METHOD OF PACKING ANALYSIS FOR THE SOLUTION OF THE CRYSTAL STRUCTURE OF CYCLO (GLY-L-TYR-GLY)$_2$. 2H$_2$O*

N. SHAMALA

(Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012)

Received on May 4, 1977

ABSTRACT

*This paper describes the development of a new technique for the solution of crystal structures from packing considerations, using contact distance criteria, knowing approximately the shape of the molecule. The method is applied to a complicated structure, namely, that of a cyclic hexapeptide, cyclo (GLy–L–Tyr–GLy)$_2$. 2H$_2$O, which crystallizes in the trichnic space group P*1. *Although this is a difficult space group for solution by conventional method, it turns out to be particularly facile for the new method. The trial structure thus obtained could be refined by a series of successive SFLS refinement, increasing the number of reflections used for the refinement at each stage.*

Key Words: Molecular packing, contact criteria, cyclic peptide, least-squares refinement.

INTRODUCTION

It has been well recognized that the three dimensional structure of biological macro-molecules plays a crucial role in their biochemical activity. Undoubtedly x-ray crystallography is the most potent method to give the complete three dimensional structure of these molecules in the solid state. However, in recent years, with the advent of fast and big computers, the solution of crystal structures without x-ray intensity data has become an interesting challenge. Towards this end, methods have been reported in the literature concerned with the determination of energetically favoured structures of biologically important molecules by computer calculations. Typically, in these calculations, energies of interactions of atoms influenced by the conformational changes of the molecule (non-bonded interactions, torsional energies, etc.) are considered and structures are sought that would optimize the total energy of these interactions in a crystal. By using the semi-emperical potential functions describing Van der Waals and electrostatic interactions between non-bonded atoms and ions and the formation

of hydrogen bonds, some successful attempts to solve the phase problem in crystals of unknown structures have been made by Williams [1], Kitaigorodsky and Mirasakaya [2], and Coiro, Giacamello and Giglio [3].

Extensive work has been carried out in this laboratory to predict the conformation of isolated biological macromolecules by theoretical calculations using potential functions. As early as 1969, Ramachandran [4] pointed out the interrelation between the theory of biopolymer conformation and the stability of a crystal structure and that it is not difficult to use the potential functions which have been successfully applied in this laboratory to predict the protein conformation, for the purpose of studying crystal structures. In fact such calculations have been done here for two simple examples of benzene and sulfurdioxide [5]. The results show that the potential functions adopted for biopolymer conformations in our laboratory can predict the crystal structures with reasonable accuracy. However, since the energy of a macromolecule in its crystal lattice will be influenced by the interactions of the molecule with the neighbouring molecules, such a calculation will present seemingly unsuromountable computational problems because of the large size of these molecules. Hence an alternative idea of using the contact distance criteria of Ramachandran, Ramakrishnan and Sasisekharan [6] for predicting the best packing of the molecules in their crystal lattice was employed. This technique of using contact criteria has been well developed in our laboratory [7] and particular reference may be made to the recent analysis of poly (Gly-Pro-HyP) [8]. Hence an attempt was made to employ this method of packing the molecules to solve the crystal structure of a hexapeptide cyclo $(Gly\text{-}L\text{-}Tyr\text{-}Gly)_2 \cdot 2H_2O$.

In this paper we shall discuss the method adopted for fixing the molecules and analysing their contacts, leading up to the revelation of a satisfactory packing arrangement which is probably accurate to about 1Å in the location of the outermost atoms in the molecule. The advantage of using low resolution data at the initial stages and how these could be improved as the refinement continued are also pointed out. A brief analysis of the power and the limits of the method on the basis of our experience is dealt with. A preliminary report of this work has appeared recently [9].

METHOD

In this section we describe the principle of the method and how various steps are programmed for a computer.

Principle of the Method

It is well known that the normal interatomic contact distances in a crystal are always larger than certain minimum values and that information regarding the magnitudes of these minimum values can be used to predict the conformation of various biomolecules. Such an analysis of the minimum contact distance criteria was made by Ramachandran and coworkers [10] for predicting the allowed and disallowed ranges of dihedral angles (ϕ, ψ) for a pair of peptide units. It is reasonable to assume that the criteria thus worked for a dipeptide unit can work equally well for the packing of larger peptides in their crystal structures.

If the stereochemistry of the molecule is known, or can be postulated from known bond lengths, bond angles and dihedral angles, its crystal structure can be defined by six parameters: three rotational parameters, which define the orientation of the molecule and three translational parameters to represent its position within the unit cell. The rotational parameters can be conveniently defined by means of the three Eulerian angles ϕ, θ, ψ. The translational parameters are taken to be the fractional coordinates X, Y, Z of the origin used for defining the rotations. By stepwise variation of these six parameters and by applying the symmetry operations of the space group on the rotated and translated unit, we can derive a set of trial structures. Not all of them can exist in reality due to steric overlaps between some of the constituent atoms in neighbouring molecules. An acceptable solution is one that involves no predominant short-contacts between the neighbouring molecules and for which the condition for the formation of hydrogen bonds, if any, are also satisfied.

General Considerations

The prerequisites for applying this method are (*a*) The geometry of the asymmetric unit. The knowledge of molecular conformation is often necessary. In fact, bond lengths and angles are available in the literature with accuracy adequate for this purpose. The conformation of a molecule is established either from NMR study or intramolecular potential energy calculations. (*b*) The unit cell parameters and the space group. This is a matter of standard routine and is easily determined from preliminary studies of x-ray diffraction.

Calculational Procedure Employed

The details of the procedure followed in carrying out the packing analysis is explained in this section by describing a standard computer program. Since the main objective of this paper is to bring out the success of this

method in solving the crystal structure of cyclo (Gly-L-Tyr-Gly)$_2$ by using only contact criteria, we restrict our discussion on the program only to the particular space group P1. However, the framework of the program is such that it can be easily modified to suit other space groups as well. The flow chart of the program used in our study of the cyclic hexapeptide molecule is given in Fig. 2.

The main input data required to carry out the calculations are:

(i) The unit cell parameters of the crystal.

(ii) The angular parameters ϕ, θ, ψ defining the regions to be scanned in the unit cell and their relative increments.

(iii) Atomic coordinates of the first asymmetric unit as obtained from conformational study.

(iv) Elements of the rotation matrix and components of unit cell translational vectors for generating the neighbouring molecules.

(v) The table of normal and extreme contact limits for every pair of atoms of known chemical species (Table I).

TABLE I

Values of limiting distances in Angstrom units for various interatomic contacts

Type of contact	Extreme limit (Column A)	Normal limit (Column B)
H ... H	1·9	2·0
H ... O	2·2	2·4
H ... N	2·2	2·4
H ... C	2·2	2·4
O ... O	2·6	2·7
O ... N	2·6	2·7
O ... C	2·7	2·8
N ... N	2·6	2·7
N ... C	2·8	2·9
C ... C	2·9	3·0

(*vi*) The number of bad contacts which will classify a particular region as being disallowed.

The major steps involved in this program can be summarised as follows:

(*a*) Establishment of the cartesian coordinates of the atoms of the asymmetric unit.

(*b*) Generation of the neighbouring molecules as required by the space group symmetry.

(*c*) Determination of the sterically allowed orientations and positions of the molecule in the unit cell by systematic variation of the angular parameters ϕ, θ, ψ and the translational parameters (X, Y, Z).

Since the space group P1 has full translational degree of freedom within the unit cell, all the translational parameters can be taken to be zero.

(*d*) Elimination of the regions that contain steric overlaps between some of the constituent atoms, found using the contact distance criteria.

(*e*) Ranking of all these allowed regions according to number of bad contacts.

The details of the procedural steps employed for the cyclic hexapeptide molecule are discussed in the following sections.

Application of the Method to the Structure of a Cyclic Hexapeptide

In this section we explain how the method was applied to work out the preliminary structure of cyclo (Gly-L-Tyr-Gly)$_2$.2H$_2$O in the space group P1, which could be subsequently refined by employing the usual techniques normally adopted in x-ray crystallography. The space group P1 is the most difficult one for the solution of the structure by the standard methods used for solving the phase problem. It is worth mentioning here that there was only one equal atom structure in P1 reported in the literature[11] when this work was started.

For two reasons the crystal structure sought to be solved was an ideal example for us to start our work. Firstly, in the space group P1, the molecule does not have any symmetry-related molecule in its unit cell. Secondly,

the solution conformation of this molecule has been determined by Kopple and coworkers [12], so that this is a fine example to verify if the solution and solid state conformations are the same.

X-ray Data and Processing

Crystals of the chosen compound, $C_{25}H_{31}N_6O_8 \cdot 2H_2O$ are colourless and needle-shaped, crystallising in the triclinic system. Cell constants and other pertinent physical data are listed in Table II.

TABLE II

Molecular formula	$C_{26}H_{31}N_6O_8 \cdot 2H_2O$
Molecular weight	588
Space group	P1
a	$6{\cdot}270 \pm 0{\cdot}002$ Å
b	$8{\cdot}808 \pm 0{\cdot}002$
c	$13{\cdot}350 \pm 0{\cdot}002$
α	$104{\cdot}18^\circ \pm 0{\cdot}01$
β	$97{\cdot}19 \pm 0{\cdot}01$
γ	$98{\cdot}46 \pm 0{\cdot}01$
V	698·1
Radiation	CuK_α
Number of independent reflections	2176
Number of molecules per unit cell Z	1

The diffracted intensities of 2176 reflections were measured on a CAD-4 diffractometer, of which 204 reflections were considered to be unobserved, since their intensity was less than 3 σ. Three standard reflections were measured after every 50 reflections and there were no significant changes in their intensities during the period of data collections. Intensities were then corrected for Lorentz and Polarization factors, but not for absorption ($\mu R = 0{\cdot}2$). An overall temperature factor of 2·2 A^2 and the scale factor were evaluated. The chemical structure of the molecule is given in Fig. 1

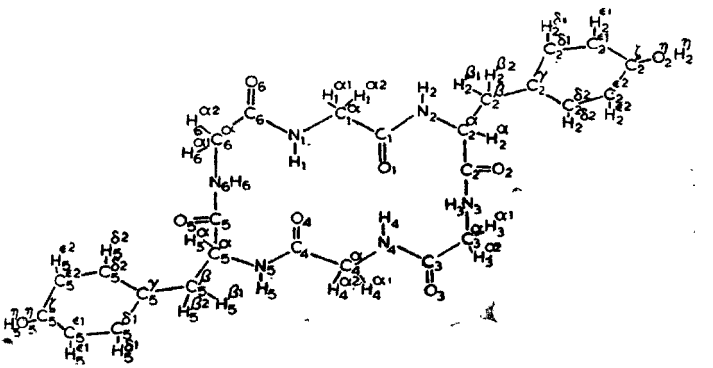

FIG. 1. Structural formula of the molecule cyclo (Gly-L-Tyr-Gly)$_2$.

Generation of the Theoretical Model

The theoretical model was constructed on the basis of the conformation of the backbone given by NMR studies. The solution conformation of this molecule has been studied by Kopple and coworkers [12] from the proton magnetic resonance spectra. They noticed that the molecule has C_2 symmetry at least in the NMR time average. The side chains, namely the tyrosine rings, occur at the second and the fifth position. The dihedral angles for the three peptide units and the side chain conformational angles obtained from NMR are given in Table III (*a*). These values of dihedral angles were used for the construction of the computer model used for packing calculation. The method adopted is same as that used by Ramakrishnan and Sarathy [13] for their study of cyclic hexapeptides with two fold symmetry. The bond lengths and angles used are the standard values for the planar and *trans* peptide units [10]. The method can be summarised as follows.

TABLE III

Chosen parameters for the cyclic hexapeptide molecule

(*a*) Conformation angles obtained from NMR data

Residue i	Nature of residue	Conformational angles*			
		Backbone angles		Side-chain angles	
		ϕ_i	ψ_i	χ_i^1	χ_i^2
1, 4	Gly	—150	—170		
2, 5	Tyr	—80	120	180	90
3, 6	Gly	70	0		

* All angles are in degrees.

(*b*) Hydrogen bond data and conformational parameters of the model chosen

Number of residue+ i	Backbone parameters* τ_i	ϕ_i	ψ_i	Hydrogen bonds
1, 4	111·0	—132·0	—162·0	Nature
				$N_1H_1 \ldots O_4$ (1→4)
				$N_4H_4 \ldots O_1$ (4→1)
2, 5	112·0	—75.0	125·0	
				Parameters for both
3, 6	112·0	70·0	1·00	Length 2·97 Å
				Angle 18°

\+ same as in Table III (*a*).

* All angles are in degrees. Side chain dihedral angles are as in Table III (*a*).

Assuming the peptide units to be planar and *trans*, given the value of τ, ϕ, ψ at two C^α atoms (C_2^α and C_3^α), three peptide units starting from C_1^α, and ending with C_4^α, can be generated. The two-fold axis is taken to pass through the centre of line joining C_1^α and C_4^α and perpendicular to the plane formed by C_1^α, C_2^α and C_4^α. Two similar sets of such three units (namely C_1^α–C_2^α–C_3^α–C_4^α and C_4^α–C_5^α–C_6^α–C_1^α) can then be linked to each other at the two terminal alpha carbon atoms C_1^α and C_4^α. They are then tilted about the line joining these two linking alpha carbon atoms such that the τ values at these carbon atoms lie close to the tetrahedral values. The resulting structure will be a hexapeptide with two-fold symmetry. This procedure was used in getting a cyclised structure.

In generating the three peptide units, six conformational parameters namely τ_2, ϕ_2, ψ_2, τ_3, ϕ_3, ψ_3 were varied. τ_2 and τ_3 were varied in the range 108° to 112°, at intervals of 2°, while (ϕ_2, ψ_2) and (ϕ_3, ψ_3) were varied in the range (−90° to −70°, 110° to 130°) and (60° to 80°, − 10° to 10°) respectively, at intervals of 10°, and the tilt was given in the range from 0° to 180°. Only those values of (ϕ_1, ψ_1) for which τ_1 lies in the range of 108° to 112° and occurs in the range (−160° to − 140°, − 180° to − 160°) were printed. A large number of possibilities were thus obtained. However, and important feature, which is found to occur in cyclic hexapeptide structures, is the existence of a 4 → 1 type of internal hydrogen bond which brings about a reversal in the direction of progress of the polypeptide chain and thus leading to the closure of the ring [13, 14]. When this additional criterion was introduced for the formation of good cyclic structures, the number of possibilities were reduced appreciably. Out of these, the one which had good hydrogen bond parameters and for which the conformational angles were close to the ones reported from NMR studies was chosen. The data for this trial structure are shown in Table III (*b*).

Having obtained the backbone with two-fold symmetry, the tyrosine side chains were fixed at C_2^α and C_3^α using the values of $\chi^1 = 180°$ and $\chi^2 = 90°$, the same as those given by NMR studies. The average bond length for C—C in the phenyl ring was taken to be 1·40 Å and the bond angles were taken to be 120°. The C—O bond length was given a value of 1·45 Å. These data were taken from the paper by Ponnuswamy and Sasisekharan [15]. For one of the tyrosine side chains, namely the one at C_5^α, a positive rotation of 155° was made about $C^\alpha{}_5 — C_5^\beta$, increasing χ^1 by this value, in order to break the two-fold symmetry of the molecule, since the molecule as a whole is expected to be asymmetric in its triclinic unit cell.

In the computer model of the molecule, the hydrogen atoms H^{η} of the tyrosine rings were not included. Due to the flexibility of these two hydrogens it would have been unrealistic to assign precise coordinates to these atoms. Also if rigid coordinates had been assigned to them, some legitimate orientations might have been eliminated during the screening for sterically allowed orientations.

Packing of the Model in the Unit Cell

Although this molecule is fairly large and has several degrees of freedom in terms of dihedral angles of rotation in its internal molecular structure, this space group P1 is particularly favourable for the packing analysis method. The molecule in the unit cell has full translational degree of freedom and any atom on point in the molecule can be taken as the origin of the unit cell. There are only three rotational degrees of freedom. Hence the problem reduces to a three parameter case and it is sufficient if only an orientation search is made and the translational parameters will all be zero, and need not be varied.

The calculational steps which are explained in this section are all programmed for an IBM 360/44 computer for the space group P1. The flow chart of this program is given in Fig. 2.

The cartesian coordinates of the atoms of the generated model in the so-called starting orientation were calculated. In this orientation the centre of the hexapeptide ring, O, which is the intersection of the lines $C_1^{\alpha}C_4^{\alpha}$ and $C_2^{\alpha}C_5^{\alpha}$ is taken to be the origin of a right handed coordinate system with the Z-axis, perpendicular to the plane of these four atoms and coinciding with the two-fold axis. The X-axis is taken along OC_1^{α} and the Y-axis perpendicular to it in the above plane.

Using the coordinates of all the atoms in this initial frame, corresponding to $\phi = 0$, $\theta = 0$, and $\psi = 0$, the coordinates of the atoms, for any set of values for ϕ, θ, ψ, are calculated using the transformation matrix R as given below:

$$\begin{pmatrix} \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi & \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi & \sin\psi\sin\theta \\ -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi & -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi & \cos\psi\sin\theta \\ \sin\theta\sin\phi & -\sin\theta\cos\phi & \cos\theta \end{pmatrix}$$

The number of neighbouring molecules that have to be considered for the packing analysis were precalculated using a simple program. The

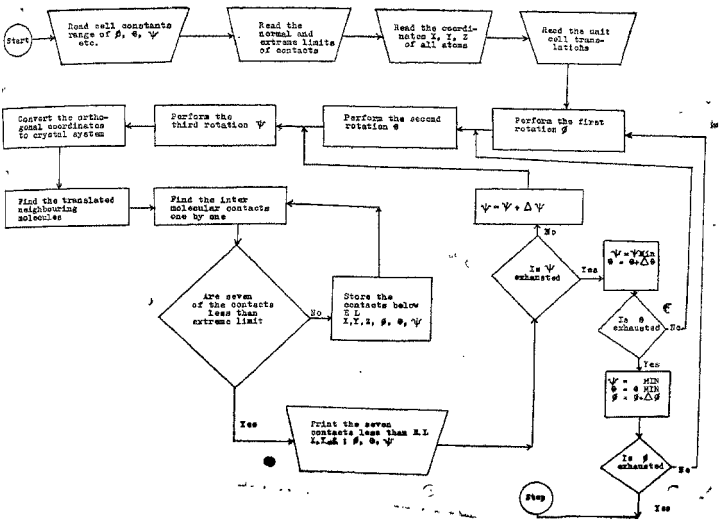

FIG. 2. Flow chart of the packing program.

interactions of only those molecules with their centres at a distance of less than 15 Å from the first molecule were found to be significant. This amounted to taking seven translated neighbours, the positions of which are given in Table IV. Due to the uncertainty in the measured density, the presence or otherwise of the water molecules could not be concluded; hence only the hexapeptide molecule with 80 atoms was taken for the packing analysis.

In the beginning, we tried to eliminate the obviously disallowed regions by packing the molecules by trial and error in the unit cell constructed for this purpose. But our attempts were not successful due to the large size of the molecule. Furthermore, the formation of hydrogen bonds were also not very clear.

Starting from the arrangement of atoms characterized by $\phi = \theta = \psi = 0$, and systematically varying the three Eulerian angles ϕ, θ, ψ at 20° intervals, all the conceivable orientations of the molecule were generated on the computer one by one. The range of angles covered was 0° to 180° in ϕ,

TABLE IV

Equivalent points of the hexapeptide molecules which are taken into account in the packing calculations

Molecule number	Coordinates
1	x, y, z
2	$1 + x, y, z$
3	$x, 1 + y, z$
4	$x, y, 1 + z$
5	$1 + x, 1 + y, z$
6	$1 + x, y, 1 + z$
7	$x, 1 + y, 1 + z$
8	$1 + x, 1 + y, 1 + z$

0° to 180° in θ and 0° to 360° in ψ. However, many of the orientations could be eliminated on steric grounds. The sterically allowed orientations were selected as follows.

After performing each rotation (ϕ, θ, ψ), the seven neighbouring molecules were generated. The intermoleclar contacts between the first asymmetric unit and its neighbours were calculated one by one and compared with the contacts, given for every pair of atoms in Table I, by comparing with the "extreme limits" for the corresponding pair of atoms. The "extreme" contact limits are somewhat shorter than the normal contact distances between various atoms, which are shown in column B of Table I. This conservative choice of contact limit values allowed the retention of those orientations which represented only marginal overlaps, between some of the atoms. In the beginning, we had some difficulty in fixing the number of bad contacts which will designate a particular orientation as

being disallowed. When this was taken to be ten contacts, we found that a large number of regions could be categorized as being allowed. The number was then reduced to seven, and the permissible orientations were then reduced to a much smaller number. Thus, if in a certain generated orientation there are more than seven intermolecular bad contacts between the molecule at the origin and its neighbours, the orientation was rejected as being sterically impossible.

All the allowed orientations which were obtained in this way could be grouped into six independent regions which were relatively free of short contacts. These are shown in Table V. The time taken for these calculations on an IBM 360/44, with a memory of 32 K words, was about 6 hours, for the triclinic unit cell, with one molecule in the unit cell. If the symmetry is higher, much larger time would be required.

Having obtained these approximate regions, using a large increment of 20°, a further scan was made, in these six regions using a smaller scan of 5°, in order to locate the nearly correct allowed regions. This did not lead to any appreciable reduction in the number of contacts in each region, and, therefore, the six cases A to F given in Table V were used for further

TABLE V

The six allowed regions

Case	ϕ	θ	ψ	No. of bad contacts
A	40	0	120	3
B	40	140	20	3
C	80	40	120	4
D	140	120	100	5
E	120	20	220	6
F	180	20	20	6

studies. The details of these contacts are presented in Table VI. From an analysis of these results, it was not immediately obvious which of these regions represented the correct model. It was at this stage that x-ray data were utilised to pick out the most probable model.

Selection of the Trial Model

The selection of the refinable trial model was accomplished using the observed x-ray diffraction data. The atomic coordinates corresponding to all local minima A to F were used for structure factor calculation and compared with the observed x-ray data. Using all the observed reflections, the R-values were computed for the six allowed regions. As can be seen from Table VII, four of these orientations gave a high R-value in the region of 0·85, while the other two regions A and B gave it in the neighbourhood of 0·60, which is not far from the theoretical value of 0·59 for R-index given by Wilson [16] for non-centric and randomly distributed atoms. Since this did not give any indication as to the correctness or otherwise of the structure, it was decided to carry out the least squares refinement on the structures A and B, since both of them showed only a small difference in in the R-value. In anticipation of what follows, we may say that one of the two choices turned out to be a good trial model.

Verification of the Correctness of the structure

The refinement was carried out using the block diagonal least-squares refinement program of Shiono [17]. The quantity minimized is $\Sigma\omega\ (F_o^2 - S^2 F_c^2)$ where $1/S$ is the scale factor obtained from the Wilson plot and $\omega = 1$ initially. Two cycles of least-squares refinement was carried out on both the structures A and B, using all the 2,176 observed reflections. At the end of the second cycle, there was no noticeable change in the R-value, which oscillated in the region of 0·60. We then decided to carry out the refinement, using initially a few low angle reflections and then using this refined model for further refinement with all the reflections. Hence, about 300 reflections, with $\sin\theta \leqslant 0{\cdot}3$, were used to refine both structures A and B and the final coordinates after two cycles were the input for refinement in the next stage with all the reflections. This did not improve the situation any better. Therefore, we continued our attempts and proceeded in the following manner.

The refinement was carried out initially using only a few low angle reflections and adding extra number of reflections in steps in the subsequent stages. This procedure resulted in the establishment of the correct trial

TABLE VI

Details of bad contacts

Structures	Nature of bad contact	Contact distance*
A	O2 ... O4	2·47 (2·6)
	O5 ... N2	2·43 (2·6)
	O5 ... H2	1·94 (2·2)
B	O5 ... C3	2·63 (2·7)
	O6 ... H2	2·07 (2·2)
	C5 ... H3	2·04 (2·2)
C	O2 ... O1	2·56 (2·6)
	H4 ... H3	1·89 (1·9)
	O5 ... N2	2·23 (2·26)
	C5 ... H2	2·07 (2·2)
D	O4 ... O2	2·30 (2·6)
	O5 ... C3	2·64 (2·7)
	O5 ... C4	2·06 (2·7)
	O6 ... H2	1·90 (2·2)
	C5 ... H3	1·92 (2·2)
E	N5 ... N2	2·30 (2·6)
	O4 ... O5	2·01 (2·6)
	N5 ... O2	2·11 (2·6)
	H5 ... N2	1·93 (2·2)
	C3 ... N5	2·33 (2·8)
	C6 ... O2	2·43 (2·7)
F	N2 ... C1	2·40 (2·8)
	N4 ... O6	2·53 (2·6)
	O3 ... C2	2·42 (2·7)
	O3 ... C1	2·27 (2·7)
	H2 ... H1	1·81 (1·9)
	C3 ... N4	2·10 (2·8)

*The distances are in Å and the numbers in brackets are the extreme limit values.

TABLE VII

R-values for the six probable orientations

Structures	ϕ	θ	ψ	R (2,176 reflections)
A	40	0	120	0·594
B	40	140	20	0·614
C	80	40	120	0·885
D	140	120	100	0·750
E	120	20	220	0·921
F	180	20	20	0·830

model which was progressively refined by adding about three hundred reflections at each stage. By making a series of five steps, a structure with accuracy of 0·01 Å and with R = 0·30 was obtained. The details of the individual steps are given below.

Step 1 : The first step of the calculation was done using about 300 reflections with $\sin\theta \leqslant 0{\cdot}3$. An average temperature factor of 2·2 A^2 was assumed. The coordinates of C_1^{α} was held fixed to prevent a singular matrix. Only the positional parameters and the scale factor were refined. At the end of two cycles of refinement the R-value for structure A was 0·40 and for B, it was 0·45. The shifts in the coordinates were of the order of 0·5 Å on the average. The bond length analysis for both the structures showed a better distribution of values for Å which ranged from 0·9 A to 1·75 Å than for B, which recorded values as low as 0·6 Å for N—C^{α} and C^{α}—C bonds.

Step 2 : The second step of the calculation was done for both the models using the final coordinates of Step 1 as the input coordinates. Three hundred more reflections were added with sin θ going up to 0·45. Refinement of both the temperature factors and positional parameters showed that the

structure A is more favourable than B. There were significant changes in the coordinates of certain atoms in the structure A, in the direction which will improve the bond lengths closer to the correct value. The R-value was 0·40 and the temperature factor ranged from 1·2 Å^2 to 4·0 Å^2.

The details of these two steps of refinement for both the structures are given in Table VIII.

TABLE VIII

Analysis of the refinement at the first two stages for structures A and B

No. of reflections	Average shifts (Å)	Bond length range (Å)	Temperature factor	R
		Structure A		
300	0·5 Å	0·9–1·8	2·2	0·402
600	0·2 Å	1·1–1·7	1·5–3·5	0·400
		Structure B		
300	0·6 Å	0·9–1·9	2·2	0·472
600	0·5 Å	0·6–1·9	Temp. factor of five atoms not positive definite	0·452

Step 3: Further refinements were carried out only with structure. About 900 reflections were taken in the range of 0·0 to 0·55 for sin θ. The B factors of all the atoms were initialised to 2·2 Å^2. Convergence was obtained after two cycles of refinement and temperature factors ranged from 1·35 Å^2 to 3·8 Å^2. The R-factor dropped to 0·394.

Step 4: In the subsequent step, in which about 1,200 reflections were used, the R-value dropped down to 0·36 and the accuracy of the structure also increased. The bond lengths between the atoms were very close to their respective normal values.

Step 5: Finally a structure with an accuracy of 0·01 Å in the coordinates and R-factor of 0·30 was obtained including all the observed reflections.

Table IX gives the analysis of the refinement at these various stages.

TABLE IX

Summary of the different stages of refinement of structure A, which led to the correct structure

No. of reflections	300	592	942	1216	A 11
Total shift in each step	0·5 Å	0·7 Å	0·1 Å	0·02 Å	0·01 Å
Bond lengths	0·9–1·8	1·1–1·7	1·2–1·6	1·2–1·56	1·2–1·56
Temperature factor	2·2	1·5–3·5	1·5–3·2	1·7–3·2	1·7–3·1
R Initial	0·540	0·441	0·484	0·504	0·440
R after 2 cycles	0·402	0·400	0·394	0·360	0·300

Refinement of the structure

Further refinement on this structure did not bring down the R-value below 0·30. Hence the possibility of the presence of water molecules was thought of and the difference-Fourier map was calculated. There were two major peaks of strength about ·6e/Å^3 and six peaks of comparatively low strength distributed very close to both the tyrosine ring atoms. Considering the two major peaks as water molecules gave a very satisfactory hydrogen bonding scheme. Incorporation of these two major peaks as water oxygens readily led to a refined structure with an R-value of 0·131 at the end of three cycles.

The hydrogen atoms in the molecule were fixed theoretically and were all found to appear at reasonable positions on peaks of electron densities between 0·35 to 0·66 eA^{-3} in the difference-Fourier calculated at an R-value of 0·131. The positions of water hydrogens and that of H$^{\eta}$ of tyrosine rings were also fixed from the directions of the hydrogen bonds they are involved in. Further refinement of the structure was carried out using the full matrix refinement program [18]. Employing Hughes weighting scheme [19] and including all the hydrogens at theoretical positions brought the R-factor down to 0·081. The hydrogens were used only in the structure factor calculations and were assigned the temperature factors of the heavy atoms to which they are bonded. Refinement with anisotropic temperature factors for the non-hydrogen atoms brought down the R-factor to the final value of 0·051. The shifts in the coordinates of the atoms in the last cycle of refinement were less than 1/20th of their estimated standard deviation. The goodness of fit, $\Sigma\omega(F^0 - F_c)^2/(n-p)$, for $n = 2176$ reflections and $p = 376$ parameters, was 1·02 [20].

The final fractional coordinates of the structure will be given in a succeeding paper, in which the conformational aspects are discussed. The average shifts in the coordinates of the atoms between the starting structure and the final one was about 0·8 Å. The movements of the atoms from the trial structure to the final structure is shown in Fig. 3, in which both the structures are superposed on each other with their centres superposed.

It is interesting to note that, although a structure with the symmetry of a 2-fold axis was assumed, to start with, for the hexapeptide molecule, the

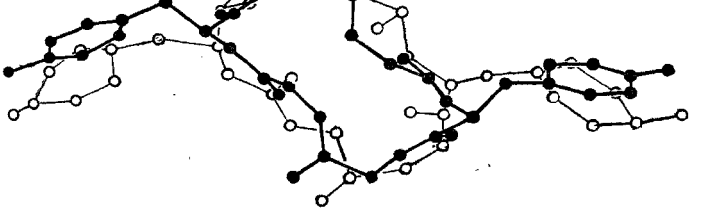

FIG. 3. Movement of the atom from (*a*) the trial structure (thin lines) to (*b*) the final structure (thick lines).

refined structure does not have this approximate symmetry. Instead, the backbone has an approximate inversion symmetry and the two tyrosine side-groups also have entirely different dihedral angles, namely $\chi_2^1 = 80°$, $\chi_2^2 = -72°$ and $\chi_5^1 = -80°$, $\chi_5^2 = -80°$, at residues 2 and 5 respectively. This shows that the refinement, using successively larger number of reflections, has succeeded, even though the trial molecule had quite a different conformation from the correct one.

After completing this study we were interested to know how the structure, C, D, E, F would behave if refined by the same process using a stepwise addition of reflections. We carried out the same procedure as was done for A and B. The results obtained are shown in Table X. All these structures behaved erratically at the stage of even 600 reflections and did not indicate that they would converge to the correct structure. The bond lengths and temperature factor distribution also were not satisfactory.

TABLE X

Refinement details of the first two stages for the four structures C, D, E, F

Structure	No. of reflections	Average shifts (Å)	Bond length range (Å)	Temperature factor	R
C	300	0·11	0·7–1·90	2·2	0·589
	600	0·07	0·5–1·8	—ve for 4 atoms	0·562
D	300	0·18	0·8–1·9	2·2	0·532
	600	0·07	0·6–1·8	—ve for 8 atoms	0·494
E	300	0·13	0·6–1·7	2·2	0·601
	600	0·05	0·5–1·8	—ve for 5 atoms	0·564
F	300	0·10	0·7–1·8	2·2	0·653
	600	0·02	0·5–1·8	—ve for 6 atoms	0·641

Conclusion

The packing method, described in this paper, could be applied with reasonable success for molecules which are fairly rigid and have only small regions which have freedom of motion in their structure. The advantage of employing a stepwise addition of reflections in the least squares refinement procedure is to be particularly noted. This procedure will be useful in cases wherein the trial structure is accurate only to about 0·5 Å to 1·0 Å in the location of the atoms in the molecule. Attempts are being made in this laboratory for solving structures with space groups of higher symmetry using this packing analysis method.

Acknowledgement

It is a pleasure to thank Prof. G. N. Ramachandran for suggesting the problem and for his invaluable discussions. The assistance from grants given by CSIR (Silver Jubilee Grant "Protein Food from Cellulosic Plant Materials") and DST/SERC (India) is gratefully acknowledged.

References

1. Williams, D. E. — A method of calculating molecular crystal structures. *Acta Cryst.*, 1969, **A2**, 464–470.
2. Kitaigorodsky, A. I. and Mirsakaya, K. V. — Quadrupole Interaction in a molecular crystal. *Soviet Physics. Crystallography*, 1965, **10**, 121–124.
3. Coiro, V. M., Giacomello, P. and Giglio, E. — Determination of the molecular packing in the crystal of N, N'-dicyclo hexyl urea by means of potential energy calculations. *Acta Cryst*, 1971, **B27**, 2112–2119.
4. Ramachandran, G. N. — Molecular forces in protein structure and crystallography. *Protein Research*, 1969, **1**, 1.
5. Ramachandran, G. N., Sarathy, K. P. and Kolaskar, A. S. — Interatomic potential functions and crystal packing. *Zeitschrift fur Kristallographie*, 1973, **138**, 299–312.
6. Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. — Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, 1963, **7**, 95–99.
7. Ramachandran, G. N. — *Treatise on Collagen*, 1967, Vol. 1, Edited by G. N. Ramachandran, pp. 103–183, New York: Academic Press.
8. Bansal, M. Ramakrishnan, C. and Ramachandran, G. N. — A triple helical model for (Gly-Pro-Hyp)$_n$ with *cis* peptide units. *Biopolymers*, **14**, 2457–2466.
9. Ramachandran, G. N. and Shamala, N. — Crystal structure from packing: Application to the triclinic structure of the cyclic hexapeptide cyclo (Gly-L-Tyr-Gly)$_2$. *Acta Cryst.*, 1976, **A32**, 1008–1009.

10. Ramachandran, G. N. and Sasisekharan, V. — Conformation of polypeptides and proteins in *Advances in Protein Chemistry*, 1968, **23**, 283–438.

11. Kistenmacher, T. J., Hunt, D. J. and Marsh, R. E. — The crystal and molecular structure of L-N-Acetylhistidine monohydrate: An application of direct methods to space group P1. *Acta Cryst.*, 1972, **B28**, 3352.

12. Kopple, K. D., Go, A., Logan, A. R. and Savrda, J. — Conformations of cyclic peptides. VI. Factors influencing mono-, 1, 4-di and 1, 2, 4-trisubstituted cyclic hexapeptide backbones. *J. Amer. Chem. Soc.*, **94**, 973–981.

13. Ramakrishnan, C. and Sarathy, K. P. — Stereochemical studies on cyclic peptides. IV. Conformational analysis of cyclopentapeptides. *Int. J. Protein Research* 1969. **1**, 63–71.

14. Chandrasekaran, R., Lakshminarayanan, A. V., Pandya, U. V. and Ramachandran, G. N. — Conformation of the LL and LD hairpin bends with internal hydrogen bonds in protein and peptides. *Biochem. Biophys. Acta*, 1973, **303**, 14–27.

15. Ponnuswamy, P. K. and Sasisekharan, V. — Studies on the conformation of amino acids. V. Conformation of amino acids with δ-atoms. *Int. J. Protein Research*, III, 1971, 9–18.

16. Wilson, A. J. C. — Largest likely values for the reliability index. *Acta Cryst.* 1950, 3, 397.

17. Shiono, R. — SFLS Program, Private Circulation.

18. Gantzel, P. K., Sparks, R. A. and Trueblood, K. N. — University of California Program, UCLALSI, 1961.

19. Hughes, E. W. — The crystal structure of melamine. *J. Amer. Chem. Soc.* **94**, 973–981.

20. *International Tables for X-ray Crystallography* — Vol. III, 1962, Birmingham: Kynoch Press.