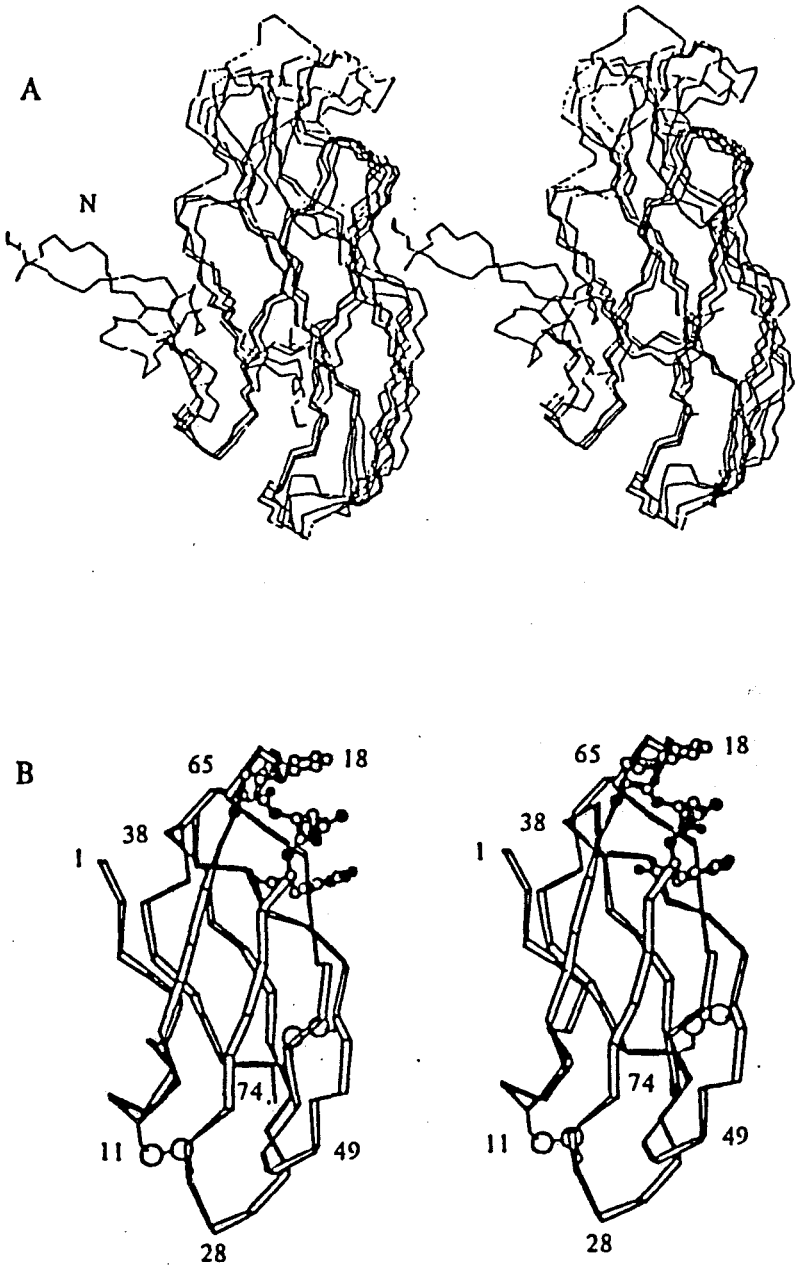


Fig. 1. Stereo drawings of the backbone structures of a-amylase inhibitor Tendamistat in solution and crystal as determined by NMR (a) and X-ray diffraction techniques (b).

Input distance constraints

Conotoxin GI

The distance constraints of conotoxin GI derived from actual NOE data observed in a NOESY spectrum[25] of mixing time 120 msec are shown in Chap II. A total of 81 NOE's were collected; that is, 22 NOE's between intraresidual protons, 31 NOE's between protons of neighboring residues and 28 NOE's between non-adjacent residual proton's. NOE's which are between intraresidual protons, and between backbone hydrogen atoms and C^{β} protons of the adjacent residues were treated with the rigid model and interpreted to upper distance constraints of $2.5\mathring{A}$, $3.0\mathring{A}$ and $3.5\mathring{A}$ according to their intensities. For the other NOE's, distance constraints were set to $5.0\mathring{A}$[5,31]. Additional distance constraints for the formation of the two disulfide bridges of $2.04\mathring{A}$ on the S-S distance and of $3.05\mathring{A}$ on the S-C^{β} distances across each bridge were used[26].

Tendamistat

The distance constraints of Tendamistat were derived from X-ray crystal structure[24] by following procedure. In order to obtain the simulated distance constraints which could in principle be collected in the NOESY spectra, first hydrogen atom must be attached to crystal structure, and then the problem of prochiral protons and resolution and precision of the NOESY experiment should be taken into account. Hydrogen atoms were added to the crystal structure of Tendamistat using a standard procedure of the regularization[9] with the standard geometry of

the amino acid residues in the Empirical Conformational Energy Program for Peptide (ECEPP)[26,27]. In the course of the regularization, all variable dihedral angles including ω were treated as variable and the positions of unconstrained protons were determined at the stereochemically comfortable one. Root-mean-square distance (r.m.s.d.)[28] of the backbone atoms between regularized and crystal structure was within 0.1$\mathring{A}$ for backbone atoms. As an NOE should be observable when the corresponding distance of proton pair is at most within 4$\mathring{A}$ or 5$\mathring{A}$[29], only distances less than 4$\mathring{A}$ among number of proton distances from the regularized crystal structure were in consideration. In the case of groups of protons for which stereospecific assignments are impossible in the standard procedure, such as methylene protons, peripheral two amide protons, methyl protons, and aromatic protons of Tyr and Phe[1], $(\langle r^{-6} \rangle)^{1/6}$ mean distances were used in the evaluation of distances[30]. A $(\langle r^{-6} \rangle)^{1/6}$ mean distance is represented as follows:

$$(\langle r^{-6} \rangle)^{1/6} = \Sigma' (rij^{-6}/n)^{-1/6}$$

where n is the number of protons in the group.

First, all of the labile protons which are expected to vanish in the aqueous solution, were omitted in the list[1]. Furthermore, for the case of proton pairs which involved at least one amide, aromatic or methyl proton, distances within 4.0A were picked up and for the other case distances within 3.0A were picked up. This reflects that ^{1}H-NMR peaks of these protons are expected to locate at well separated regions and easily assigned in the two-dimensional NMR spectra. At this stage, total number of distance constraints derived from crystal structure of

Tendamistat was 960. Since this number might exceed the one found in the NOESY spectra of such a protein, we diminished the number of constraints in order of 401, which was reported in the paper of solution structure determination of Tendamistat with NMR[10]. After all, a total number of 452 distance constraints was used for the input data of each programs which include 75 intra-residue distance constraints, 142 sequential distance constraints and 235 non-adjacent residual distance constraints. As the nature of the distance evaluation from NOE intensity is semi-quantitative, the intra and sequential distances were divided into three classes; for all distances within 2.5$\mathring{A}$, upper distance constraints were set to be 2.5$\mathring{A}$, for those from 2.5$\mathring{A}$ to 3.0$\mathring{A}$ were to be 3.0$\mathring{A}$ and for those from 3.0$\mathring{A}$ to 4.0$\mathring{A}$ were to be 4.0$\mathring{A}$. These distances were corresponding to strong, medium and weak NOE's. For all distances between non-adjacent residual protons, distance constraints were set to 4.0$\mathring{A}$[31]. That takes account of the fluctuation of distances between far separated protons along the primary structure.

Computational methods

Metric matrix method

In metric matrix method, distances between all pairs of atoms are converted into atomic coordinates elegantly by use of the metric matrix[14,15]. The metric matrix G for an actual molecule with N atoms is defined by

$$G_{ij}=r_i \cdot r_j \quad (i,j=1, \cdots, N)$$

where r_i represents the vector from the geometric center of mass

to atom i. Thus G_{ij} is an N x N symmetric matrix. The matrix G_{ij} is calculated from the distances between atoms i and j, d_{ij} as follows:

$$G_{ii} = \frac{1}{N}\sum_{j}^{N} d_{ij} - \frac{1}{2N^2}\sum_{j}^{N}\sum_{k}^{N} d_{jk}$$

$$G_{ij} = \frac{1}{2}(G_{ii} + G_{jj} - d^2_{ij})$$

A spectral representation of metric matrix by diagonalization is,

$$G_{ij} = \sum_{\alpha}^{N} \lambda_{\alpha} E_{i\alpha} E_{j\alpha}$$

where λ_{α} is the eigenvalue and E_{α} corresponding eigenvector.
A comparison of the two different representations of G_{ij} leads to

$$r^{\alpha}_{i} = \sqrt{\lambda_{\alpha}} E_{i\alpha}$$

If all distances between all pairs of atoms are exactly known, the metric matrix has the important properties of rank 3. This means that number of non-zero eigenvalues is three. Then coordinate of atom; (x_i, y_i, z_i) is embedded as follows;

$$x_i = \sqrt{\lambda_1} E_{i1},\ y_i = \sqrt{\lambda_2} E_{i2},\ z_i = \sqrt{\lambda_3} E_{i3}$$

Therefore atomic coordinates can be calculated from interatomic distances. This procedure is usually named as embedding.

But distance information elucidated from NMR measurements takes a form of a distance constraint and is obtained only for a small subset of all possible atom pairs. Therefore it is an

unrealistic assumption to know all the distances from NMR data exactly. In DISGEO program[6,8], this problem was solved as follows. To calculate a metric matrix, distances are chosen at random within the given distance constraints. A distance large enough to cover whole molecule is set to upper distance constraints for the interatomic distances of which distance informations are not available from experimental data. The range of distance constraints is narrowed by lowering the upper bounds and raising the lower bounds using following triangle inequalities[5].

$$U_{ij} < U_{ik} + U_{jk}$$

$$L_{ij} > L_{ik} - U_{jk}$$

Actually rank of a metric matrix calculated from randomly chosen distances is not three so that the approximation of the three greatest eigenvalues is used in the embedding procedure.

Least-square minimization method

Least-square minimization method[5] minimize a suitable target function. In order to obtain the structure which satisfies imposed distance constraints , an arbitrary molecular structure is generated and is changed the conformation in such a way that a target function becomes zero. The target function is a measure of how good the distance constraints are satisfied. Braun and Gō defined the following error function as a target function to be minimized.[5]

$$T = \Sigma' \frac{(U_{ij}^2 - r_{ij}^2)^2}{U_{ij}^2} + \Sigma' \frac{(L_{ij}^2 - r_{ij}^2)^2}{L_{ij}^2} + \Sigma' w * [(s_{io} + s_{jo})^2 - r_{ij}^2]^2$$

where Σ' means summation only over the terms which violate

distance constraints, and s_i, s_j are repulsive core radii[5] of i and j atoms , and w is a relative weight of the third term to the first and the second terms.

During the minimization of the target function, only the dihedral angles ϕ,ψ and χ's are treated as independent variables by fixing the bond angles and bond lengths of peptide residues at the values listed in ECEPP[26,27]. A structure is generated for a set of given values of dihedral angles and r_{ij} are calculated. Reductions of the calculation time and computer memory space were achieved by treating only dihedral angles as independent variable.[5,32,33] if one want to minimization a target function rapidly. It is a strong demand to know the first derivative of a target function with respect to a variable dihedral angle $\theta_a(\delta T/\delta\theta_\theta)$, This quantity is derived basing on the idea of Noguti and Gō[33] as follows:

$$\frac{\delta T}{\delta \theta_a} = -(e_a,\ e_a \wedge r_a)\ F_a,$$

$$F_a = \frac{1}{|r\alpha - r\beta|} \cdot \frac{\partial T}{\partial |r\alpha - r\beta|} \begin{bmatrix} r\alpha \wedge r\beta \\ r\alpha - r\beta \end{bmatrix}$$

where $\mathbf{r}_\alpha$ and $\mathbf{r}_\beta$ are position vectors of atoms α and β, respectively; r_a is the position vector of the atom at the C terminus end of bond a; e_a is the unit vector along the bond a from the N terminus to the C terminus; θ_a is the dihedral angle in consideration; Ma is a set of atoms on the moving side which are movable according to the rotation of θ_a under the condition that atoms on the N-terminal side are fixed.

In order to avoid the multiple minimum problem and to search

for the global minimum of the target function T, the distance constraints were applied stepwise following the idea of Ooi et al. as follows.[34] At first, the distance constraints were classified according to the distance k between the residues to which each of paired atoms under consideration belongs, where k is defined as the difference of the residue numbers. Next, minimization was carried out for the target function with the distance constraints of only smaller values of k. Then, by starting from the conformation obtained in this minimization, calculation was continued to minimize the target function with the constraints of up to slightly larger values of k. Starting from the previously minimized conformation, the target function was then minimized with the constraints of the second range between the residues aparted by four residues. A series of the minimization was continued by increasing the range of constraints. This program was written by Braun and Gō, and coded as DADAS[5].

Restrained molecular dynamics

Restrained molecular dynamics[11,12,17] dissolves the classical equations of motion for all atoms in the molecule for a suitable time period and trace the trajectory of the molecular motion under the condition that distance constraints are incorporated into the energy function as the pseudo potentials[35]. These pseudo potentials are main driving forces to obtain the structures which are consistent with NMR data. The simple integration algorithm used to solve classical Newton's equation of motion for molecular dynamics is that due to

Verlet[36]. The co-ordinates $r_i(t+\Delta t)$ of atom i can be described by the co-ordinates $r_i(t)$ and its derivations at time t using a Tailor expansion:

$$r_i(t+\Delta t) = r_i(t) + V_i\Delta t + m_i^{-1}F_i(t)(\Delta t)^2 + \cdots\cdots$$

The expansion for $-\Delta t$ yields:

$$r_i(t-\Delta t) = r_i(t) - V_i\Delta t + m_i^{-1}F_i(t)(\Delta t)^2 - \cdots\cdots$$

When adding these two expansions:

$$r_i(t+\Delta t) = 2r_i(t-\Delta t) + m_i^{-1}F_i(t)(\Delta t)^2 + o[(\Delta t)^4]$$

So, once the co-ordinate r_i and an interaction potential $V(r_1, r_2, \cdots, r_N)$ at time t and $t-\Delta t$ are given, the coordinates at time $t+\Delta t$ can be calculated by this equation. The time period Δt should be chosen so that $F_i(t)$ is essentially constant over Δt. This value is $10^{-16}<\Delta t<10^{-14}$sec. Actually the calculated structure in the olecular dynamics moves toward the energy minima with shaking at a setting temperature.

Many molecular dynamics programs[3] are available now. The program which we adopt in this study is CHARMM[16] which is one of well prevailed molecular dynamics programs. In CHARMM, pseudo potential $V_{NOE}(r_{ij})$ term which represents the interproton distance constraints is added to the total energy function of the molecule in following quadratic biharmonic form[11,12]:

$$E_{NOE}(r_{ij}) \begin{cases} = C_1(r_{ij}-r_{ij}^{o})^2, & \text{if } r_{ij}>r_{ij}^{o} \\ = C_2(r_{ij}-r_{ij}^{o})^2, & \text{if } r_{ij}<r_{ij}^{o} \end{cases}$$

where r_{ij} and r_{ij}^{o} are the calculated and target interproton distances respectively, and C_1 and C_2 are force constants given by:

$$C1 = \frac{k_B TS}{2(\Delta_{ij}^{+})^2}, \quad C2 = \frac{k_B TS}{2(\Delta_{ij}^{-})^2}$$

where S is a scale factor, and Δ_{ij}^{+} and Δ_{ij}^{-} are the positive and negative error estimates of r_{ij}^{o}, respectively. When S is unity, C_1 and C_2 mean that if the deviations of the actual distances r_{ij} from their target distances r_{ij}^{o} are equal to the error estimates Δ_{ij}^{+} or Δ_{ij}^{-}, the pseudo potential $V_{NOE}(r_{ij})$ increases by $1/2k_BT$.

[Result]

Calculational strategy

DISGEO program

Because prochiral protons such as methyl or methylene protons, are not able to be stereo-specifically assigned by standard procedures of NMR experiments[1]. Then input distance constraints for these protons were referred to pseudo atom positions, suitable distances were added on the distance constraints obtained from NOE's for prochiral protons as described by Wuthrich et al[37]. In order to reduce the calculation time and computer-memory space, these prochiral protons were replaced by pseudo atoms with suitable atomic radii. Bond length and bond angles of amino acid geometries which are standardized in the ECEPP[26,27], chirality constraints which define the chiral centers and planarity constraints of all sp^2 centers provide an additional set of constraints . Non-biased sampling of the initial structures was carried out by making a random selection of the interatomic distances within the given limit of distances [6,7].

The calculations of DISGEO program proceeded in following four phases[8].

1) All distance constraints were determined so as to satisfy the requirement of triangle inequality.

2) A substructure consisting of only backbone atoms was embedded first with the distances between paired atoms belonged to

backbone in order to avoid the entanglement of side chains.

3) Metric matrix for all distances was calculated using the distances between the atoms of backbone obtained in the calculated substructure, then projected from the higher-order distance space into the three dimensional space so as to convert distances into coordinates.

4) The resulting three dimensional structure was refined by restrained least-squares minimization.

A hundred of the structures were calculated with different initial values for the random number seeds. Five structures which satisfy the constraints well were picked up according to the value of residual violations for the NOE constraints.

DADAS program

The input distance constraints for prochiral protons in DADAS program were treated by the pseudo atom method similar to the case of DISGEO. In this study DADAS was modified to set the atomic radii of pseudo atoms to zero in order to incorporate all protons including prochiral protons into the calculation. This means that prochiral protons were not replaced by pseudo atoms. The latter were treated only as reference points on which distance constraints are imposed with the same distance corrections as the previous pseudo atom treatment did. This modification achieves correct packing of the molecules without loss of the merit of the pseudo atom treatment. Distance constraints involving residues further apart along the primary structure were gradually incorporated into the target function as described in Material and Methods.

Judging from our experience in the previous model study of BPTI, we had to keep the following two points in mind during the DADAS calculation. (Matsukuma, J. & Gō, N. unpublished data).
1) Local structure in early stage of the calculation sometimes affects seriously to final structure. Only correct construction of the local conformation leads to the correct overall folding.
2) Annealing of the calculated structure in the refinement is useful. Annealing means the re-calculation of the improperly folded structure in order to get a correct folding.

A hundred of the structures were calculated with randomly chosen initial structures. Five structures which satisfy the constraints well were picked up according to the residual value of target function.

CHARMM program

The treatment of the distances involving prochiral protons in CHARMM was quite different from those in DISGEO and DADAS. These distances were calculated as $(\langle r^{-6} \rangle)^{-1/6}$ mean distances(30) which are directly related to the experimental data of NOE, and hence no correction for the distance constraints were done. We modified the program to change a biharmonic form of effective pseudo potential of NOE to a form of square well potential, as follows.

$$E_{NOE}(r_{ij}) \begin{cases} = C(r_{ij}-U_{ij})^2 & , \text{ if } r_{ij}>U_{ij} \\ = 0 & , \text{ if } L_{ij}<r_{ij}<U_{ij} \\ = C(r_{ij}-L_{ij})^2 & , \text{ if } L_{ij}>r_{ij} \end{cases}$$

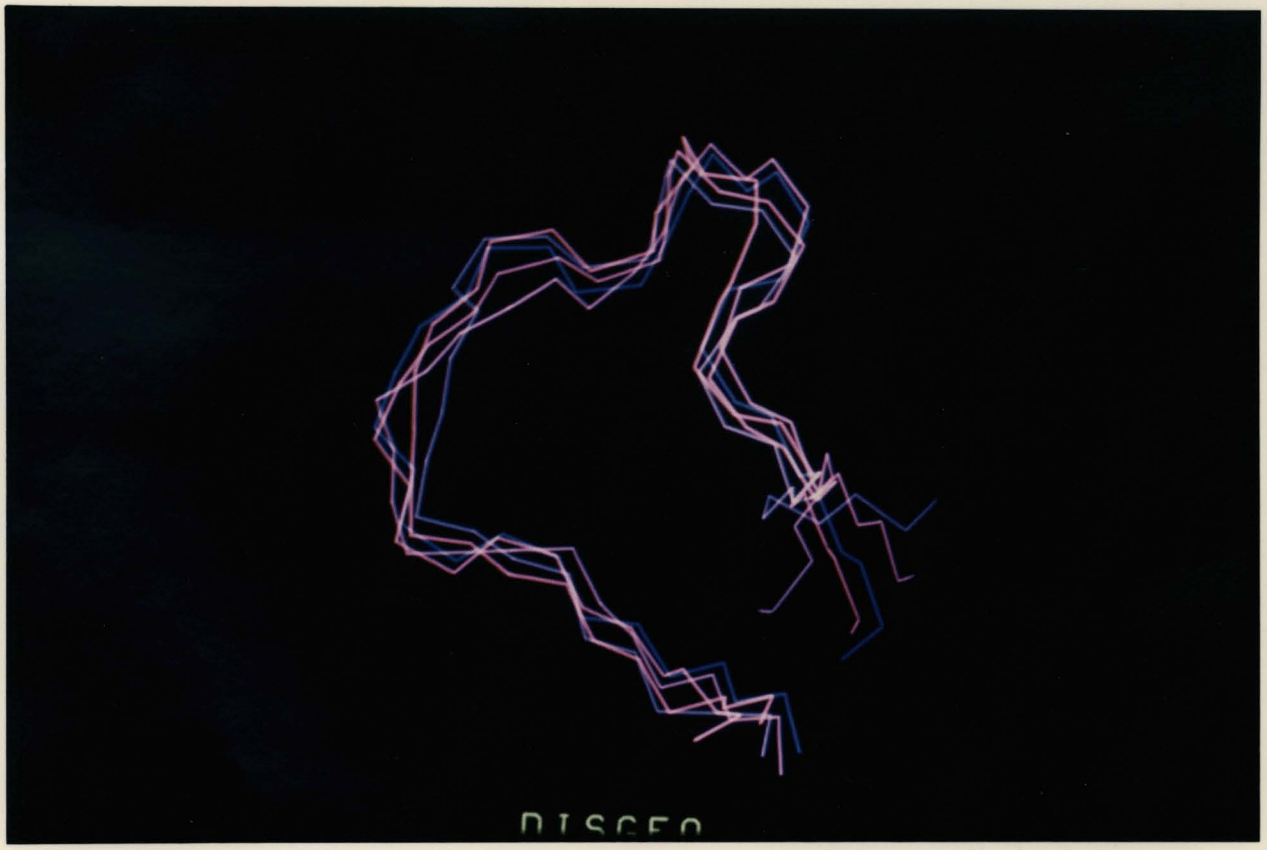
where U_{ij} and L_{ij} are the upper and lower distance constraints, respectively. Force constant C is described in Material and

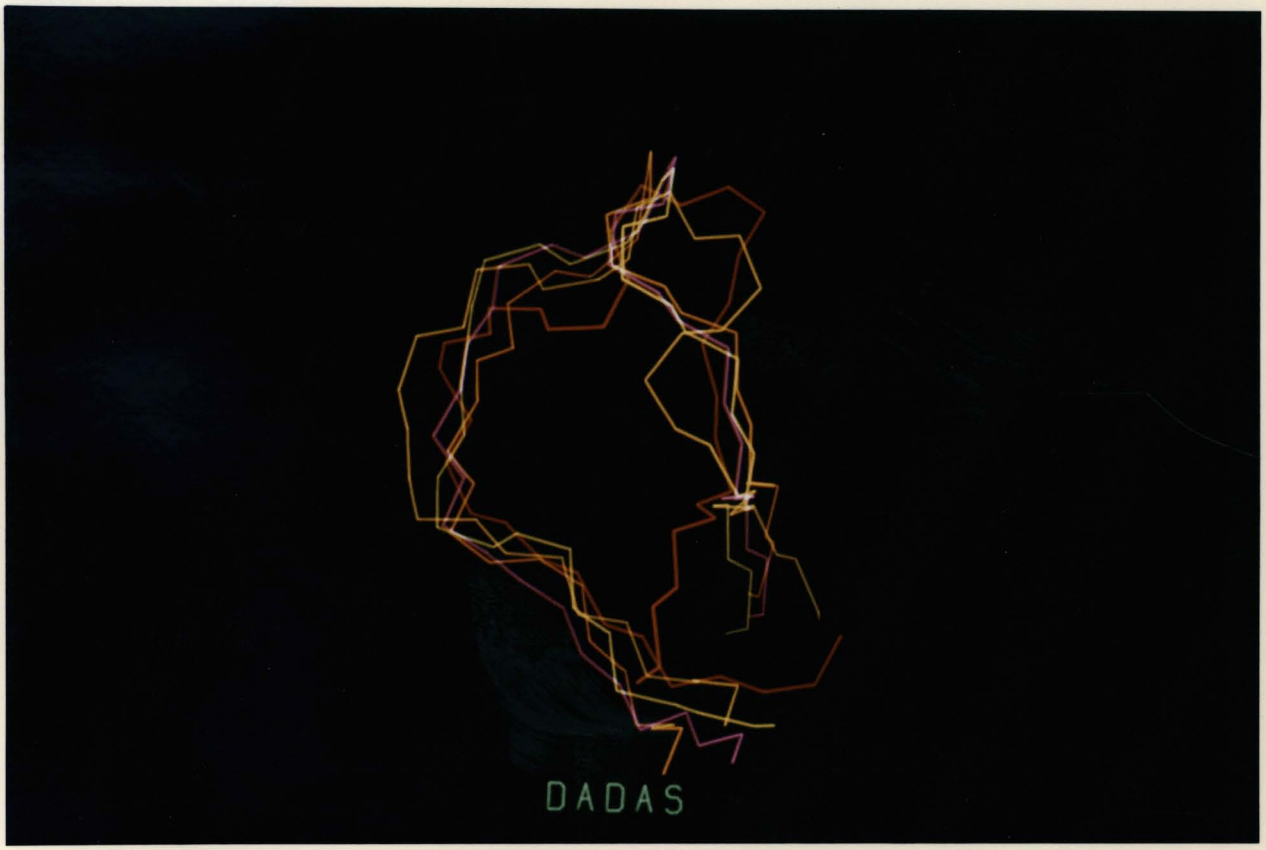
Methods. The calculations carried out first a structure determination and then proceeded to its refinement[12]. In the structure determination stage, constraints concerning to the s-s bridges were not incorporated into the calculations. The distance constraints involving residues separated along the primary structure were gradually incorporated into calculation like in the case of DADAS. In each step, trajectories of 5 through 10ps were calculated with the S value of the scale factor of 1.0 to 5.0. In the refinement stage all the distance constraints including S-S bridges were incorporated into calculation. The value of S was gradually increased from 1.0 to 5.0. Each 10ps molecular dynamics calculation was followed by the restrained energy minimization. A completely extended structure was chosen for the initial structure of the calculations and subjected to conjugate gradient energy minimization[38]. Calculations were carried out with changing the some calculational parameters, such as S.

Structure of conotoxin GI

In order to achieve quantitative evaluation of resulting structure, criteria were provided by residual NOE constraint violation values and r.m.s.d. values.[2,3].

First, each of the conotoxin GI conformations obtained by DISGEO, DADAS and CHARMM (Fig. 2) were evaluated by values of a sum over the residual violations which are listed in Table 1. Five structures from DISGEO and DADAS, and three structures from CHARMM were listed respectively there according to these values.


DISCEO


DADAS

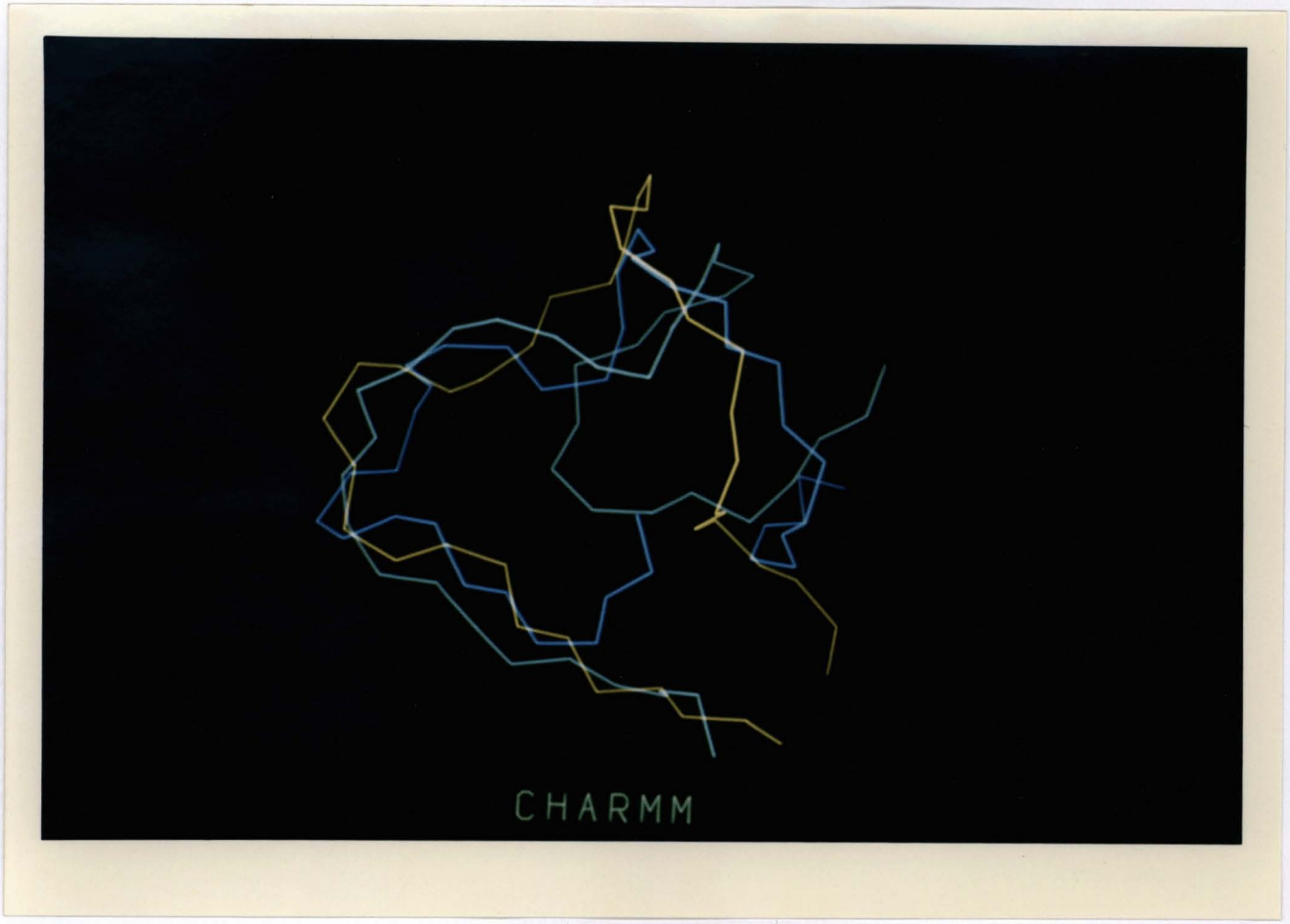


Fig. 2. Computer drawings of the backbone structures of conotoxin GI calculated by DISGEO, DADAS and CHARMM using experimentally obtained NMR data. They are superimposed to best fit in space. (a) Five DISGEO structures. (b) Five DADAS structures. (c) Three CHARMM structures.

Table 1

Statistics of the residual NOE constraints violations for the conotoxin GI conformations computed from the NMR data using DISGEO, DADAS and CHARMM

run	Sum of violations (Å)				
	DISGEO (C-DS)		DADAS (C-DD)		CHARMM (C-CM)
1	1.18	(1.69)	1.67	(3.64)	(2.60)
2	1.30	(2.86)	1.82	(2.08)	(3.12)
3	1.56	(2.21)	2.05	(3.38)	(5.46)
4	1.56	(2.34)	2.11	(1.69)	-
5	1.69	(1.69)	2.11	(2.73)	-

The numbers in parentheses indicate that violations for the equivalent protons are calculated using $(\langle r^{-6} \rangle)^{-1/6}$ mean distances.

They are denoted as C-DS-1, C-DS-2, ---,C-DD-1,---,C-CM-1,--, so on. These structures from DISGEO and DADAS were picked up from a hundred conformers. Thus CHARMM structures do not have qualities same to those of DISGEO and DADAS structures. Because of the long calculation time, only three structures of conotoxin GI were obtained from CHARMM. In CHARMM structures, violations for the prochiral protons were calculated using $(\langle r^{-6} \rangle)^{-1/6}$ mean distances. Although the pseudo atom treatment for the prochiral protons was used in both DISGEO and DADAS, violations for these structures were calculated using$(\langle r^{-6} \rangle)^{-1/6}$ mean distances same as CHARMM structure for comparison. Variations for the DISGEO and DADAS structures were not seriously large. This means that these structures satisfied the demand of the $(\langle r^{-6} \rangle)^{-1/6}$ mean distance treatment and that these treatments seemed to be a realistic approach for distances of the prochiral protons.

Table 1 shows that qualities of both DISGEO and DADAS structures with respect to consistency to the constraint is nearly the same. As the violation of the C-CM-3 structure is remarkably large, this structure should be trapped by a local minimum of the energy function. DADAS structures have slightly larger values of violation than DISGEO structures. Thus the DADAS structures should have a tendency to locate at some local minima which are in the vicinity of the global minimum. Although the violation values in the CHARMM structure are spread out, that of C-CM-1 structure is comparable with those of the other two method's structures. Thus all three methods have the ability to generate structure consistent with NMR data.

The r.m.s.d. values among different conformers of conotoxin

Table 2

R.m.s.d. values of the backbone atoms between different conformations of conotoxin GI (Å)

	1	*											
	2	1.3	*										
DISGEO	3	1.0	1.6	*									
	4	1.0	1.0	1.2	*								
	5	1.1	1.2	1.2	1.0	*							
	1	1.9	1.8	1.6	1.9	1.4	*						
	2	1.8	1.7	1.7	1.7	1.2	1.1	*					
DADAS	3	3.5	4.0	3.6	4.0	3.9	3.8	4.1	*				
	4	1.9	1.7	1.7	1.7	1.4	1.5	0.7	4.2	*			
	5	0.8	1.0	1.1	1.0	0.9	1.7	1.6	3.7	1.7	*		
	1	1.5	1.7	1.8	1.8	1.6	1.7	1.8	3.8	1.9	1.7	*	
CHARMM	2	2.7	2.6	2.6	2.4	2.5	2.7	3.1	4.6	3.0	2.8	2.6	*
	3	3.4	3.3	3.5	3.3	3.6	4.0	3.9	3.1	3.9	3.3	4.0	3.9	*

Table 3

Average r.m.s.d. (Å) values of conotoxin GI

	DISGEO	DADAS	CHARMM
DISGEO	1.16		
DADAS	1.96	2.41	
	(1.50)	(1.38)	
CHARMM	2.55	3.02	3.50
	(2.12)	(2.34)	(2.60)

The r.m.s.d. values were calculated including all the backbone atoms. Averages were taken over the 5 DISGEO or 5 DADAS or 3 CHARMM structures.

The numbers in parentheses indicate that averages were taken except for C-DD-3, C-CM-2 and C-CM-3 structures.

GI are listed in Table 2 and their averaged values are listed in Tables 3. Table 2 shows that the r.m.s.d. values of C-DD-3, C-CM-2 and C-CM-3 structures are remarkably large, although other r.m.s.d. values are always within 2.0$\mathring{A}$. C-CM-3 structure is a typical case of the structure which was trapped by a local minimum because both convergence and consistency were poor. This might be the consequence of the competition between pseudo energy of NOE and non-bonded energy. For the C-DD-3 and C-CM-2 structures, the values of violations are small. This suggests that these structures are candidate for the solution structure of conotoxin GI. We cannot determine which is more probable so far. The r.m.s.d. values among these structures except C-DD-3, C-CM-2 and C-CM-3 structures are always within 2.0$\mathring{A}$. This values are good enough to discuss the structure-activity relationship as shown in Chapter II.

Structure of Tendamistat

The residual NOE constraints violations of Tendamistat conformers calculated by DISGEO and DADAS are summed up respectively. Five of them were listed according to the values. These conformers are denoted as T-DS-1, T-DS-2, ---,T-DD-1,---,T-CM-1,--, so on, respectively. Table 4 shows that the distance constraints were almost fulfilled in both five structures by DISGEO and DADAS. As the size of molecule and the number of distance constraints are larger than those of conotoxin GI, the values of the violations in the calculated Tendamistat structures are larger than in the conotoxin GI structures. Although

Table 4

Statistics of the residual NOE constraints violations for the Tendamistat conformations computed from the simulated X-ray data using DISGEO and DADAS

run	Sum of violations (Å)	
	DISGEO (T-DS)	DADAS (T-DD)
1	4.00	12.58
2	4.29	12.95
3	6.73	14.43
4	6.96	16.28
5	7.33	25.90

Table 5

R.m.s.d. values of the backbone atoms between different conformations of Tendamistat and between different conformations and X-ray structure (Å)

		1-74	7-73								
X-ray											
	1	2.4	1.8	*							
	2	2.7	2.0	1.3	*						
DISGEO	3	2.7	2.0	1.2	1.3	*					
	4	2.4	1.9	1.2	1.4	1.4	*				
	5	2.5	1.9	1.2	1.1	1.4	1.4	*			
	1	3.2	2.0	3.0	3.4	3.2	3.2	3.2	*		
	2	2.4	1.7	1.9	2.3	2.2	2.2	2.2	2.9	*	
DADAS	3	2.8	2.0	3.1	3.4	3.2	3.2	3.4	2.5	3.0	*
	4	3.2	2.5	3.4	3.3	3.5	3.4	3.4	3.1	3.2	3.1 *
	5	4.0	3.8	3.9	3.8	3.7	3.9	3.8	4.6	4.3	3.6 3.8 *

violations in each structure by DISGEO and DADAS were within almost same values, the values in the DADAS structures were always greater than those in the DISGEO structures. This means that the DADAS structures have a larger tendency to be trapped by local minima than the DISGEO structures.

None of the structure from CHARMM comparable to those from DISGEO and DADAS were obtained. Main reason of this might come from the fact that during the CHARMM calculation it is hard to escape from local minima and to get a correct folding. The most successful CHARMM structure is shown in Fig. 3. Comparing to the X-ray structure of Tendamistat, it is clear that, even in this best structure, some of β-sheets strands are twisted. Thus the CHARMM structure of Tendamistat is not discussed anymore.

The extent to which the Tendamistat structures differs from each other, and how they relate to the X-ray structure were measured by the r.m.s.d. values which are listed in Tables 5 and 6. The five conformers of Tendamistat by DISGEO and DADAS, respectively were superimposed in Fig. 4 so as to get best-fit in space. In spite of the complexity of the folding, the global feature of the X-ray structure including the arrangement of the each strands of two β sheets was well reflected in both DISGEO and DADAS structures. Therefore the ability of DISGEO and DADAS to construct the correct structure should be almost same. Fig. 4 shows that both β-sheets in the structures from both programs are well defined. However, polypeptide segments of 1 to 6 (N-terminus,) 36 to 45 (around the junction between the β-strands III and IV) and 61 to 68 (around the junction between the β-strands V and VI) are deviated from X-ray structure. In these

Fig. 3. Computer drawing of the backbone structures of Tendamistat calculated by CHARMM using simulated NOE data from X-ray structure. The most successful CHARMM structure is shown.

Table 6

Average r.m.s.d. values (Å) of Tendamistat		
X-ray	- DISGEO	2.54
X-ray	- DADAS	3.12
DISGEO	- DISGEO	1.29
DADAS	- DADAS	2.97
DISGEO	- DADAS	3.17

The r.m.s.d. values were calculated including all the backbone atoms. Averages were taken over the 5 DISGEO or 5 DADAS structures.

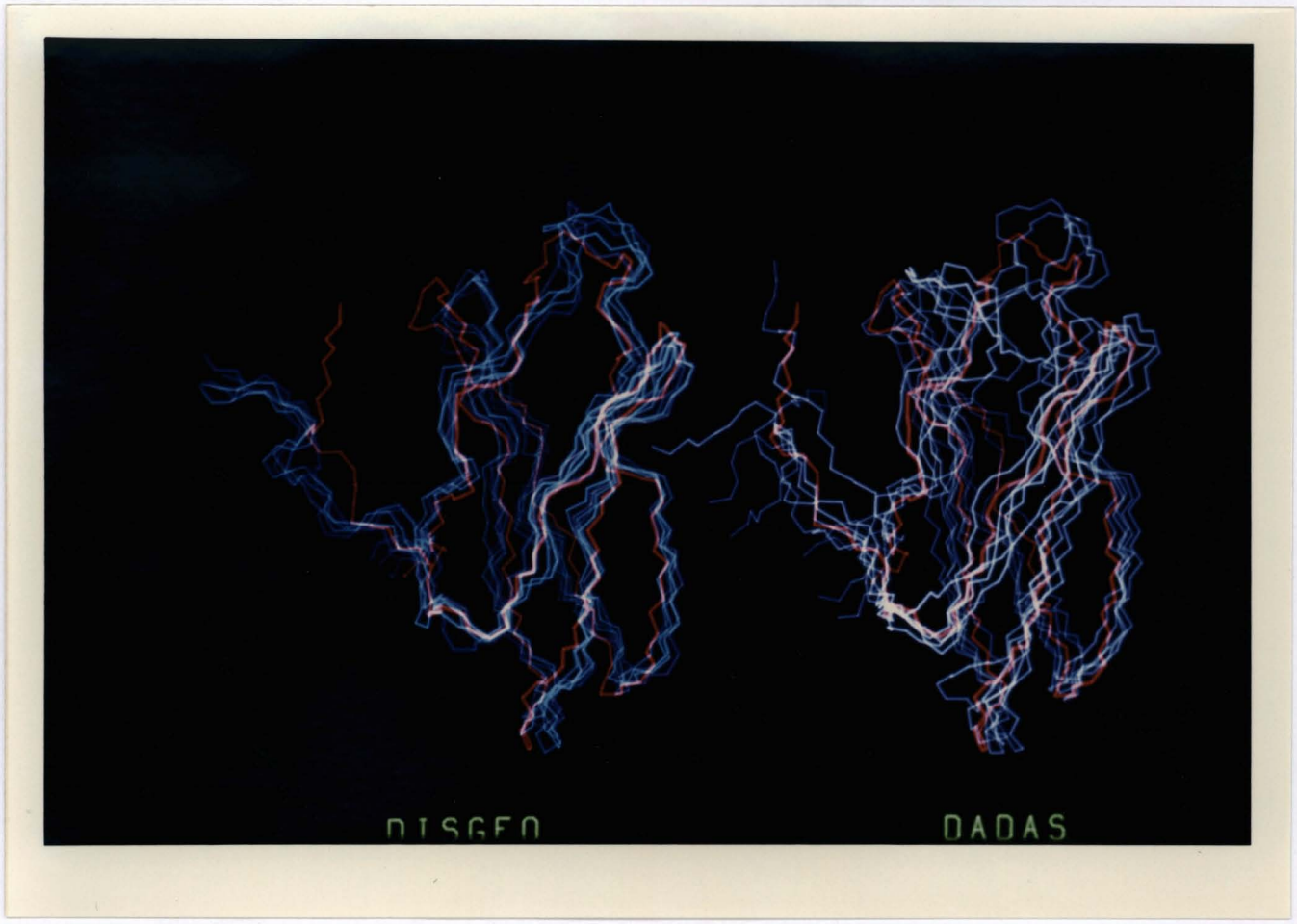


Fig. 4. Computer drawings of five DISGEO structures and five DADAS structures with the X-ray structure. They were calculated by DISGEO or DADAS using simulated NOE data from X-ray structure and were superimposed on the X-ray structure to best fit in space. The red line represents the X-ray structures.

polypeptide segments, the number of the distance constraints are rather small. Figs. 5 and 6 show some correlation between the distribution of distance constraints and the convergence of the structures. The close relationship between the quality of constraints and the convergence is clear in a comparison of these figures. Thus, both programs reflect well the quality of the distance constraints.

From the aspect of r.m.s.d.'s between any two conformers, DISGEO structures have smaller values than DADAS structures. This results in the better convergence of DISGEO structures than that of DADAS ones as shown in Fig. 4. However the degrees of the deviation from the X-ray structure measured by r.m.s.d. are almost same for both structures. These facts imply that the sampling space of the DISGEO structures was smaller than that of DADAS. DISGEO structures might lead to the misunderstanding of the some feature of solution structure. For example, N-terminus segments of different Tendamistat structures derived from DISGEO were well converged to a unique structure which has a quite different orientation from X-ray data. In contrast to this, these segments of DADAS structures were widely spread in space. It can easily concluded that this segment cannot be determined uniquely from the NMR data. It should be emphasized that the DISGEO structure dose not mean a wrong structure, because some of the DADAS structures also take the similar orientation in DISGEO structures.

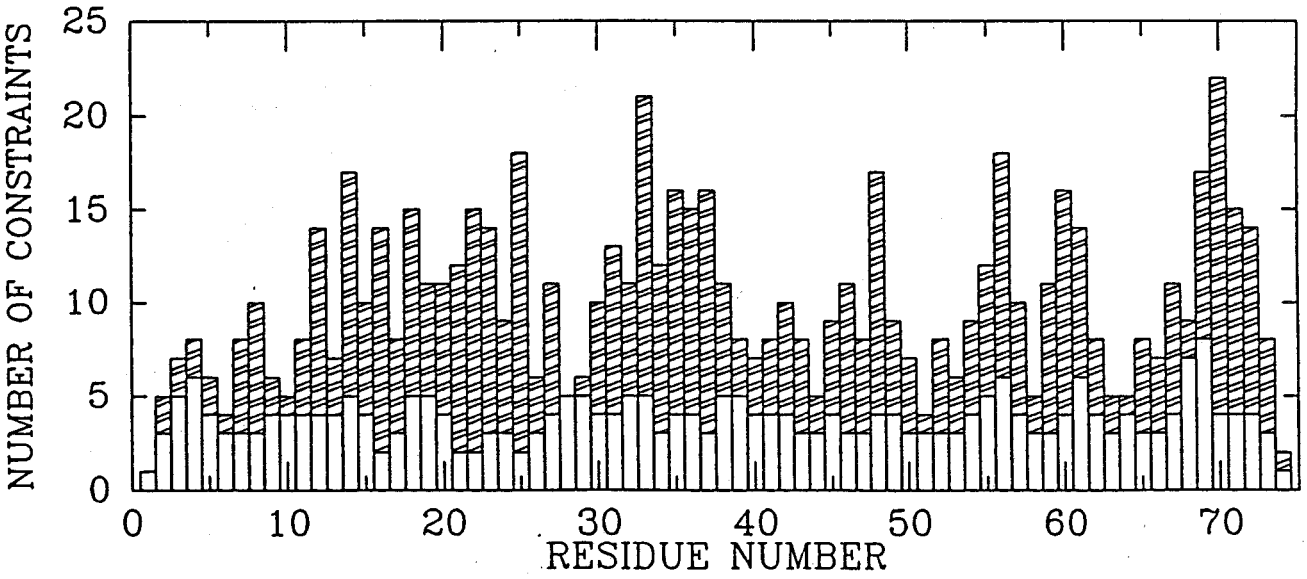


Fig. 5. Plot of the total number of distance constraints per residue versus the amino acid sequence of Tendamistat.

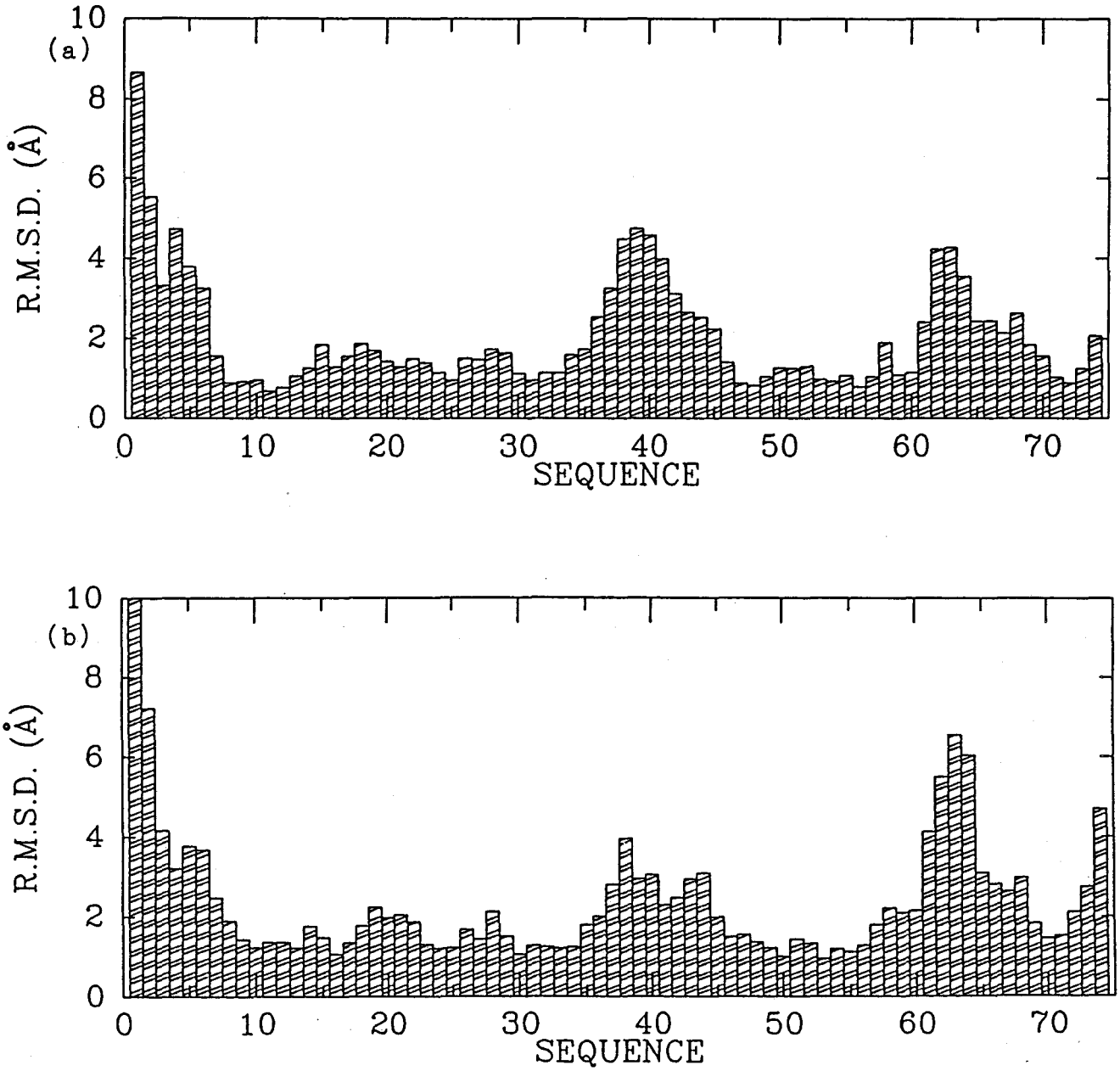


Fig. 6. plots of the average r.m.s.d. per residue calculated from pairwise global fits of the individual structures versus the amino acid sequence. The global fit was over all backbone atoms and r.m.s.d. per residue is also for these atoms. (a) 5 pairs between X-ray and T-DS-1 to 5 structures. (b) 5 pairs between X-ray and T-DD-1 to 5 structures.

[Discussion]

In this section, we discuss the merits and demerits of these computer programs; DISGEO, DADAS and CHARMM, judging from the aspect of the application to the structure determination by the NMR.

The first subject is so called local minimum. The ability of these programs largely depends on how they can overcome local minima to find the global minimum. This can be measured by the residual violations of imposed constraints in calculated structures. All three methods showed to have the ability to generate the structure consistent with NMR data in the case of conotoxin GI. DISGEO and DADAS also showed the ability in a complex molecule such as Tendamistat. The deviations of both structures by DISGEO and DADAS from the X-ray structure are about 2.5 $\mathring{A}$. As shown in cases of conotoxin GI and Tendamistat, almost all of the DISGEO structures have small values of residual violations. One of the merits of DISGEO is that it can be easily handled with suitable parameters. The other two programs are more sensitive to calculation parameters. However, there was no serious violation in DADAS conformers. The strategies of DADAS and CHARMM are the minimizations of a target function and of pseudo NOE potentials, respectively. In this sense both programs have a resemblance each other. The minimizer of DADAS is the conjugate gradient method and that of CHARMM is molecular dynamics. If the CHARMM structure is trapped by some local minima, it is hard to escape from them. The reason might come

from the fact that non-NOE terms in CHARMM obstruct the way to correct folding from local minima.

As discussed in the general introduction, the important requirement for these programs are unbiased search of the all allowed conformational space. One way to realize this is a introduction of the randomness into a starting condition for calculation. In DISGEO, randomly selected distances were used and in DADAS a randomly chosen initial structure was used for this purpose. Although the DISGEO structures have a small value of violation, the searching area of DISGEO is limited. This might mean that a kind of bias sampling in stead of completely random selection of a starting condition, is done at the initial stage. In this respect the DADAS achieves the most unbiased sampling of the conformational space. The CHARMM structures starting from randomly chosen structures are easily trapped in local minima at the early stage. It might be a demerit of CHARMM that the initial structure of the calculation is restricted in a narrow range. Another way for unbiased search might be possible in CHARMM that is tracing the trajectory of molecular dynamics through efficiently long period to travel all allowed conformational space. But it is very difficult because of the limit of CPU time of computer.

References

1. Wüthrich, K. (1986) "NMR of Proteins and Nucleic Acids", Wiley, New York.
2. Braun, W. (1987) Quart. Rev. Biophys., 19, 115-157.
3. Clore, G.M. & Gronenborn, A.M. (1987) Protein Engineering, 1, 275-288.
4. Noggle, J.H. & Schirmer, R.E. (1971) "The Nuclear Overhauser Effect", Academic Press, New York.
5. Braun, W., Bösch, C., Brown, L.R., Gō, N., & Wüthrich, K., (1981) Biochim. Biophys. Acta., 667, 377-396.
6. Havel, T.F. & Wüthrich, K. (1985) J. Mol. Biol., 182, 281-294.
7. Williamson, M.P., Havel, T.F. & Wüthrich, K. (1985) J. Mol. Biol., 182, 295-315
8. Havel, T.F. (1986) DISGEO, Quantum Chemistry Program Exchange, Program No. 507, Indiana University.
9. Braun, W. & Gō, N. (1985) J. Mol. Biol., 186, 611-626.
10. Kline, A.D., Braun, W. & Wüthrich, K. (1986) J. Mol. Biol., 189, 377-382.
11. Brünger, A.T., Clore, G.M., Gronenborn, A.M. & Karplus, M. (1986) Proc. Natl. Acad. Sci. USA., 83, 3801-3805.
12. Clore, G.M., Brünger, A.T., Karplus, M. & Gronenborn, A.M. (1986) J. Mol. Biol., 191, 523-551.
13. Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, A.T., Karplus, M. and Gronenborn, A.M. (1986) EMBO J., 5, 2729-2735.

14. Blumenthal, L.M. (1970) "Theory and Applications of Distance Geometry", Chelsea, New York.

15. Crippen, G.M. (1981) "Distance Geometry and Conformational Calculation", Research Studies Press, New York.

16. Brook, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) J. Comput. Chem., 4, 187-217.

17. van Gunsteren, W.F. & Berendsen, H.J.C. (1982) Biochem. Soc. Trans., 10, 301-305.

18. Clore, G.M., Gronenborn, A.M., Brünger, A.T. and Karplus, M. (1985) J. Mol. Biol., 186, 435-455.

19. Clore, G.M., Brünger, A.T., Karplus, M. & Gronenborn, A.M. (1987) FEBS Lett., 213, 269-277.

20. Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Gō, N. & Wüthrich, K. (1987) J. Mol. Biol., 196, 611-639.

21. Gray,W.R., Luque,A., Olivera,B.M., Barrett,J. and Cruz,L.J. (1981) J. Biol. Chem., 256, 4734-4740

22. Vertesy, L., Oeding, V., Bender, R., Zepf, K. & Nesemann, G. (1984) Eur. J. Biochem., 141, 505-512.

23. Nishiuchi,Y. and Sakakibara,S (1982) FEBS Lett., 148, 260-262

24. Pflugrath, J., Wiegand, E., Huber, R. & Vertesy, L. (1986) J. Mol. Biol., 189, 383-386.

25. Jeener, J., Meier, B., Bachman, P. & Ernst, R.R. (1979) J. Chem. Phys., 71, 4546-4553.

26. Momany, F.A., McGuire, R.F., Burgess, A.W. & Scheraga, H.A. (1975) J. Phys. Chem., 79, 2361-2381.

27. Nemethy, G., Pottle, M.S. & Scheraga, H.A. (1983) J. Phys. Chem., 87, 1883-1887.

28. McLachlan, A.D. (1979) J. Mol. Biol., 106, 983-994.

29. Jardetzkey, O. & Roberts, G.C.K. (1981) "NMR in Molecular Biology", Academic Press, New York.

30. Clore, G.M. & Gronenborn, A.M. (1985) J. Magn. Reson., 61, 158-164.

31. Braun, W., Wider, G., Lee, K.H. & Wüthrich, K (1983) J. Mol. Biol., 169, 921-948.

32. Abe, H., Braun. W., Noguti T. & Gō. N. (1984) Comp. Chem., 8, 239-247.

33. Noguti, T. & Gō, N. (1983) J. Phys. Soc. Jpn., 52, 3685-3690.

34. Ooi, T., Nishikawa, K., Oobatake, M. & Scheraga, H.A. (1978) Biochim. Biophys. Acta., 536, 390-405.

35. Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. & van Gunsteren, W.F. (1985) J. Mol. Biol., 182, 179-182.

36. Verlet, L. (1967) Phys. Rev., 159, 98-105.

37. Wüthrich, K., Billeter, M. & Braun, W. (1983) J. Mol. Biol., 169, 949-961.

38. Fletcher, R. (1980) "Practical Methods of Optimization : Unconstrained Optimization", Wiley, New York.

General Conclusion

In this thesis, the author developed a method to determine solution structures of protein and polypeptide with three dimensional atomic coordinates by using distance informations from NMR experiments. There had not been any direct method to do so at the beginning of the present study. Our strategy is the combined use of 1H-NMR and one of the distance geometry algorithms; DADAS. Semi-quantitative estimation of interatomic distances from NOE data and pseudo atom treatment of prochiral protons were introduced in order to treat the experimental data with DADAS. The tools of the quantitative evaluation of calculated structures were also developed and used to clarify their features.

Resulted structures with our method were not a unique but an assemble of somewhat different conformers. Although distance informations from NMR measurements were obtained only for a small subset of all possible proton pairs, the calculated structures were well converged. These results showed two things that (1) distance information obtained from currently available NMR techniques such as NOESY experiment is enough to delineate the molecular structure, and (2) our method has the capability to delineate the boundary of the conformational space defined by NMR data.

Limitation on the size of molecules to which our method can be applied is particularly interested, because proteins which play an important biological role in vivo have molecular weight

ranging widely up to 1000000. The limit of weight handled by NMR is said to be about 10000 so far. As the line-width of the individual NMR peaks is approximately proportional to molecular weight, decrement of the S/N ratio in the NMR spectrum caused by broadening of the peaks is a serious problem in large proteins. The overlapping of the peaks is also serious one because assignments of these peaks are very difficult. One answer for these problem is the increase of the magnetic field of NMR, because high magnetic field improves the S/N ratio and resolution. Another answer is the use of newly developed NMR techniques such as three-dimensional NMR. By these improvements, maximum molecular weight will be about 20000 in near future.

In this thesis information used to construct molecular structure is only the distance constraints from NOE data. But NMR is able to offer another structural information: J-coupling constants, which are related to values of dihedral angles. Incorporation of these informations into our method should promise a further improvement. Incorporation of stereospecific assignment improves the procedure as well because our treatment of prochiral protons loses some distance information.

Although our interest has been focused in the distance informations of macromolecules, NMR provides informations of molecular motion as well. Molecular dynamics calculation or Monte Carlo simulation of macromolecules using the structure which is determined with NMR data as above mentioned should give us a new aspect of molecular motion. The results of these calculations could be evaluated by a comparison to the experimental parameters such as T_1 and T_2.

List of Publications

The contents of this thesis have been published or will be published in the following papers.

Chapter 1

A Conformational Study of Polypeptides in Solution by ^{1}HNMR and Distance Geometry

Ohkubo, T., Kobayashi, Y., Shimonishi, Y., Kyogoku, Y., Braun, W. and Gō, N.

Biopolymers 25 (1986) S123-S134

Chapter 2

Conformational Analysis of Conotoxin and its Analogue by NMR Measurement and Distance Geometry Algorithm

Kobayashi, Y., Ohkubo, T., Kyogoku, Y., Nishiuchi, Y., Sakakibara, S., Braun, W. and Gō, N.

Peptides : Structure and Function, Proceedings of the 9th American Peptide Symposium (1985) 101-104

Solution Conformation of Conotoxin GI

Kobayashi, Y., Ohkubo, T., Kyogoku, Y., Nishiuchi, Y., Sakakibara, S. and Gō, N.

Peptides : Proceedings of the 19th European Peptide Symposium (1986) 331-334

Solution Conformation of Conotoxin GI by ^{1}H-Nuclear Magnetic

Resonance and Distance Geometry

Kobayashi, Y., Ohkubo, T., Kyogoku, Y., Nishiuchi, Y.,
Sakakibara, S., Braun, W. and Gō, N.

Protein Engineering, submitted

Chapter 3

A Comparison of the Polypeptide and Protein Structures Determined by Three Different Methods; DISGEO, DADAS and CHARMM

in preparation

List of the related papers

Conformational Analysis of Ancovenin, Angiotensin I Converting Enzyme Inhibitor, by NMR and Distance Geometry

Nishikawa, M., Teshima, T., Wakamiya, T., Shiba, T.,
Kobayashi, Y., Ohkubo, T., Kyogoku, Y. and Kido, Y.

Peptide Chemistry 1987 (1988) 71-74

The solution structure of α-Human Atrial Natriuretic Polypeptide(α-hANP)

Kobayashi, Y., Ohkubo, T., Kyogoku, Y., Koyama, S.,
Kobayashi, M. and Gō, N.

Peptide Chemistry 1987 (1988) 81-84

Two-Dimensional ^{1}H-NMR Investigation of the Tertiary Structure of -Conotoxin

Kobayashi, Y., Ohkubo, T., Nishimura, S., Kyogoku, Y., Shimada, K., Minobe, M., Nishiuchi, Y., Sakakibara, S. and Gō, N.
Peptide Chemistry 1987 (1988) 85-88

The Conformation of α-Human Atrial Natriuretic Polypeptide in Solution
Kobayashi, Y., Ohkubo, T., Kyogoku, Y., Koyama, S., Kobayashi, M. and Gō, N.
J. Biochem. (1988), 104, 322-325.

Acknowledgments

The present works described in this dissertation has been performed under the direction of Professor Y. Kyogoku, Division of Molecular Biophysics, Institute for Protein Research, Osaka University. The author would like to express his sincere gratitude to Professor Y. Kyogoku for his cordial guidance, discussion, and intimate encouragements throughout the course of his study. He also wishes his sincere thanks to Dr. Y. Kobayashi for his cordial guidance, continuing discussion, hearty instruction and warm encouragements.

During the course of the study the author has received a number of fruitful discussions, helpful comments and hearty encouragement from many persons. The author is deeply indebted to Professor N. Gō, Kyoto University, and Dr. W. Braun, Eidgenossische Technische Hochschule for their continuous encouragements and discussions of distance geometry. His thanks are also due to all the members of Division of Theoretical Physics, Kyushu University.

In Chapter I the author is deeply grateful to Professor Y. Shimonishi, Dr. H. Ikemura and Mr. Y. Hidaka, Institute for Protein Research, for kindly supplying the synthetic polypeptide; STh [6-19]. In Chapter II the author is also grateful to Drs. S. Sakakibara and Y. Nishiuchi, Peptide Institute, Protein Research Foundation, for kindly supplying the synthetic polypeptide; conotoxin GI. In Chapter III the author is deeply grateful to Mr. S. Nishimura, Institute for Protein Research, and Drs. H.

Nakamura and W. Yoshikawa and Mr. T. Nakai, Protein Engineering Research Institute, and Mr. J. Matsukuma, Kyoto University, for their useful suggestions and discussions.

The author wishes to express his cordial appreciation to Drs. H. Sugeta, Institute for Protein Research, and H. Akutsu, Yokohama National University. His thanks are also due to all the members of Division of Molecular Biophysics, Institute for Protein Research. Furthermore the author wishes to thank Miss T. Kondo who skillfully took on the long task of typing. The author's thanks are also due to all the members of Protein Engineering Research Institute for their encouragements. Finally the author thanks sincerely to his parents for their unfailing understanding and affectionate encouragements.

Tadayasu Ohkubo