Bull. Inst. Chem. Res., Kyoto Univ., Vol. 48, Nos. 2-3, 1970

Tertiary Structures of Proteins: Analysis of Conformations

Ken NISHIKAWA, Motohisa OOBATAKE and Tatsuo OOI*

Received May 16, 1970

Outlines of the tertiary structures of myoglobin and lysozyme determined by x-ray analysis are expressed simply by diagrams of a distance from the center of gravity to each residual C^{α} atom plotted against the residue number. These diagrams exhibit some characteristic features of these proteins, *e. g.* hydrophobic residues and α-helical regions are located at rather inner part of the molecules. The tertiary structure of a protein can also be computed by dihedral angles about N-C^{α} bonds and C^{α}-C′ bonds, φ and ψ. The calculated conformations using several sets for φ and ψ, including those given by Kendrew and by Phillips, show little similarity to the native ones, especially in their overall shapes. Refinement of the dihedral angles has been tried to reproduce the native conformations, and we have been able to obtain reasonable sets for φ and ψ; the calculated conformations are in good agreement with the native one, the agreement which is shown in the diagrams.

INTRODUCTION

During the last two decades, remarkable developements have been made on investigations for structures of proteins; the determination of amino acid sequences begun with Insulin has now been extended to a larger molecule of more than 200 residues,[1] the stable secondary structures of α-helix and β-structure which could account for x-ray diffraction patterns of fibrous proteins, have been found actually in the tertiary structure of protein molecules and the complete determination of steric arrangement of the atoms in some proteins, was successfully done on myoglobin, haemoglobin, lysozyme, RNase, and so on.[2] One of the most important features revealed so far on the structure of a protein, may be specific arrangement of constituent atoms, the arrangement which gives rise to appearance of corresponding specific function of the protein, *e. g.*, enzymic activity. There are experimental results which strongly suggest that a protein has one and only one structure specified for the protein, although a large number of other conformations are possible.

For the purpose to elucidate the formation of the stable structures of proteins and polypeptides, various experiments such as renaturation experiments have been done.[3,4] Apparently, it seems to be quite strange that a native protein molecule of a definite primary structure determined by genetic codes, chooses one specific conformation. The completely denatured molecule, presumably corresponded to the linear chain molecule just synthesized on ribosomes, has been shown to be folded up by itself into the native compact conformation under a physiological environment. Therefore, it is quite likely that the native con-

* 西川　建, 大島　玄久, 大井　龍夫: Laboratory of Physical Chemistry of Enzyme, Institute for Chemical Research, Kyoto University, Uji, Kyoto.

formation of a protein is thermodynamically stable,[5] thus forming a specific structure without aid of other materials. When this is true, the native conformation must be at a minimum of free energy.[6] This view of thermodynamically stable structure of a protein molecule has been believed to be true since complete chemical synthesis of a protein was in success to show enzymatic activity specific for the protein.[7,8]

In order to verify the above postulate theoretically, a high speed computer enables us to calculate coordinates of atoms in a protein, and interatomic energies, and therefore the conformation stability of the protein. For polypeptides, simpler model compounds of proteins, the method would be shown to be useful, since stability of α-helix and its helical sense for various homopolymers of L-amino acid have been explained in terms of total energy or energy minimum,[9] taking into account of non-bonded energy, tortional energy, electrostatic energy, and hydrogen bond energy. Such conformational analyses on the structures of biopolymers are now successfully being developed[10] and some trials have been made on protein molecules.[11] However, the stability of protein molecules is not clearly elucidated yet from the energetic point of view, due probably to difficulty of generating a native conformation by given rotational freedoms φ and ψ about bonds.

In this article, some analyses of the known tertiary structures of lysozyme and myoglobin and an approach to derivation of native conformations will be described.

Molecular structure

The tertiary structures or complete atomic coordinates of myoglobin[12] and lysozyme[13] were determined by x-ray analysis. Because experimental accuracy in determination of atomic coordinates would be of an order of 1 A or less,[11] the coordinates are not so reliable to calculate the total internal energy summing over all the atomic pairs in the molecule, although arrangements of the atoms, or the conformation of the molecule as a whole, is correct. The stereographic

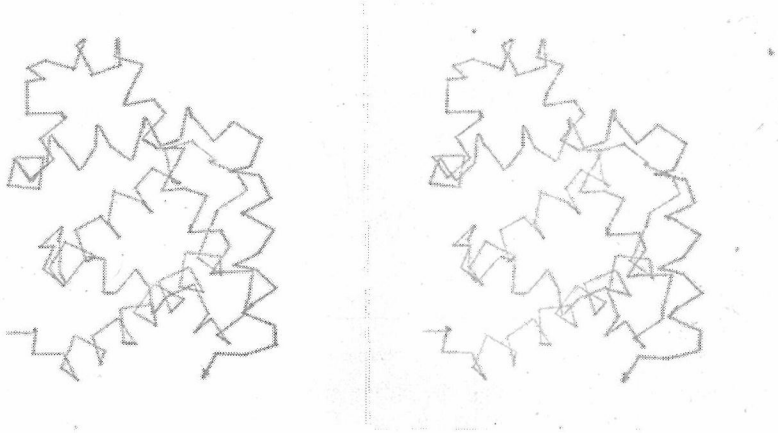

Fig. 1. The native structure of myoglobin, drawn by connecting successively the C^{α}-carbon to neighboring one. N-terminal end is designated as N. One can see the stereographic picture with the aid of a stereoscope.

formation of a protein is thermodynamically stable,[5] thus forming a specific structure without aid of other materials. When this is true, the native conformation must be at a minimum of free energy.[6] This view of thermodynamically stable structure of a protein molecule has been believed to be true since complete chemical synthesis of a protein was in success to show enzymatic activity specific for the protein.[7,8]

In order to verify the above postulate theoretically, a high speed computer enables us to calculate coordinates of atoms in a protein, and interatomic energies, and therefore the conformation stability of the protein. For polypeptides, simpler model compounds of proteins, the method would be shown to be useful, since stability of α-helix and its helical sense for various homopolymers of L-amino acid have been explained in terms of total energy or energy minimum,[9] taking into account of non-bonded energy, tortional energy, electrostatic energy, and hydrogen bond energy. Such conformational analyses on the structures of biopolymers are now successfully being developed[10] and some trials have been made on protein molecules.[11] However, the stability of protein molecules is not clearly elucidated yet from the energetic point of view, due probably to difficulty of generating a native conformation by given rotational freedoms φ and ψ about bonds.

In this article, some analyses of the known tertiary structures of lysozyme and myoglobin and an approach to derivation of native conformations will be described.

Molecular structure

The tertiary structures or complete atomic coordinates of myoglobin[12] and lysozyme[13] were determined by x-ray analysis. Because experimental accuracy in determination of atomic coordinates would be of an order of 1 A or less,[11] the coordinates are not so reliable to calculate the total internal energy summing over all the atomic pairs in the molecule, although arrangements of the atoms, or the conformation of the molecule as a whole, is correct. The stereographic

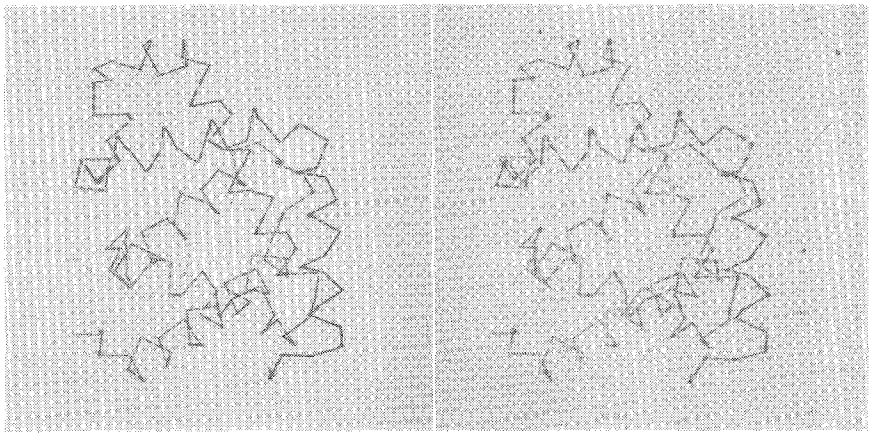

Fig. 1. The native structure of myoglobin, drawn by connecting successively the C^{α}-carbon to neighboring one. N-terminal end is designated as N. One can see the stereographic picture with the aid of a stereoscope.

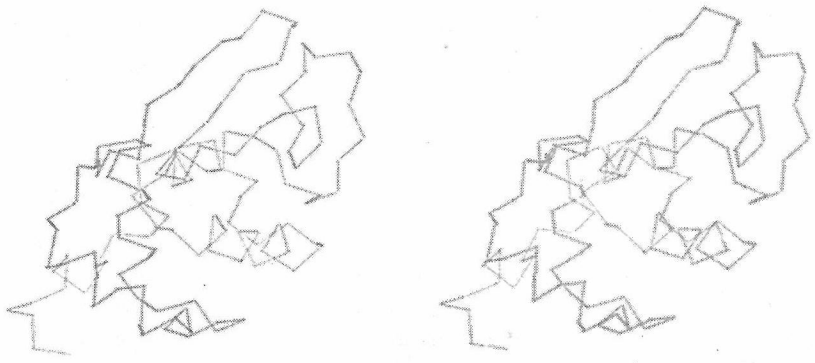

Fig. 2. The native structure of lysozyme, drawn as Fig. 1.

views of both proteins are shown in Figs. 1 and 2, where the peptide main chain is represented by the successive lines connected C^{α} atom to neighboring C^{α} atom. The molecules are nicely folded in compact shape. As can be seen in the figures, myoglobin contains high portion (about 75 %) of α-helix[14] in it, whereas lysozyme has nearly half as much as helical content (30 %)[13] in addition to antiparallel β-structure, and non-helical parts of the peptides make the conformation globular by delicate orientation of each peptide. In the main chains, the peptide unit, -CONH-, has approximately planar structure. However, angles about C^{α} carbons are not always tetrahedral.[10]

The globular shape of the molecules may be expressed by comparing a distance of each C^{α} carbon from the center of gravity. In Figs. 3 and 4, those distances are plotted against the residue number for myoglobin and lysozyme, respectively. Roughly speaking, every point is not so apart from the mean radius

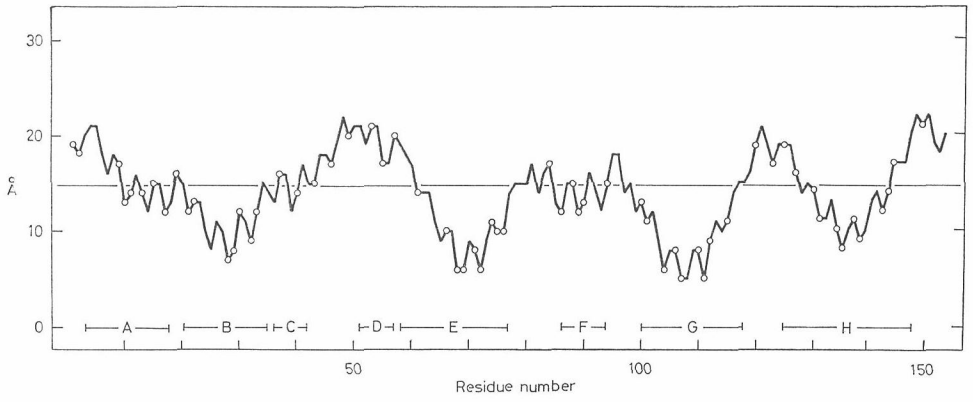

Fig. 3. The diagram representing the native structure of myoglobin, in which the distance from the center of gravity to each C^{α} atom is plotted against the residue number, and the root-mean square distance averaged over all residues is shown by the thin line. Open circles indicate hydrophobic residues and α-helical regions are shown with the letters A, B, C *etc.*, following Kendrew's notation.

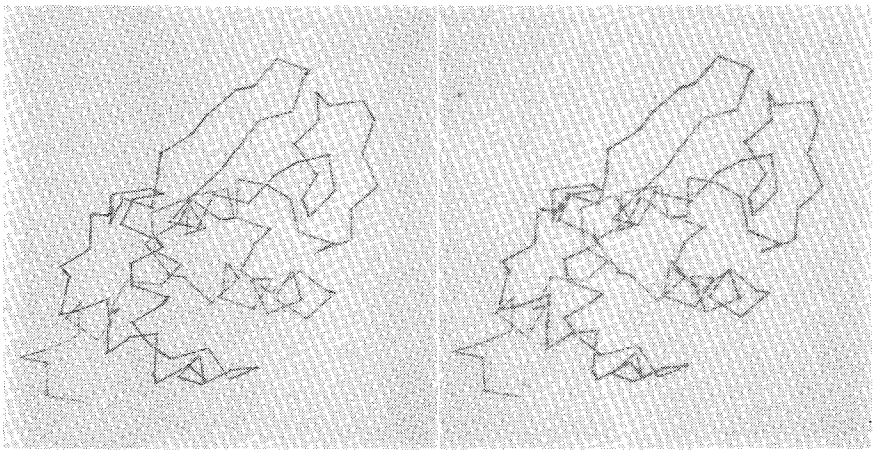

Fig. 2. The native structure of lysozyme, drawn as Fig. 1.

views of both proteins are shown in Figs. 1 and 2, where the peptide main chain is represented by the successive lines connected C^{α} atom to neighboring C^{α} atom. The molecules are nicely folded in compact shape. As can be seen in the figures, myoglobin contains high portion (about 75 %) of α-helix[14] in it, whereas lysozyme has nearly half as much as helical content (30 %)[13] in addition to antiparallel β-structure, and non-helical parts of the peptides make the conformation globular by delicate orientation of each peptide. In the main chains, the peptide unit, -CONH-, has approximately planar structure. However, angles about C^{α} carbons are not always tetrahedral.[10]

The globular shape of the molecules may be expressed by comparing a distance of each C^{α} carbon from the center of gravity. In Figs. 3 and 4, those distances are plotted against the residue number for myoglobin and lysozyme, respectively. Roughly speaking, every point is not so apart from the mean radius

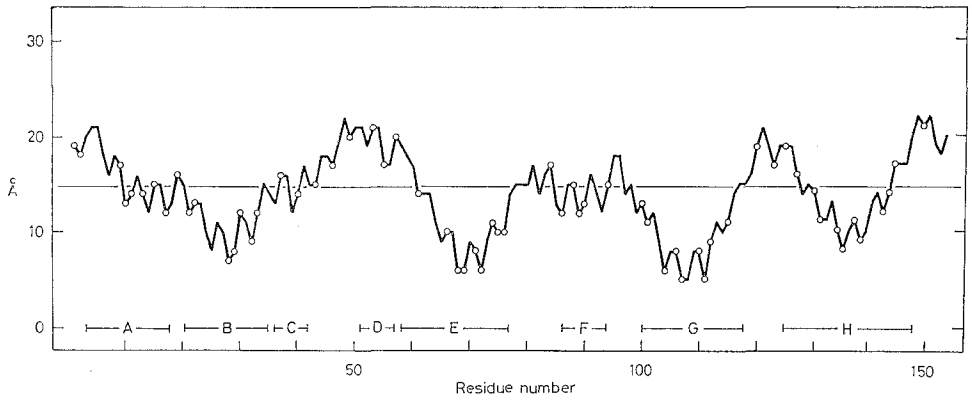

Fig. 3. The diagram representing the native structure of myoglobin, in which the distance from the center of gravity to each C^{α} atom is plotted against the residue number, and the root-mean square distance averaged over all residues is shown by the thin line. Open circles indicate hydrophobic residues and α-helical regions are shown with the letters A, B, C *etc.*, following Kendrew's notation.

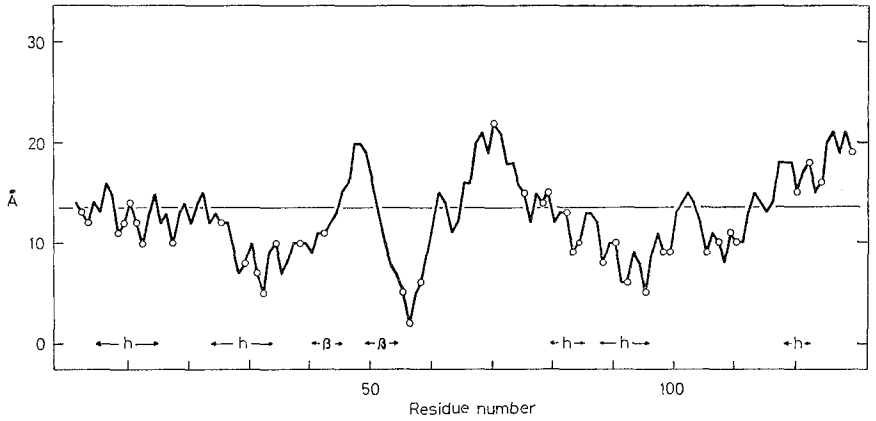

Fig. 4. The diagram for the native structure of lysozyme. The symbols, h and β, stand for α-helix and β-structure, respectively. The thin line represents the root-mean square distance and open circles indicate hydrophobic residues.

of gyration, 15.13 Å for myoglobin and 13.73 Å for lysozyme; in other words, some residues may be buried inside of the molecule and others exposed to the surface. Therefore, the residues having shorter distances than the mean radius are regarded as located at inner part of the molecule. It is of interest to see whether hydrophobic groups may exist inside or not, because hydrophobic residues generally constitute the inner core of the protein.[14] Circles in the figures represent hydrophobic residues in the molecules, most of which have shorter distances than the mean radius. To examine more quantitatively, the mean radius of gyration is calculated for hydrophobic groups separately. The result is that the mean radius for hydrophobic residues are 14.44 Å for 63 groups of myoglobin, and 11.81 Å for 39 groups of lysozyme, respectively. The corresponding values for hydrophilic 90 groups for both proteins are 15.52 Å and 14.32 Å. This result listed in Table 1 shows apparently the afore-mentioned nature. Interestingly, helical parts of the molecules are found in rather inner region except one short helix D of myoglobin. Incidentally, this helix region is deleted in α-chain of haemoglobin, which has a similar structure to myoglobin.[15]

Table 1. Rootmean Square Radius of Gyration.

	Myoglobin	Lysozyme
$\langle r_g \rangle$	15.13 Å (153)	13.73 Å (129)
$\langle r_g \rangle_{h\phi}$	14.44 Å (63)	11.81 Å (39)
$\langle r_g \rangle_h$	15.52 Å (90)	14.32 Å (60)

There is some difference in detailed patterns shown in the figures for myoglobin and lysozyme; *i. e.*, deviations from the mean value are much larger for lysozyme and especially leucine 56 located very close to the center of gravity seems to be deeply buried. On the other hand, there is no such a minimum for myoglobin but it has a periodic structure. The difference may be due to that in helical content, and to that in its biological function, the former has a capacity

to bind oxygen and the latter acts as an enzyme. It is, in addition, of interest to mention that locations of catalitic sites, glu 35 and asp 52 are interier of the lysozyme molecule.[13] The difference is also clearly seen by the deviation of $<r_g>_{h\phi}$ from $<r_g>$ or 0.7 Å for myoglobin and 1.9 Å for lysozyme.

Calculation of protein conformation

Once a sequence of a protein is known, the chemical structure can be determined completely. Therefore, a certain conformation of the protein may be calculated as a function of freedoms in the molecule. When rotational freedom about a bond, bending of bonds, and stretching of a bond are taken into account, the complete description of conformational analysis may be made.[11] However, there are too many variables to deal with the problem, so that it may be better to impose some reasonable restrictions. First of all, all the bond lengths are fixed according to crystal data for small molecules as listed in Table 2. Second, all the bond angles are chosen to fit the average ones of the crystal data for small molecules which have similar chemical structures as listed in Table 2 also. The geometry of a peptide is taken from the one obtained by the crystal structure as usually employed in the calculation, *i. e.*, transplanar structure of frozen rotation about C′-N bond.[16] Therefore, only freedoms left are rotation about N-C$^{\alpha}$, and C$^{\alpha}$-C′ bond, φ and ψ, respectively. Under such restrictions, every conformation can be computed as a function of all φ_i and ψ_i, which reproduce the conformation. When the set of φ and ψ could be found, the conformational analysis of a protein molecule would be hopefull, because we will be able to examine whether the conformation corresponds to energy minimum or not by calculating energy as a function of φ and ψ. If we fail to find any appropriate set of φ and ψ, some loosening of the restrictions imposed must be taken into consideration in next step.

Table. 2. Equilibrium Values for Bond Lengths and Bond Angles.

Bond lengths (Å)		Bond angles	
C$^{\alpha}$−C′	1.53	τ (C$^{\alpha}$C′N)	114°
C′=O	1.24	τ (NC′O)	125°
C$^{\alpha}$−N	1.47	τ (OC′C$^{\alpha}$)	121°
C′−N	1.32	τ (C′NC$^{\alpha}$)	123°
N−H	1.00	τ (C$^{\alpha}$NH)	114°
C−H	1.00	τ (HNC′)	123°
		τ (NC$^{\alpha}$C′)	112.5°

To begin with, several sets of φ and ψ are available for calculation; that is, the values for myoglobin given by Kendrew,[17] those for lysozyme by Phillips,[13] both being approximate values with uncertainty of about 10°.[10] These values are thought to be given by measuring each angle by model building.[13] Other sets of dihedral angles could be computed from atomic coordinates of C′$_{i-1}$, N$_i$, C$_i^{\alpha}$, C$_i$′, and N$_{i+1}$. Atomic coordinates corresponding to the conformation determined by the above angles could be obtained according to computation of poly-

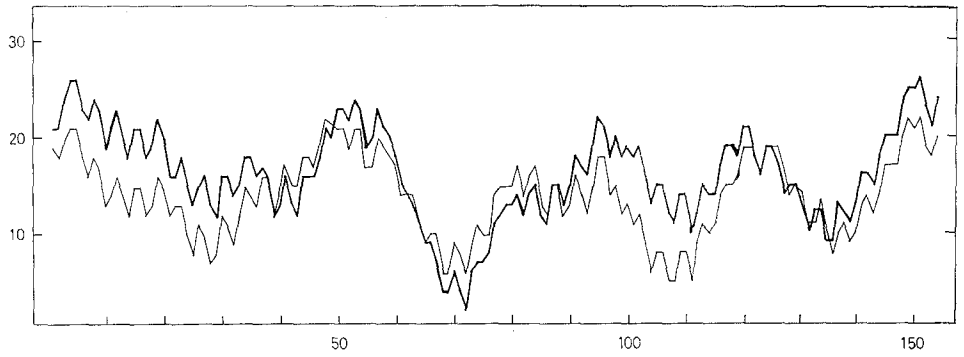

Fig. 5. The diagram for the conformation of myoglobin calculated from the dihedral angles of Kendrew is shown by the thick line, as well as its native structure, the thin line, for comparison.

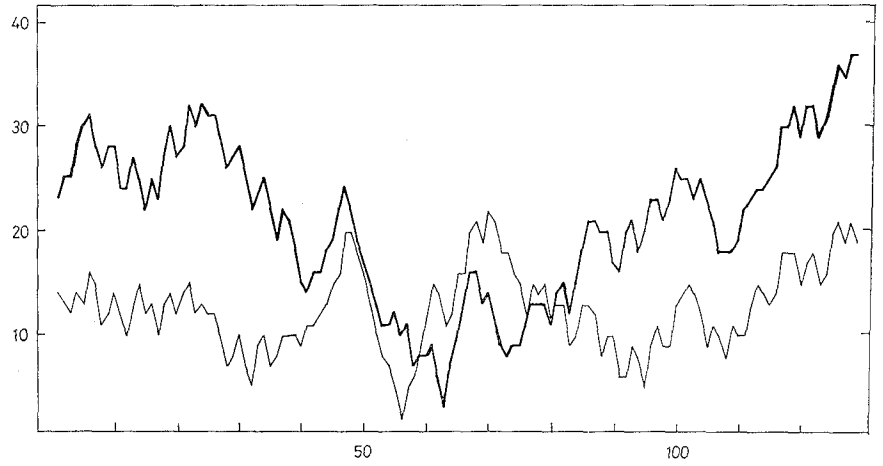

Fig. 6. The diagram for the conformation of lysozyme calculated from the values of Phillips is shown by the thick line, comparing the native one (the thin line).

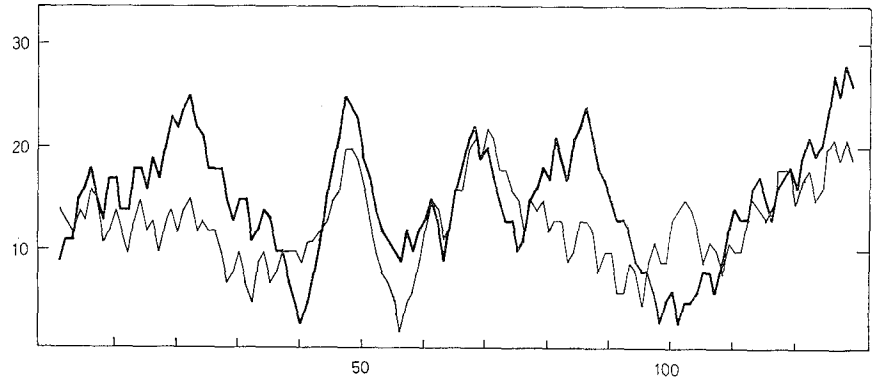

Fig. 7. The diagram for the conformation of lysozyme calculated from the computed dihedral angles (See the text), comparing the native one (the thin line).

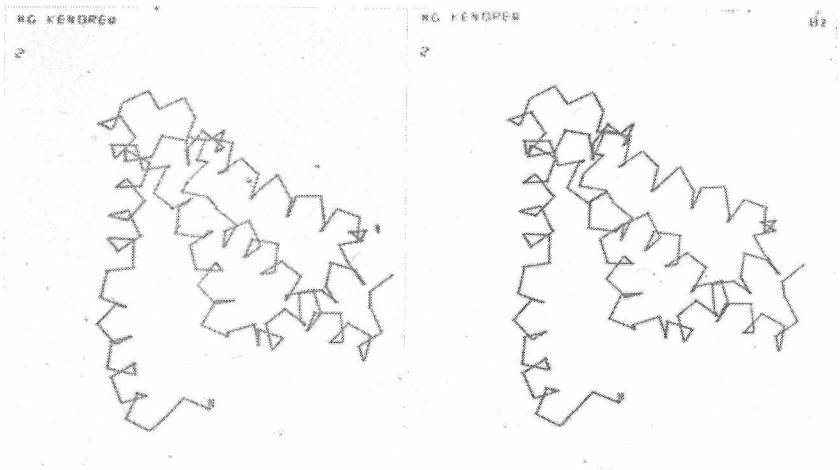

Fig. 8. Stereographic view of the myoglobin structure calculated from the dihedral angles of Kendrew.

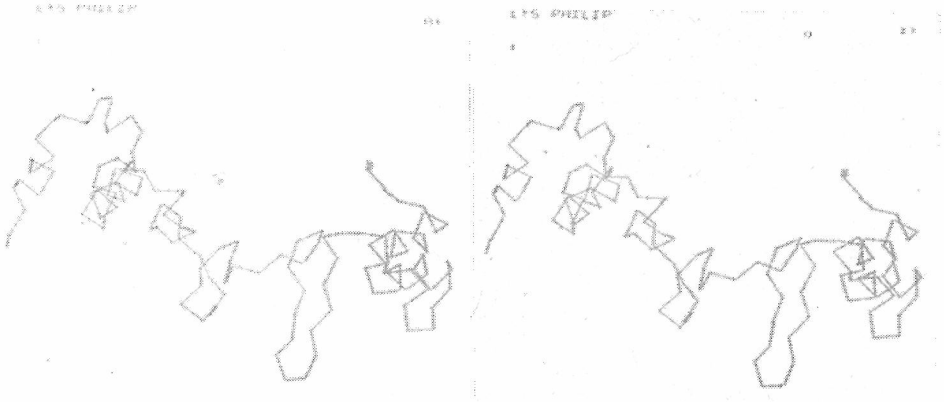

Fig. 9. Stereographic view of the lysozyme structure calculated from the dihedral angles of Phillips.

peptide chains. The calculated distance of each C^{α} carbon from the center of gravity for myoglobin by the use of the values of Kendrew is shown in Fig. 5. Apparently, the shapes of the molecule seems to be quite different from the native one. The similar figures for lysozyme by applying the values of Phillips and the computed angles are shown in Figs. 6 and 7, respectively. The conformation computed by the use of the dihedral angles given by Phillips seems to be extended form, because both ends are far apart from the mean radius of gyration. The one obtained by the computed angles seems to be folded better. The stereographic views of the conformations for the proteins are shown in Figs. 8 and 9. Even though the global conformations computed by the dihedral angles do not resemble the native ones, conformations of short segments are reproduced by computation, *e. g.*, α-helices and β-structure as seen in the figures. The principal origin of the discrepancy is due to the variey of the angle $NC^{\alpha}C'$ along the main chain. Despite of the use in the calculation of the mean value for this angle, 112.5°, for both proteins, the crystal angle show maximum deviation of

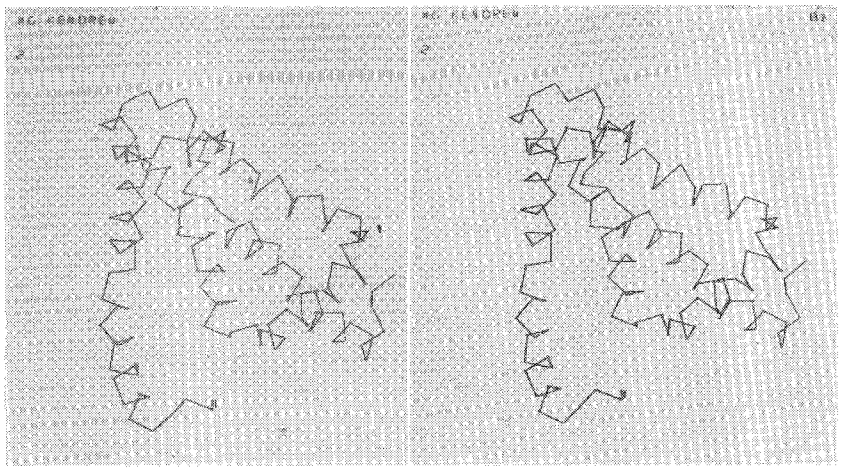

Fig. 8. Stereographic view of the myoglobin structure calculated from the dihedral angles of Kendrew.

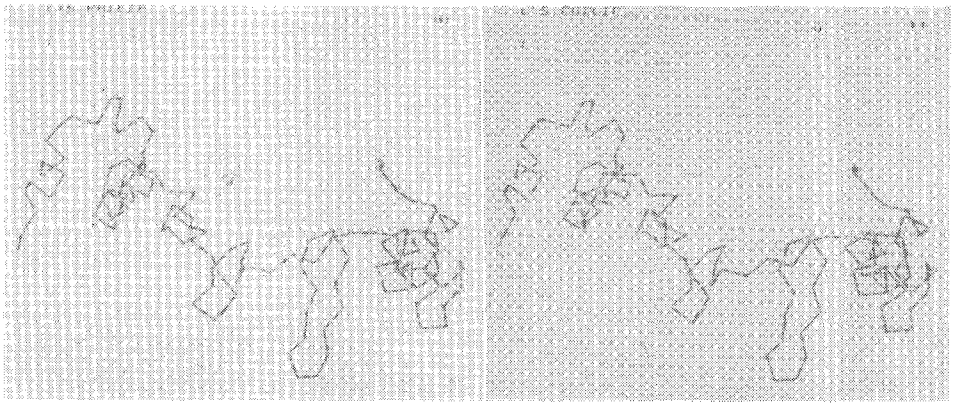

Fig. 9. Stereographic view of the lysozyme structure calculated from the dihedral angles of Phillips.

peptide chains. The calculated distance of each C^{α} carbon from the center of gravity for myoglobin by the use of the values of Kendrew is shown in Fig. 5. Apparently, the shapes of the molecule seems to be quite different from the native one. The similar figures for lysozyme by applying the values of Phillips and the computed angles are shown in Figs. 6 and 7, respectively. The conformation computed by the use of the dihedral angles given by Phillips seems to be extended form, because both ends are far apart from the mean radius of gyration. The one obtained by the computed angles seems to be folded better. The stereographic views of the conformations for the proteins are shown in Figs. 8 and 9. Even though the global conformations computed by the dihedral angles do not resemble the native ones, conformations of short segments are reproduced by computation, *e. g.*, α-helices and β-structure as seen in the figures. The principal origin of the discrepancy is due to the variey of the angle $NC^{\alpha}C'$ along the main chain. Despite of the use in the calculation of the mean value for this angle, 112.5°, for both proteins, the crystal angle show maximum deviation of

20° from this value. When all the angles of $NC^{\alpha}C'$ derived from crystal data and the computed φ and ψ are used in the calculation, for lysozyme, approximate conformation can be reproduced. However, we could not obtain correct conformation for myoglobin even by putting the angles of crystal structure into computation. This discrepancy is due to twists of peptide planes. Although we can improve the conformations by putting more variables in a calculation, the calculation seems to be less meaningful, since there are sever strains in the crystal structure which may be absent in the possible native conformations.

Staring from these values, we have tried to obtain a reasonable set of φ and ψ for native conformation. Details of procedures and results will be described elsewhere. The computations done so far show that a reasonable set of dihedral angles with the fixed $NC^{\alpha}C'$ angle of 112.5° does exist for myoglobin and for lysozyme. The generation of the native conformation is shown in Figs. 10 and 11, where the distances of C^{α} carbon from center of gravity coincide with the one for the native conformation although slight differences are present. Since we have not imposed any stress for bending of bonds, and stretching of bonds,

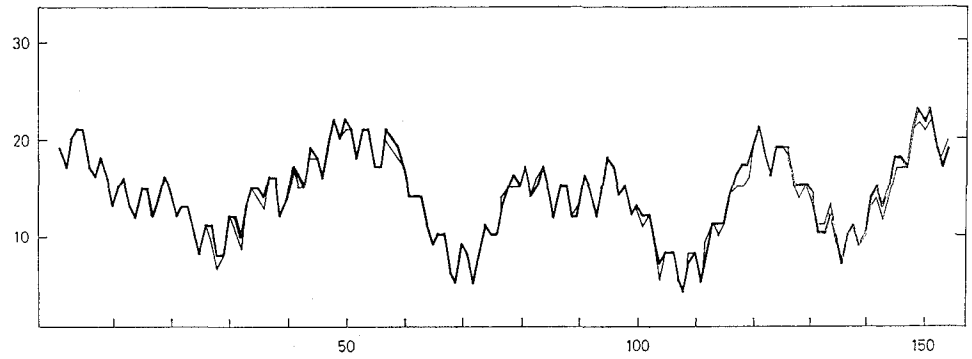

Fig. 10. The diagram for the conformation of myoglobin calculated from the refined dihedral angles (See the text). The thin line is the same one in Fig. 3, for comparison.

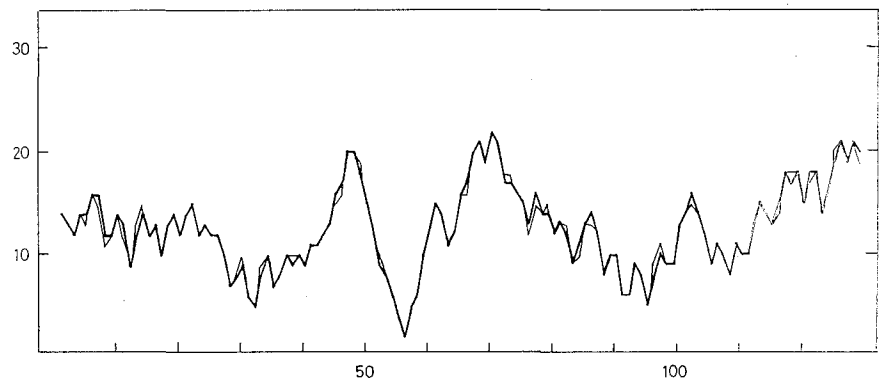

Fig. 11. The diagram for the conformation of lysozyme calculated from the refined dihedral angles (See the text). The thin line is the same one in Fig. 4, for comparison,

the energy of the structure may be calculated only by mutual interatomic energies, such as non-bonded energy, hydrogen bond energy *etc.*. Preliminary energy calculation showed, however, a few steric hindrances occur for the conformation computed by the use of refined dihedral angles, so that further studies to remove the hindrances are necessary. The study along this line is now in progress.

ACKNOWLEDGMENTS

The authors are grateful to express their sincere thanks to Dr. J. C. Kendrew and to Dr. D. C. Phillips for sending us the atomic coordinates of the proteins, which enable us to perform the present study. This work supported partially by the research grant from the Institute of Fundamental Physics, Kyoto University.

REFERENCES

(1) "Atlas of Protein Sequence and Structure" Vol. 4 (1969).
(2) M. F. Perutz, *Europ. J. Biochem.*, 8, 455 (1969).
(3) C. B. Anfinsen and E. Haber, *J. Biol. Chem.*, **236**, 1361 (1961).
(4) T. Isemura, T. Takagi, Y. Maeda and K. Yutani, *J. Biochem.*, **53**, 155 (1963).
(5) C. B. Anfinsen, *Brookhaven Symposia in Biology*, **15**, 184 (1962).
(6) K. D. Gibson and H. A. Scheraga, *Physiol. Chem. and Phys.*, **1**, 109 (1969).
(7) B. Gutte and R. B. Merridield, *J. Am. Chem. Soc.*, **91**, 501 (1969).
(8) R. G. Denkewalter, R. Hirshmann *et al*, *J. Am. Chem. Soc.*, **91**, 502, 503, 505, 506, 507 (1969).
(9) T. Ooi, R. A. Scott, G. Vanderkoii and H. A. Scheraga, *J. Chem. Phys.*, **46**, 4410 (1967).
(10) G. N. Ramachandran and V. Sasisekharan, Adv. in Port. Chem., **23**, 283 (1968).
(11) M. Levitt and S. Lifson, *J. Mol. Biol.*, **46**, 269 (1969).
(12) C. L. Nobbs, H. C. Watson and J. C. Kendrew, *Nature*, **209**, 339 (1966).
(13) C. C. F. Blake, G. A. Mair, A. C. T. North, D. C. Phillips and V. R. Sarma, *Proc. Roy. Soc., Lond.*, **B 167** 365 (1967).
C. C. F. Blake, L. N. Johnson, G. A. Mair, A. C. T. North, D. C. Phillips and V. R. Sarma, *Proc. Roy. Soc., Lond.*, **B 167**, 378 (1967).
(14) J. C. Kendrew, *Brookhaven Symposia in Biology*, **15**, 216 (1962).
(15) M. F. Perutz, H. Muirhead, J. M. Cox and L. C. G. Goaman, *Nature*, **219**, 131 (1968).
(16) L. Pauling and R. B. Corey, *Proc. Natl. Acad. Sci. U, S.*, **37**, 241 (1951).
(17) Personal communication.