CONFORMATIONAL ANALYSIS

OF

CYCLIC TETRA-PEPTIDES

BY

GLOBAL ENERGY MINIMISATION CALCULATIONS

A Thesis presented for

the degree of Doctor of Philosophy

in the

Faculty of Science

of the

University of Glasgow

by

Christopher Morrow

October 1978

ACKNOWLEDGEMENTS

Glasgow, 1978   C. Morrow

Error is a hardy plant — it flourishes in every soil.

Anon.

CONTENTS

Chapter Five - Correlation of Observed and Calculated Conformations

Appendix - GLOMIN

SUMMARY

An algorithm, based on Molecular Mechanics calculations, was developed to locate the Global Minimum Energy Conformation (GMEC) of cyclic molecules. This was used to find the GMEC of cyclotetraglycyl and also those conformations within 20 kcal/mole of this conformation. The resulting conformation is composed of four trans amino acids and has S4 symmetry. The calculated GMEC corresponds to an observed crystal structure of dihydrochlamydocin and to a proposed conformation of cyclotetraglycyl itself.

The ring conformations of all observed cyclic tetra -peptides and -depsipeptides were found to correspond, more or less exactly, to some conformation in the calculated low energy set of cyclotetraglycyl. Only one exception was found and this seems to be an intermediate between two calculated conformations.

The algorithm has now been further streamlined into a single computer-program and is currently being used to locate the GMEC of cycloalkanes (e.g. cycloundecane) in this department and elsewhere.

Calculations were also performed on various derivatives of cyclotetraglycyl. Protonation and N-Methylation were found to profoundly alter the relative energies of the various conformations in the low energy set. It was observed that for these derivatives the cis, trans, cis, trans sequence of amino acids is more stable than the all trans sequencs.

In addition to peptides the conformations of other 12-membered rings were found to correspond to members of the calculated set of conformations. These included 12-crown-4 ethers and cyclododecanes.

In order to facilitate all of these operations a Chemical Graphics System (CGS) was developed. This is an extensive suite of computer programs which utilises a PDP11/40 minicomputer with a VR17 graphics terminal. The CGS allows interactive building and manipulation of molecular models and includes various advanced features. Foremost of these is the dynamic rotation facility which allows the user to observe a structure as it rotates about the X,Y or Z axis. Measurement of distances,

angles and torsion angles is easy and the constructed structure can be directly used as input to a Molecular Mechanics calculation.

CHAPTER ONE

An Introduction to Molecular Mechanics Calculations

1.1 Introduction

Conformational analysis and the linking of three dimensional structure of a molecule with reactivity were not generally appreciated concepts prior to 1950. In that year D.H.R.Barton published his work [1] relating the chemical reactivity of steroids and terpenoids to conformation. However it was in the last century that the seeds of this idea were sown by Dalton's atomic theory and the concept of isomerism proposed by Berzelius; and perhaps of greater import and significance to modern structural chemistry the idea of the three-dimensional arrangement of atoms propounded by Le Bel and van't Hoff in 1874.

Before these ideas were proposed chemistry had been merely the science of labelling compounds according to their gross observable properties. We can now appreciate how these measurable quantities and phenomena are related not only to the atomic composition of molecules (the theory of valency has led to structural formulae) but also, in many instances, to the three-dimensional arrangement of the constituent atoms and groups of atoms in the molecule. This latter idea is playing an increasingly important role in the understanding of the behaviour of molecules, especially those which take part in the chemistry of life. It is now realised that the biological activity of proteins, enzymes etc. is intimately connected to the three- dimensional structure of all of the species involved.

Thus over the last quarter-century increasing emphasis has been placed upon those methods (both experimental and theoretical) which are capable of determining the detailed structure of molecules — large and small. The aim of all of these methods is to provide information about the geometry and/or energy of the molecule under examination. None of the existing experimental methods gives a complete picture of BOTH the energy and geometry of a molecule. Thus diffraction methods (X-ray, neutron and electron) give excellent descriptions of geometry but no information whatsoever about energy while spectroscopic methods (NMR, IR and ORD/CD) give incomplete information about both geometry and energy.

Having obtained, from whatever source, the structure of a molecule it is often of interest to determine how far the various internal coordinates of the molecule (bond lengths, valence angles and torsion angles) deviate from their optimal values. Where deviations of this kind exist the molecule is said to be "strained". Strain can, intuitively, be appreciated in terms of deformed functional groups - non-planar aromatic and olefinic groups are indicative of strain in a molecule. Qualitatively, therefore, strain is a readily appreciable phenomenon. Difficulties arise in attempting to quantify this strain; how does one compare, for example, the degree of strain in two essentially dissimilar molecules which both happen to contain more or less non-planar benzene rings?

The experimental difficulties encountered in measuring strain may be overcome by various methods of calculating the molecular properties. Quantum mechanical calculations, for instance, may be used to determine energy, geometry and electron density. The most theoretically sound type of quantum mechanical calculation are undoubtedly the ab-initio calculations, unfortunately these are impractical for use beyond twenty or so atoms due to the enormous amount of computer resources required. The run-time requirement of this type of calculation rises as n^4 where n is the number of orbitals involved. Simplified, or semi-empirical, types of quantum mechanical calculations are available (EHT, CNDO, MINDO) but are of restricted utility due to the approximations inherent in them.

An alternative, wholly empirical, type of calculation is utilised in the present work. Molecular mechanics (empirical Force-Field) calculations are far less costly of computer time (time increases as N^2 where N is the number of atoms) and are relatively easy to perform. This method requires, however, the existence of a substantial body of experimental data from which to derive a force-field. This force-field can, by extrapolation, be used to provide information on new, or indeed unobserved compounds. The results obtained in this way are found to have comparable accuracies to many experimental techniques.

Molecular mechanics calculations have been performed, in recent years, on an increasing number and type of organic molecules and have been most useful in the interpretation and understanding of various experimental results. The elucidation of molecular conformation, thermodynamic properties, reaction mechanisms and kinetics have been undertaken as well as interpretation of spectroscopic (both vibrational and dynamic NMR) results.

Further discussion of the general aspects of molecular mechanics may be found in the reviews contained in Refs. 2-8.

1.2 Potential Functions

There are a variety of different techniques used in molecular mechanics calculations all of which, however, express the molecular strain energy as the sum of contributions from the various geometric deformations (e.g. bond length stretching, bond angle bending etc.). In most current methods the geometry of the molecule is described by cartesian (orthogonal) or crystal coordinates from which the internal coordinates can be derived quite simply (see chapter three). It is assumed that these internal coordinates have an equilibrium (or strain-free) value, deviations from which cause increases in strain (normally referred to as steric or molecular potential energy) according to the relevant potential function. These potential expressions are partially derived from experimental results and may be used to calculate the strain energy of a molecule.

The total molecular potential energy, V_S, is given by the sum of six components, which are assumed to be largely independent of one another in this treatment, and whose precise form depends on the particular treatment being used:-

$$V_S = V_\ell + V_\theta + V_\omega + V_\chi + V_q + V_r , \qquad (1)$$

where the various terms are the components of energy arising from bond

length deformation, bond angle bending (Baeyer strain), torsional (Pitzer) strain, out-of-plane bending at trigonal atoms, charged (coulombic) and non-bonded (van der Waals') interactions respectively.

In the above treatment coupling between the components (cross terms such as stretch-stretch and stretch-bend) is small [9], and is therefore neglected. The following sections describe each of these terms in more detail.

(a) Bond length deformation

Bonds may be regarded as the equivalents of molecular springs and therefore they have an equilibrium (or strain-free) length. The energy required to lengthen or shorten a bond (and hence the potential energy stored by the deformed bond) can be calculated from Hooke's Law.

$$V_\ell = \sum_\ell \frac{1}{2} k_\ell (1 - 1_0)^2, \qquad (2)$$

where 1_0 is the equilibrium bond length, 1 is the "observed" bond length, and k_ℓ is the force constant for the type of bond being considered (k_ℓ will vary in proportion to the bond order). The deformation of a bond is costly in terms of energy; lengthening or shortening of, for example, a typical single bond by 0.03Å involves expending 0.3 kcal/mole.

(b) Bond angle bending

"For small displacements from the equilibrium valence angle one may again utilise Hooke's Law to calculate the potential energy involved. However for larger deviations (> 25°) it is necessary to introduce a cubic term into the expression in order to take into account the vibrational anharmonicity of molecules. An expression such as equation (3) is found to be suited to our model [10].

$$V_\theta = \sum_\theta \frac{1}{2} k_\theta (\Delta\Theta^2 - k_\theta' \Delta\Theta^3), \qquad (3)$$

where $\Delta\Theta = \Theta - \Theta_0$, Θ is the "observed" angle, Θ_0 is the equilibrium angle, k_θ is a force constant depending on the stiffness of the valence angle under consideration and k'_θ is the cubic (anharmonic) force constant.

Bond angle bending is much less difficult than bond length deformation and is consequently energetically cheaper. Variation of bond angles from their equilibrium values is a fairly common occurrence since a deviation of 1^0 from the strain-free angle in a C-C-C system requires only 0.01 kcal/mole.

(c) Torsional strain
___ _____ _____

Pitzer strain arises from 1,2 interactions between groups attached to bonded atoms. The changes in energy arising from variation of torsion angles from their minimum energy values may be calculated using equation (4).

$$V_\omega = \Sigma \frac{1}{2} k_\omega (1 + S \cos n\omega) , \qquad (4)$$

where ω is the "observed" torsion angle, k_ω is the barrier to free rotation, S is -1 for rotations where the eclipsed form is the energy minimum e.g. C=C-C-C or C=C-C-H (rotation around central C-C bond) and S = $+1$ for rotations which have a staggered conformation as minimum; and n is the periodicity i.e. the number of times a similar conformation occurs in one revolution of the torsion angle. Equation (4) becomes, for the C=C-C-H system,

$$V_\omega = \frac{1}{2} k_\omega (1 - \cos 3\omega) .$$

Torsion angles are signed quantities corresponding to the Prelog-Ingold [31] convention which may be described as follows. In figure 1.1, A must be rotated clockwise to eclipse D, the torsion angle A-B-C-D is therefore positive. A must be rotated counter-clockwise to eclipse E, the torsion angle A-B-C-E is therefore negative.
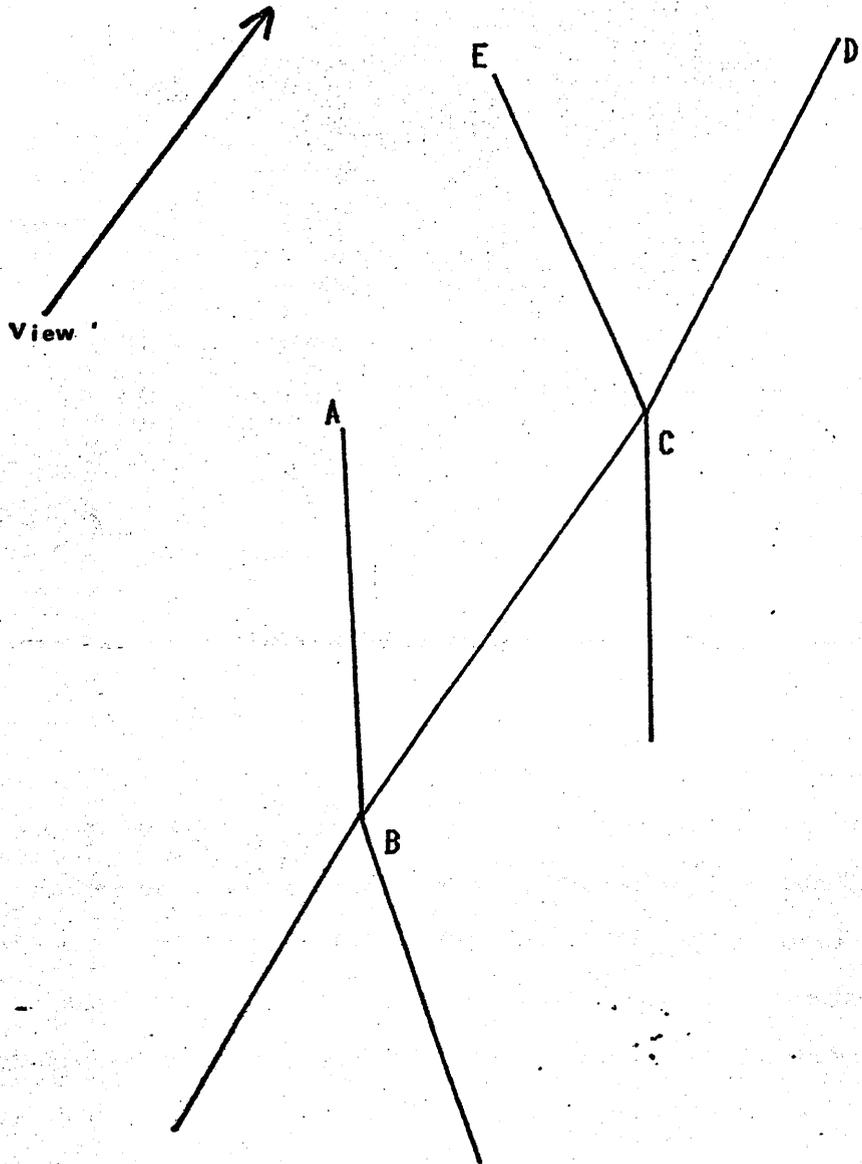
Figure 1.1 Torsion Angle Convention

The physical phenomena giving rise to these barriers to rotation have not been clearly determined, although numerous theories have been advanced [11]. Quantum mechanical calculations indicate that they arise from interactions between atoms NOT bonded to one another and it is therefore necessary to consider various terms such as substituent-substituent and electron-electron repulsions, nuclear-electronic attraction and electronic kinetic energy. Wherever these barriers come from they certainly exist and their omission from a molecular mechanics model give rise to serious and unpredictable errors in results.

(d) Out-of-plane bending

The groups (or atoms) attached to trigonal (such as olefinic and carbonyl) atoms should, ideally, be coplanar with each other and the trigonal atoms i.e. R, R', C and O, in figure 1.2, should all lie in the same plane.

R
\
C=O
/
R'
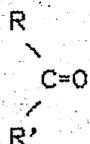
Figure 1.2

Any deviation from planarity results in a weakening of π-orbital overlap and hence of the π-bond itself. Out-of-plane bending potentials, in addition to proper torsional potentials, must be included to ensure that the geometry and energy of olefins, ketones, peptides etc. may be correctly calculated. The potential energy arising from deviations from planarity of this kind is represented by equation (4).

$$V_X = \sum_X \frac{1}{2} k_X (180 - X) , \qquad (4)$$

where k_X is the out-of-plane bending force constant and X is the improper torsion angle at the trigonal atom [12].

(e) Coulombic interactions

Coulombic (or electrostatic) interactions occur in molecules with two or more electric dipoles, e.g. peptides, and must therefore be considered in the calculation of energy and geometry of such types of compounds. The potential energy derived from interactions between charged (or partially charged) atoms is given, from classical electrostatics, by:-

$$V_q = \sum_q q_i q_j / D r_{ij} , \qquad (5)$$

where D is the effective dielectric constant between the charged atoms, q_i and q_j are the charges on atoms i and j respectively and r_{ij} is the interatomic distance.

(f) Non-bonded interactions

The existence of solids and liquids are indicative of the presence of attractive forces between atoms which are formally non-bonded. However the effort required to compress these materials beyond a certain point indicates that repulsive forces also exist between atoms and groups of atoms. These non-bonded interactions are the most difficult contributors to the overall steric energy to deal with, as will later become clear.

The potential energy curve between an isolated pair of atoms (or molecules) has the form represented by figure 1.3. This holds for spherically symmetric units (atoms or molecules) and the potential energy (V) is a function of the separation of the unit (atomic or molecular) centres (r). It can be seen that there is an equilibrium (minimum energy) separation at r_0. For non-sperical units the potential energy is also a function of the orientation of the two units [13,14]. The functional form of the non-bonded potential should contain a term representing this dependence on orientation. It is unfortunate that it has so far proved impossible to determine the form of this term since non-bonded interactions are crucially important [7] in determining both the energy and geometry of molecules.
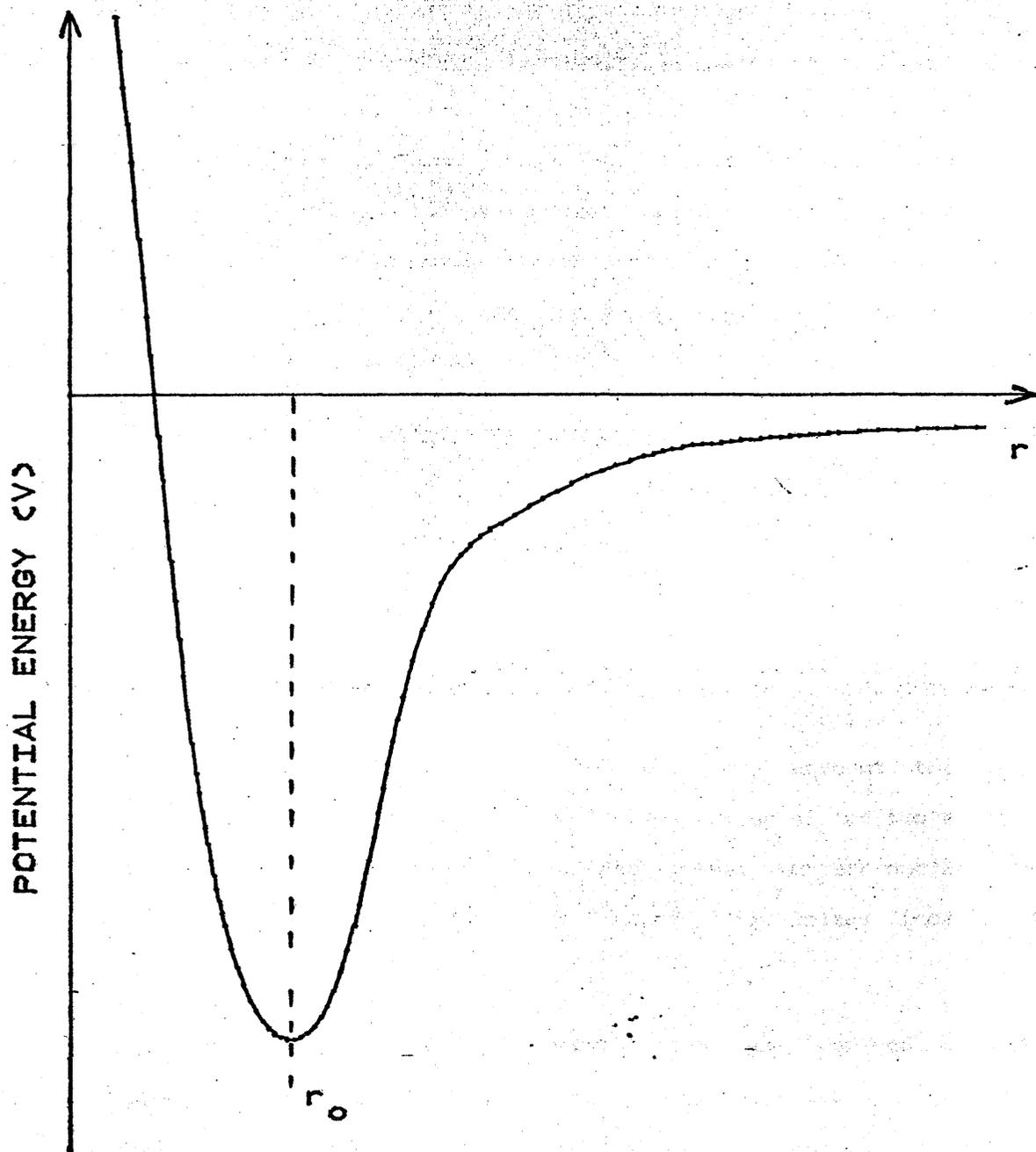
FIGURE 1.3 NON-BONDED POTENTIAL

The forces between pairs of atoms (or molecules) may be derived from the slope of the potential curve, short range interactions (< 3Å) are repulsive while longer range (> 4Å) interactions are attractive. The potential curve for any pair of atoms can, in principle, be calculated from the deviation from ideal gas behaviour of the corresponding gas. These van der Waals' curves are however difficult to obtain accurately and are available for a limited number of cases only (e.g. the noble gases [15], CH_4, and N_2). The functional form of V is therefore taken from interatomic and intermolecular interaction potentials [16]. This has proved to be reliable although the extrapolation involved lacks an unequivocable physical basis — neglecting as it does the nature of the material* between the atoms.

The most commonly used functional forms for non-bonded potentials are the Lennard-Jones potential (6) and the Buckingham potential (7).

$$V_r = \sum_r (Ar^{-n} - Br^{-6}) \; , \; n = 9 \; or \; 12 \qquad (6)$$

$$V_r = \sum_r (Aexp(-Br) - Cr^{-6}) \qquad (7)$$

The first term in both of these expressions takes into account the short range repulsive forces arising, when the separation of the two atoms is less than the sum of their van der Waals' radii, from nuclear-nuclear and electron-electron repulsion etc. The second term arises from London dispersion forces.

The precise form (8) used in the present work was derived [17] by least squares fit of observed and calculated properties of crystalline hydrocarbons (e.g. heats of sublimation, lattice constants and structural details).

$$V_r = \sum_r \varepsilon (-2 \, \alpha^{-6} + exp(12(1 - \alpha))) \; , \qquad (8)$$

*The two atoms are treated as if they existed in isolation whereas they would, in all probability be separated by other atoms.

where a is a function of the distance coordinate alone and equals $r/(r_1^* + r_2^*)$ where r is the interatomic distance and r_1^* and r_2^* are the van der Waals' radii of atoms 1 & 2; ε is the energy coordinate which varies with the size of the atoms, since it is more difficult to push two large atoms to a small separation than two small atoms. ε for a pair of dissimilar atoms is sometimes taken as the geometric mean of the values for the two pairs of similar atoms.

In this treatment the expression is summed over all 1,4 and higher pairs of non-bonded atoms (i.e. those atoms separated by 3 or more bonds) since 1,3 interactions are taken care of by the angle bending expression. Other force-fields (such as the Urey-Bradley) include 1,3 interactions in the non-bonded potential.

1.3 Force-Fields

All of the component potentials and their respective parameters, as described above, are combined to form a force-field (FF). The type of force-field described here is a Valence Force-Field (VFF). Another type commonly used in molecular mechanics calculations and conformational analysis is the Urey-Bradley FF (UBFF) [18]. The main difference between the two lies in the treatment of the non-bonded interactions - the analytical forms of the potentials are the same but the UBFF considers 1,3 and higher interactions rather than 1,4 and higher as in VFF's. The UBFF also has linear, as well as quadratic terms in $(l - l_0)$, $(\theta - \theta_0)$ etc.

As previously mentioned bilinear cross-terms, such as (9) are included in some force-fields [18,19,20].

$$V_{l\theta} = \sum_l \sum_\theta k_{l\theta}(l - l_0)(\theta - \theta_0), \qquad (9)$$

In these instances they appear to have more effect on calculated vibrational frequencies than on geometric or thermodynamic properties [7].

One other type of 'cross-term' should be mentioned here.   Hydrogen
bonding plays an important role in determining the geometry and energy of
peptides e.g.   as in Figure 1.4.
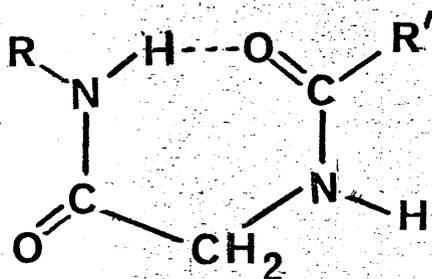
Fig. 1.4

It has been found [21] that a separate function [22] is not necessary — a
combination of of coulombic (5) and non-bonded (8) potentials suffices to
simulate H-bonding.

1.4 Derivation (or Paramaterization) of Force-fields

Using the various expressions described in section 1.2 it is possible
to calculate the steric energy of a particular molecular geometry.   It
remains, however, to assign values for the various force constants (k_{ℓ}, k_{θ},
ε etc.) and for the reference (strain-free) lengths, angles etc. which
will accurately predict a variety of physical parameters (geometric,
spectroscopic and thermodynamic).   In order to achieve this it is necessary
that there exists a large bank of experimental data on as wide a range of
compounds as possible.   The precision of the force-field is dependent on
the accuracy of the experimental data — 'good' data is therefore highly
desirable.

Initial estimates for force constants are normally obtained from
vibrational   analyses while equilibrium bond lengths and angles are
"borrowed" from microwave spectroscopy and X-ray or electron diffraction
results.   In valence angle deformation, for example, it has been shown
[18,23] that an accurate prediction of experimental energies and geometries
requires different equilibrium angles for different substitution at the

central atom.

Having obtained an estimated set of force constants it is necessary to optimise these with respect to the experimental data. Two methods of achieving this are in common use (a) trial and error [6] and (b) least squares calculations [18,23].

In method (a) the various parameters are altered on a trial and error basis in order to attempt to locate the optimal values. This process is clearly a lengthy and time-consuming business. Parameterization of an alkane/alkene FF [10] by this method took approximately two years.

Method (b) involves minimising the squares of the differences between experimental and calculated properties with respect to the force-field parameters. This method is known as the CONSISTENT force-field (CFF) method. The CFF method produces reliable force-fields giving good agreement with experimental data, the alkane/alkene CFF of Ermer and Lifson [20] produces mean absolute differences between observed and calculated bond lengths of 0.003Å, bond angles of 0.5° and torsion angles 1.0°).

Method (b), however, is no less time-consuming than method (a), nor is the accuracy of CFF's significantly better than FF's produced by method (a) [10].

1.5 Local Minimisation of Potential Energy

The starting structure used in molecular mechanics may not be (in general IS not) a very close approximation to the equilibrium (or strain-free) structure. It may have been derived from molecular models (see Chapter 3) or from diffraction studies on related compounds. It is therefore necessary to minimise the steric energy of the molecule with respect to it's coordinates (either internal or cartesian). A method of systematically adjusting the coordinates in order to reach the minimum energy is required. It is usually simpler and more convenient to operate on cartesian coordinates rather than on internal coordinates (except in the

case of very large molecules [8]).

There exist a number of algorithms [25] for performing function minimisation and some of these have been adapted for use in molecular mechanics calculations [2].Most of these procedures are adequate where the number of variables is small (< 5) whereas molecular mechanics calculations typically involve between 40 and 200 variables. Variants and adaptations of the Newton-Raphson minimisation algorithm have proved to be virtually the only useful and reliable type [8,26]. The subsequent discussion will be confined to Newton-Raphson methods applied to cartesian coordinates (x_i) as utilised in the present work. At a minimum (or maximum as the unwary find to their cost) of potential energy the first partial derivatives of potential energy with respect to each coordinate must be zero i.e. the net force acting on any atom must be zero – each atom is in its equilibrium position.

$$\frac{\delta V_s(x)}{\delta x_i} = 0 \qquad \begin{array}{l} i = 1,2,3,\ldots.3N \\ N \text{ is the number of atoms} \quad (10a) \\ \text{and } x \text{ is a coordinate.} \end{array}$$

or alternatively:

$$V_s'(x) = 0 \qquad (10b)$$

However for the initial, trial set of coordinates x this will not be the case and if the minimum is represented by x0 then:

$$x0 = x + \delta x,$$

where δx is the vector of corrections requiring to be applied to the trial coordinates x in order to reach x0. Equation (10b) then becomes:

$$V_s'(x + \delta x) = 0 \qquad (10c)$$

which by Taylor's expansion becomes:

$$V_s'(x + \delta x) = V_s'(x) + V_s''(x)\delta x + \ldots\ldots = 0. \qquad (10d)$$

This expansion, for computational simplicity is usually truncated after the linear term in x, so that equation (10d) reduces to:

$$V_s'(x) + V_s''(x)\delta x = 0, \qquad (10e)$$

where due to the truncation of the Taylor series δx is now an approximation to the previous δx. Equation (10e) now yields δx:

$$\delta x = -V_s''(x)^+ V_s'(x), \qquad (11)$$

or in a more general form:

$$\delta x = -BV_s''(x)^+ V_s'(x), \qquad (12)$$

where B is a scalar step length and $V_s''(x)^+$ is the generalised inverse (matrix) of $V_s''(x)$.

Equation (12) cannot be solved directly to give x0 due to the truncation of the Taylor series but must be iterated to convergence. The general Newton-Raphson iteration ((12) with B=1) is divergent for trial structures far from the minimum — this being the case for almost any generated structure. The trial structure must therefore be pre-optimised to a reasonable approximation to x0. One method of doing this uses (12) with B < 1. This makes use of the fact that for poor trial structures the calculated shift, δx, is in the correct direction but the magnitude is over-estimated with B = 1. The speed of convergence of (12) is reduced when B < 1 but poor approximations to x0 will eventually converge. Notice that this method involves a time-consuming calculation of $V_s''(x)$ for every iteration.

A second pre-optimisation algorithm is that of steepest descents in which $V_s''(x)$ is replaced by the identity matrix I. A proper choice of B for each iteration [27] leads to convergence of even the crudest approximation [7]. Steepest descents will only reduce the components of $V_s'(x)$ to around 0.1 kcal/mole/Å since it converges at that point. A few iterations (2 or 3) of (12) with B = 1 are then required for full convergence (usually in the region of 10^{-7} kcal/mole/Å is taken as a close approximation to $V_s'(x) = 0$).

The final approach [32] replaces the matrix of second derivatives $V_s''(x)$ by a matrix F such that:

$$F = \frac{\delta^2 V_s(x)}{\delta x_i \delta x_j},$$

where i and j are $\leqslant$ 3, for each atom so that (12) becomes:

$$\delta x = -BF^{-1} V_s'(x), \qquad (13)$$

where F^{-1} is the inverse of F. This block diagonal method overcomes the scaling problems inherent in steepest descents, since B is set to 1, but lacks its tolerance of very crude structures. Convergence is rapid to about 0.05 kcal/mole/Å, and thereafter slow. Full convergence is best achieved by using (12) with B = 1 (the Full Matrix (FM) method) as opposed to continuance of (13), the Block Diagonal (BD) method.

It has recently been found by White [8] and noticed in the course of this work that a further simplification of the BD method, the Pure Diagonal (PD), is as good as, if not better than the BD. The PD method involves replacing the matrix F (in (13)) by a matrix G such that

$$G = \frac{\delta^2 V_s(x)}{\delta x_i^2}$$

so that (12) becomes:

$$\delta x = -V_s'(x)/V_s''(x). \qquad (14)$$

The PD method is far more tolerant of poor approximations to x0 than the BD and converges just as well.

Expression (12) makes use of a 'generalised inverse' matrix and methods of calculating it are discussed elsewhere [8,28,29].

1.6 Calculation of Derivatives

Whatever method is used in minimising the potential energy one is required to calculate at least the first, and often the second derivatives of energy with respect to coordinates. Since the calculation of these derivatives is responsible for some 60 – 80% of the total computer time involved in the energy minimisation process it is essential to consider

carefully the method by which they are calculated. Two fundamental options are available (a) analytical or (b) numerical calculation of the derivatives. Analytical derivatives are difficult to program but are faster than numerical derivatives which, in turn are readily programmable. Analytical calculation requires extensive reprogramming in the event of a change in the functional form of any potential; numerical calculation of the derivatives is, of course, largely independent of the functional form of the force-field.

In general, therefore, it is best to use analytical derivatives with an established force-field where only production work is undertaken and to use numerical derivatives when development of a force-field is in progress. However with careful programming and the use of appropriate approximations it is possible to reduce the time taken to evaluate numerical derivatives so that they become preferred in all circumstances. The numerical calculation of first derivatives by central differences is a simple matter:

$$f'(x_i) = [f(x_i + \delta x) - f(x_i - \delta x)]/2\delta x, \qquad (15)$$

and the numerical evaluation of $f'(x_i)$ by evaluating $f(x_i + \delta x)$ and $f(x_i - \delta x)$ takes twice as long as the analytical calculation of $f'(x_i)$ since the time required to evaluate $f(x)$ and $f'(x)$ is comparable. Similarly the second derivatives could be calculated from:

$$f''(x_i,x_j) = [f(x_i + \delta x,x_j + \delta x) - f(x_i - \delta x,x_j + \delta x) -$$
$$f(x_i + \delta x,x_j - \delta x) + f(x_i - \delta x,x_j - \delta x)]/4\delta x . \qquad (16)$$

However numerical calculation of (16) requires four function evaluations as compared to one for the analytical form. Programs would, therefore, run four times slower if this form was used. By using appropriate approximations to $f''(x_i,x_j)$ [30], however, a more economical algorithm can be developed.

The first derivatives are evaluated using (15) and the second derivatives from (18) or (19). Let

$$f'(x_i) = [f(x_i + \delta x) - f(x_i)]/\delta x; \ (x_i + \delta x <- x_i), \qquad (17)$$

then

$$f''(x_i, x_i) = [f(x_i + \delta x) + f(x_i - \delta x) - 2f(x_i)]/\delta x , \qquad (18)$$

and

$$f''(x_i, x_j) = [f(x_i + \delta x, x_j + \delta x) - f(x_i, x_j + \delta x) -$$
$$f(x_i + \delta x, x_j) + f(x_i, x_j)]/\delta x ; \quad (x_j + \delta x \leftarrow x_j) \qquad (19)$$

This means that all of the terms in (18) and all but $f(x_i + \delta x, x_j + \delta x)$ in (19) are known from the calculation of the first derivatives. Only one function evaluation is required per second derivative — with the consequent saving in time.

1.7 Summary

The preceding sections have shown that no one method of minimising the potential energy is capable of going from the crudest initial structure to the minimum of energy. The exceptions to this are the Variable Metric Methods [33] which involve generating successivly better approximations to $V_s''(x)$ as minimisation proceeds. It is in general necessary to combine the FM procedure with one of steepest descent or diagonal (BD or PD) methods. The present work uses a two-stage Newton-Raphson with a combination of FM ((12) with B =1) and the PD (14) methods; derivatives are calculated numerically.

1.8 References

1. D.H.R. Barton, Experientia, 6, 316 (1950).

2. J.D. Williams, P.J. Stang, and P. von R. Schleyer, Annu. Rev. Phys. Chem., 19, 531 (1968).

3. E.M. Engler, J.D. Andose and P. von R. Schleyer, J. Amer. Chem. Soc., 95, 8005 (1973).

4. C. Altona and D.H. Faber, Topics in Current Chemistry, 45, 1 (1974).

5. J.D. Dunitz and H.B. Burgi in International Review of Science, Physical Chemistry Series Two, Butterworths, London, Vol. II, Ch. 4 (1976).

6. N.L. Allinger, Adv. Phys. Org. Chem., 13, 1 (1976).

7. O. Ermer, Structure and Bonding, 27, 161 (1976).

8. D.N.J. White, Computers & Chemistry, 1, 225 (1977).

9. O. Ermer, Tetrahedron, 30, 3103 (1974).

10. D.N.J. White and M.J. Bovill, J. Chem. Soc., Perk. II (1977), 1610.

11. J.P. Lowe, Science, 179, 527 (1973) compares the various postulates.

12. A. Warshel, M. Levitt and S. Lifson, J. Mol. Spect., 33, 84 (1970).

13. H. Margenau, Phys. Rev., 63, 385 (1943); 64, 131 (1943); 66, 303 (1944).

14. L. Pauling, The Nature of the Chemical Bond, Cornell University Press (1940).

15. I. Amdur and E.A. Mason, J. Chem. Phys., 22, 670 (1954).

16. G.C. Maitland and C.B. Smith, Chem. Soc. Rev., 2, 181 (1973) reviews this topic.

17. D.E. Williams, J. Chem. Phys., 45, 3770 (1966).

D.E. Williams, ibid., 47, 4680 (1967).

18. E.J. Jacob, H.B. Thompson and L.S. Bartell, ibid., 47, 531 (1967).

19. N.L. Allinger, M.T. Tribble, M.A. Miller and D.H. Wertz, J. Amer. Chem. Soc., 93, 1637 (1971).

20. O. Ermer and S. Lifson, ibid., 95, 4121 (1973).

21. A.T. Hagler, E. Huler and S. Lifson, ibid., 96, 5319 (1974).

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

23. N.L. Allinger, J.A. Hirsch, M.A. Miller, I. Tyminski and F.A. van Catledge, J. Amer. Chem. Soc., 90, 1199 (1968).

24. S. Lifson and A. Warshel, J. Chem. Phys., 49, 5116 (1968).

25. W. Murray (ed.), Numerical Methods for Unconstrained Optimization, Academic Press, London (1972).

26. D.N.J. White and O. Ermer, Chem. Phys. Lett., 31, 111 (1975).

27. M. Bixon and S. Lifson, Tetrahedron, 23, 769 (1967).

28. A. Warshel and S. Lifson, J. Chem. Phys., 53, 582 (1970).

29. C.R. Rao, Sankhya, 15, 253 (1955).

30. W.A. Busing, Acta. Cryst., A28, S252 (1972).

31. W. Klyne and V. Prelog, Experientia, 16, 521 (1960)

32. D. N. J. White and G. A. Sim, Tetrahedron, 29, 3933 (1973)

33. M. J. Hopper, Harwell Subroutine Library, HMSO Document AERE-R7477

(1973)

CHAPTER TWO

Global Minimum Energy
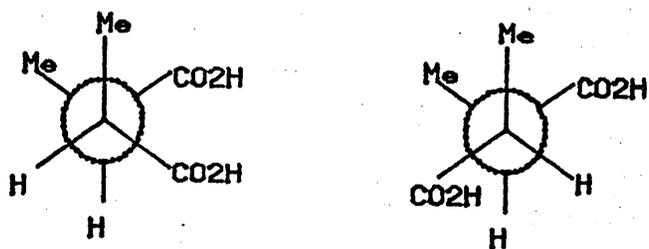
(I)
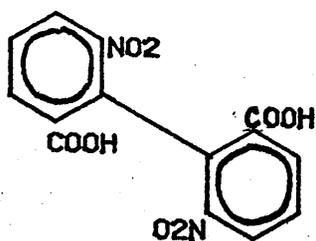
(II)

2.1 Introduction

The preceding chapter has shown how it is possible to calculate precise geometries corresponding to local minima of energy. In most cases these minima are discrete conformers of the compound under study. It was van't Hoff who proposed that rotation around a single bond was unrestricted but that rotation about a double bond was hindered. He thought that it should be possible, simply by rotation around one or more single bonds, to obtain an infinite number of stereoisomers i.e. conformers of any compound. The first indications that rotation around single bonds was not free were provided by Bischoff in 1890 and 1891. In those years he published papers [1] which proposed that the preferred conformation of butane was the staggered conformation and that restricted rotation occurred in multiple-substituted ethanes such as the isomeric disubstituted succinic acids (I).

The first experimental proof that rotation about single bonds was restricted came in 1922 when the structure of 2,2'-dinitrophenyl-6,6'-dicarboxylic acid (II) was resolved into optically active forms [2]. From 1930 onwards evidence rapidly accumulated that restricted rotation was very widespread. In ethane, for example, it was found that the observed and calculated entropies were different [3]. The most sensible explanation was that a barrier to free rotation existed although quite how it arose was not, at the time, clear.

Concurrently with the work on aliphatic molecules considerable thought was being applied to cyclic systems. Baeyer [4] suggested that all the atoms in a ring should be coplanar and therefore the strain in a ring could be deduced from the difference between polygonal angles involved in achieving this and the tetrahedral angle. According to this 'strain theory' the strain should decrease from cyclopropane to a minimum at cyclopentane and then increase indefinitely. This was not, however, in accord with experimental heats of combustion. Sachse [5] resolved these differences by returning to the tetrahedral atoms of van't Hoff and

produced, for example, the chair and boat forms of cyclohexane. Sachse also realised that these puckered rings gave rise to the possibility of non-equivalent substitution i.e. that axial and equatorial substituents were possible.

These ideas of Sachse were largely ignored until 1918 when Mohr [6] showed that the chair and boat forms of cyclohexane were readily interconvertible and could not be separated thus explaining the lack of evidence for two forms of the molecule. Hassel [7] provided the final proof when his electron-diffraction work proved that cyclohexane existed in the chair form.

A variety of conformers (local minima of energy) can exist for most molecules - which then is the most stable (the global minimum energy) conformation?

2.2 Rotation About One Bond

The simplest organic molecule which can have conformers is ethane. Kemp and Pitzer [3] suggested that the barrier to free rotation in ethane was about 3 kcal/mole. This assumption enabled the reconciling of experimental and calculated (from statistical mechanics) heat capacities and entropies. The variation of potential energy as one of the methyl groups rotates is shown in figure 2.1 in which the eclipsed form is shown, for clarity, as having a torsion angle of about 10 instead of 0 degrees.

The molecule of ethane will adopt one of the stable staggered conformations in preference to the eclipsed forms. In the case of ethane the minima (and maxima) are identical and therefore no one minimum is preferred over any other. However substitution of methyl groups for two of the hydrogens to give butane radically alters the situation. There ought still to be three minima and three maxima but they should now be distinguishable in terms of energy, as is illustrated by figure 2.2.

There are now two different types of minima and maxima. The lowest of
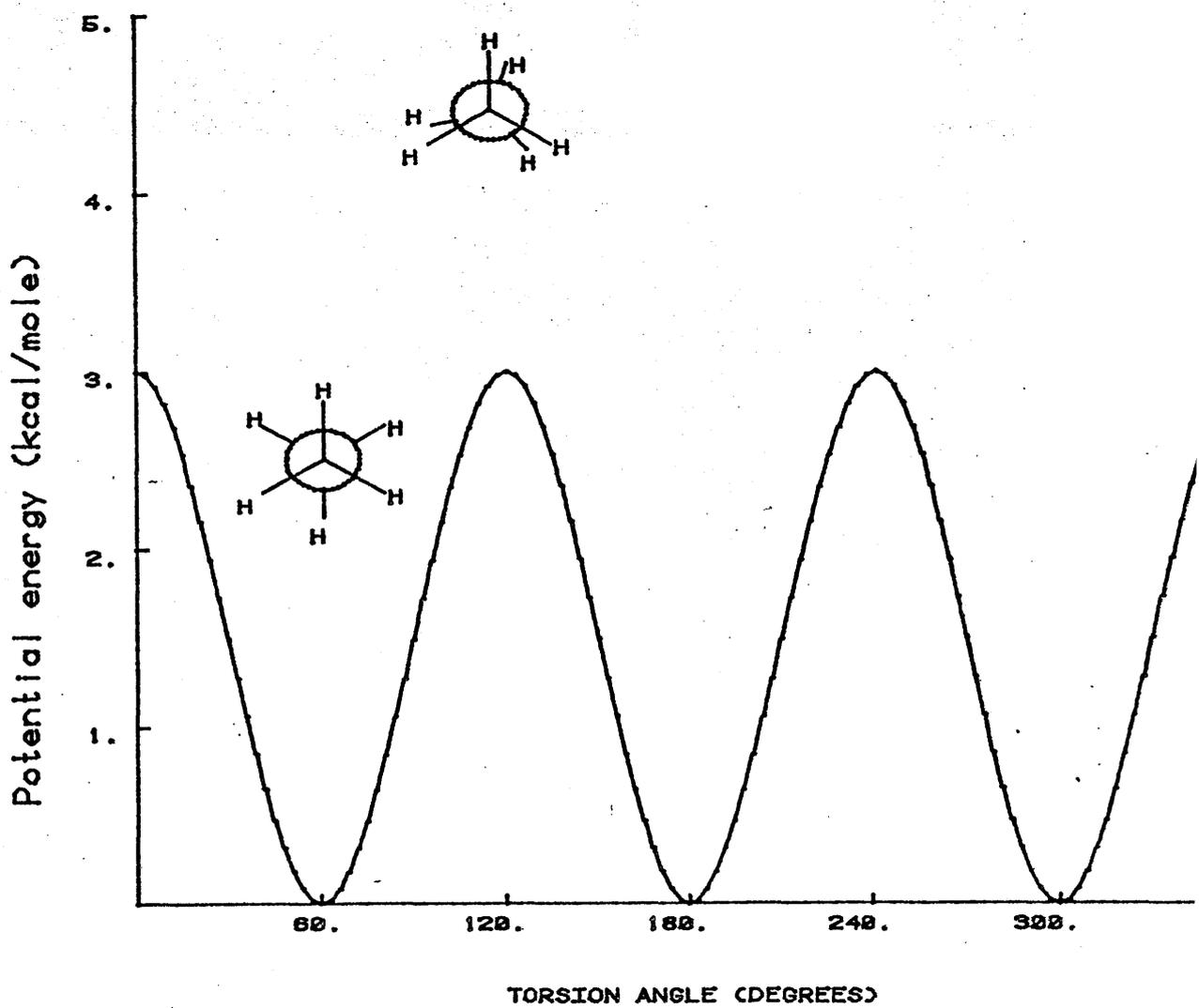
Figure 2.1 Rotation about one bond.
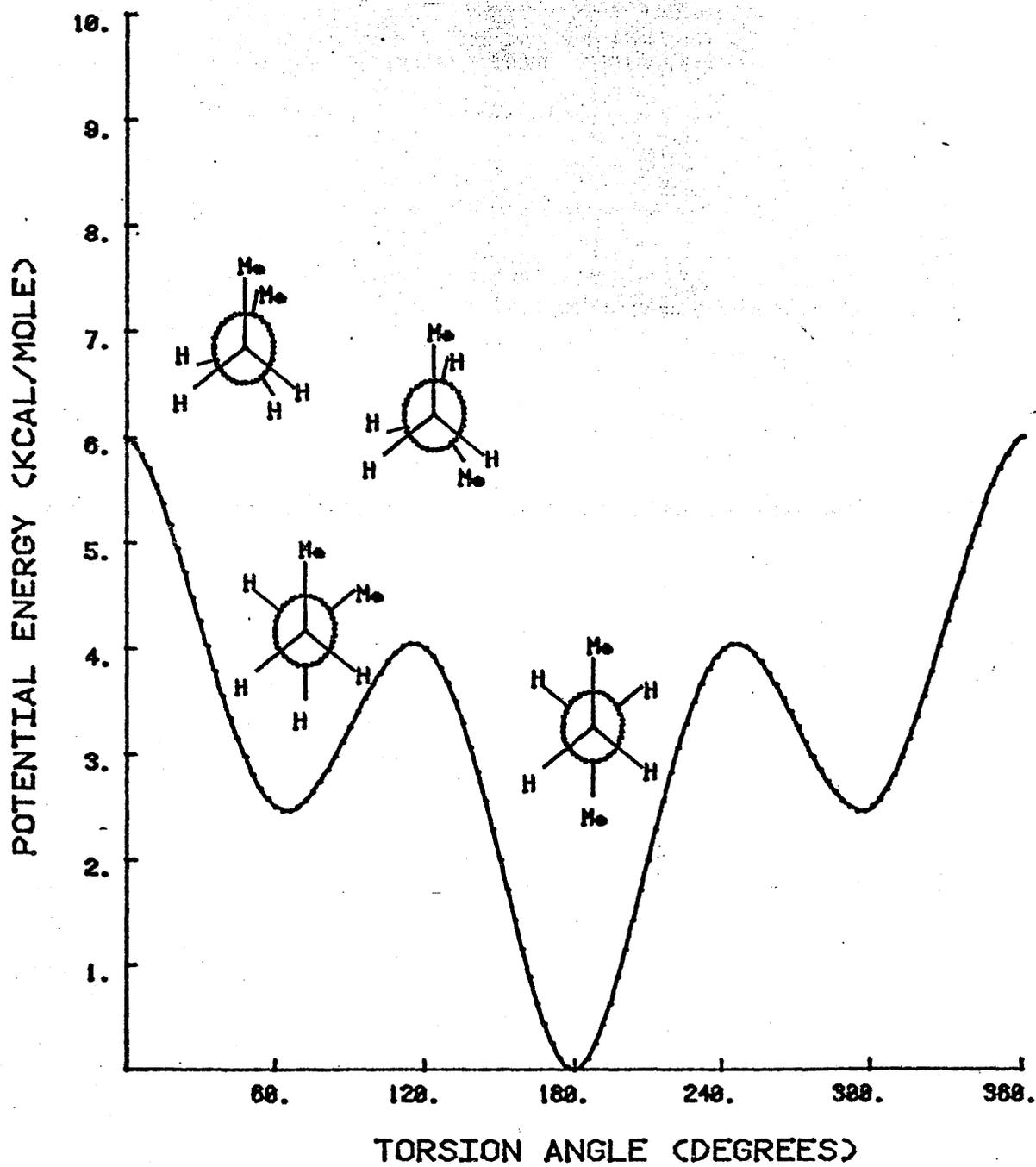
Figure 2.2 Butane Conformations.

the minima corresponds to an anti conformation (both methyls as far apart as possible) while the others represent enantiomeric gauche conformations. The highest maxima correspond to eclipsing of both methyl groups while the other two represent eclipsing of a methyl by hydrogen [8].

In both figures 2.1 and 2.2 the curve represents the potential energy at a given rotation of the central torsion angle. It is important to realise that this assumes that the other internal coordinates (bond lengths and angles) are fixed and that no other torsion angles (in the case of butane) are changing. This is a major restriction the ramifications of which are discussed later.

2.3 Rotation About Many Bonds

In most practical situations more than one torsion angle will be free to vary in such a way as to affect the potential energy. When two torsion angles are varying it is still a relatively easy matter to calculate and represent all the allowed conformers. In this instance the energy information is best displayed as a Ramachandran plot [9] and this method is extensively utilised [10,11,12]. Examples of this type of plot are used later. It is an easy matter to locate the areas of low energy from these plots and therefore the global minimum.

The drawing of the Ramachandran plots described above requires that the potential energy be calculated at a reasonable number of angle pairs if meaningful information is to be obtained. At 30 degree intervals, for example, $12^2 = 144$ evaluations are required. When more than two torsion angles are varying calculation of the potential energy hypersurface requires 12^n evaluations (where n is the number of torsion angles) and some way of displaying the n-dimensional information obtained. In view of these difficulties interest has tended to focus on cyclic molecules.

It is possible to treat chain molecules by statistical mechanics methods [36,37] although in most cases the results are not of sufficient biological interest to warrant much attention. For example long chain

hydrocarbons exist in random sequences of gauche and anti butane units. The exceptions to this are protein chains and work has been done on various types [19,20,21,22].

Ring closure constrains the values which torsion angles can have and it is possible, in simple cases, to predict the geometry of the various allowed conformers from molecular (e.g. Dreiding) models. In these cases manipulation of the models is sufficient to locate most, if not all of the allowed conformations. It should be emphasised that this applies only to small rings since larger rings involve too many variable torsion angles.

Attempts to use this method of selecting suitable-looking conformers on more complicated systems have been made [13,14,15], for example cyclodeca-1,6-diene [13,14] where eleven conformations have been located from Dreiding models and subsequently minimised. It is fairly certain that this list is incomplete although the calculated global minimum energy conformation does correspond to the geometry observed by electron diffraction [16].

It can be readily appreciated that a more systematic method of locating the global minimum energy conformer is needed and methods of doing this have been produced. These are, however restricted to two classes of system, namely cyclic oligopeptides [17,18,29] and globular proteins [19,20,21,22]. The subsequent sections of this chapter describe these methods.

2.4 Cyclic Oligopeptides

There exist two approaches to locating the global minimum energy conformation of cyclic peptides. In the first a hypothetical linear oligopeptide chain is constructed from appropriate amino-acid residues which have standard Pauling-Corey geometry [23]. If it is assumed that bond lengths and angles are fixed then a set of simultaneous equations can be formulated the solutions to which give a range of values for the torsion angles (φ_1,ψ_1), (φ_2,ψ_2)....(φ_n,ψ_n) which describe conformations of the

linear peptide which give rise to exact closure of the ring [29,30]. The mathematical basis of this is summarised in the appendix to this chapter. The amide groups are assumed to be planar and either cis ($\omega \simeq 0$) or trans ($\omega = 180$) as is appropriate. The IUB-IUPAC rules for nomenclature [24] are used here and the convention for naming torsion angles is shown in figure 2.3.
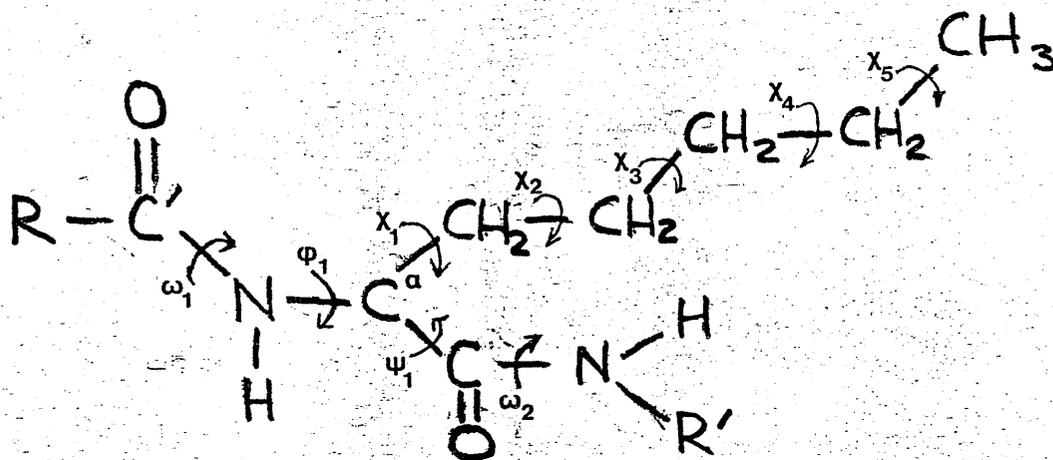
Fig. 2.3

A peptide of 5 or 6 residues will give rise to a considerable number of solutions to these simultaneous equations some of which can be eliminated by applying various energetic and geometric criteria. Firstly Ramachandran maps [9] of potential energy as a function of the two torsion angles φ and ψ are calculated for the N-acetyl-N'-methyl amides of the amino acid residues present in the peptide. Areas of low energy ("allowed regions of stereochemical space") can be located. The sets of torsion angles (φ_i, ψ_i), $i = 1,n$ obtained as solutions to the simultaneous equations are then individually compared with the low energy values for the appropriate amide. Those solutions in which one or more of the residues lies outside the appropriate range are eliminated.

Further solutions can be eliminated if only symmetric conformations are regarded as acceptable [25]. Having reduced the number of conformers to a manageable level they can then be optimised in one of the ways described above (this is necessary since ring closure does not necessarily give rise to a minimum of energy).

The above scheme was used in the investigation of the symmetric minimum energy conformations of cyclohexaglycyl [26] and some two dozen conformers were located. The major drawbacks to this method are the insistence on standard Pauling-Corey geometry for each residue and the requirement that conformers be symmetric. In effect a large proportion of the possible conformers are never considered since "real" peptides do not adhere strictly to the Pauling-Corey rules, especially in the smaller (4 or 5 residue) rings. In cyclotetraglycyl, for example, the minimum energy conformation is an all-trans configuration with S4 symmetry [17,27,28] and this requires far from standard geometry to effect ring closure. This conformation would not, therefore, have been located by this method.

The second method of locating the global minimum energy conformation of cyclic peptides uses the low energy areas of Ramachandran plots to generate possible conformations rather than eliminate impossible ones. Variations of up to 20 degrees are allowed in α-carbon bond, and all torsion, angles to achieve ring closure — there is therefore no insistence on artificially fixed geometry. This method has successfully located an S4 conformation with a non-planar (by 17 degrees) trans amide configuration as the global minimum of cyclotetraglycyl [17]. It was developed during the course of this work and is fully described in Chapter four.

2.5 Globular Proteins

The methods of energy minimisation described in Chapter one can not be used on proteins because the efficiency of the Newton-Raphson iteration is greatly reduced for problems exceeding 70 to 80 atoms. Steepest descents or pure diagonal methods would still work since these effectively treat each atom in isolation. Steepest descents has been used on proteins [31] although it converges on a point too far from the true minimum for real confidence in its results and even this convergence took a prohibitive amount of computer time for a relatively small (around 500 atom) protein. Some way of reducing the dimensionality of the problem needs to be used before molecular mechanics calculations can successfully be used on

proteins.

Several methods are available for doing this and they include: (a) fixing the Hydrogen atoms rigidly to the C, N or O atoms and only allowing them to move in concert with the parent atom; (b) keeping all bond lengths and angles fixed and only optimising the torsion angles [32]. When using this method it is normal to apply (a) as well and, in addition, to fix the ω torsion angles at 180 leaving only φ, ψ and χ to be optimised. In some treatments the χ are assumed to have standard values as well as the ω. Finally (c) treating the amide residues as fixed units represented by van der Waal's envelopes of appropriate shape (spheres, ellipsoids etc.). A force-field is then devoloped in terms of the lengths, angles, torsion angles etc between the envelopes [33]. This last method seems to be the most successful and has been used on bovine pancreatic trypsin inhibitor [33].

It is possible using the procedures outlined in the preceding paragraph to locate local minima for proteins and two methods of finding the global minimum have been developed. The first [33] allows convergence on the nearest local minimum whereupon thermal motion in the protein is simulated. This has the effect of jolting the protein out of the local minimum into another potential well which has (hopefully) greater depth. Repetition of this process should lead to the global minimum.

The second method [21] is purely empirical. It uses the fact that the probability of finding a given amino-acid sequence in a particular structural feature (pleated sheet, β-bend, α-helix etc.) can be determined from a statistical analysis of all protein structures determined by X-ray crystallography. These probabilities can then be used to predict the gross structure of a new protein which can then be optimised. The coordinate set derived from the predicted structural features should be close to, or indeed in the potential well corresponding to the global minimum. This method has been used to predict two possible conformations of glucagon [34].

2.6 Appendix — Scheraga's Method of Simultaneous Equations

This appendix contains a summary of the mathematical basis of the method of deriving symmetric cyclic peptides described in section 2.4. As an example a molecule with C3 symmetry has been chosen.

Note that throughout this appendix matrices are represented by underlined upper case characters (e.g. $\underline{A}$) while vectors are represented by underlined lower case letters (e.g. $\underline{a}$) all other variables are scalars.

C3 symmetry requires that a rotation of $2\pi/3$ about the symmetry axis leads to the same structure, the following is therefore true:

$$\tau_{i + m} = \tau_i \qquad\qquad i = 1, 2m \qquad\qquad (1)$$

where m is the number of torsion angles in a symmetry unit.
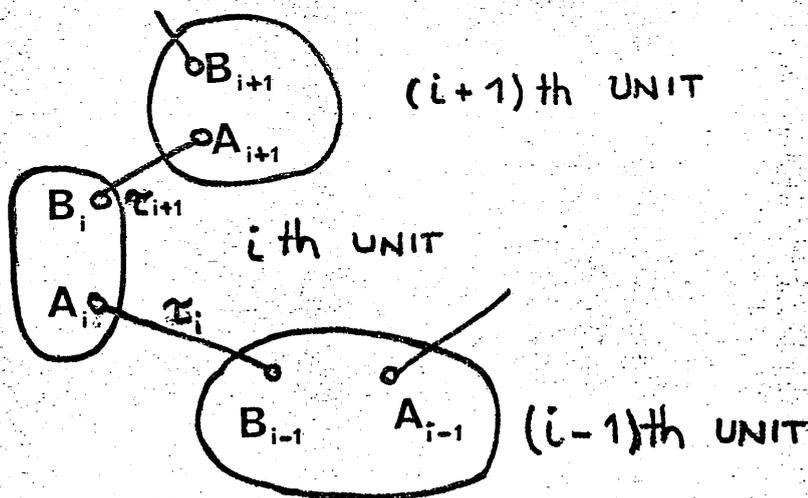
Fig. 2.4

If F_i is the (i)th unit as defined in figure 2.4 consisting of rigidly arranged atoms such as A_i and B_i then a molecule of formula (IV) can adopt C3 symmetry.

$$[(-F_1 F_2 ---F_m-)_3]$$
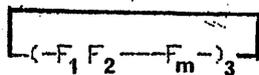
(IV)

A local cartesian coordinate system is now defined for each rigid structure with origin at B_{i-1} for unit i. The x-axis is pointed towards A_i and the y and z axes are specified by some suitable atom in the (i)th unit. If a given point in space is defined by position vectors $r(i)$ and $r(i-1)$ with respect to the (i)th and (i-1)th coordinate systems then

$$r(i-1) = T(i-1)R(i)r(i) + p(i-1) \qquad (2)$$

where $T(i-1)$ is a 3X3 orthogonal rotation matrix to bring the (i)th system into the same orientation as the (i-1)th when $\gamma_i = 0$; $R(i)$ is a matrix for a rotation of γ_i about the x-axis (the $B_{i-1} \rightarrow A_i$ vector) given by:

$$R(i) = R(\gamma_i) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma_i & -\sin\gamma_i \\ 0 & \sin\gamma_i & \cos\gamma_i \end{bmatrix} \qquad (3)$$

and $p(i-1)$ is the position vector of B_{i-1} w.r.t. the (i-1)th coordinate system. By repetitive use of (2) we can now get the position vector of any point in space w.r.t. any coordinate system if we know it's position vector w.r.t. any one system. Notice that the (3m)th coordinate system can be regarded as the (0)th system due to the cyclic nature of the molecule and this convention will be used hereafter.

To transform from the (m)th coordinate system to the (0)th requires:

$$r(0) = Ur(m) + p \qquad (4)$$

where

$$U = T(0)R(1)T(1)R(2) \ldots \ldots T(m-1)R(m) \qquad (5)$$

and

$$p = p(0) + T(0)R(1)p(1) + T(0)R(1)T(1)R(2)p(2) + \ldots \ldots$$

$$\ldots\ldots + T(0)R(1)T(1)R(2)p(2)\ldots\ldots T(m-2)R(m-1)p(m-1) \qquad (6)$$

U brings the (m)th system into the same orientation as the (0)th and p is the position vector of the origin of the (m)th system w.r.t. the (0)th.

In a molecule with C3 symmetry a rotation of $2\pi/3$ about the symmetry axis will cause the (0)th and (m)th systems to overlap (i.e. will have the same effect as (4)) and therefore U and p must be related in a specific way.

Two new coordinate systems are now defined the (0')th and the (m')th having the same orientation as the (0)th and (m)th respectively and each having origins at the same point on the symmetry axis. If r(0') and r(m') are the position vectors of a point w.r.t. the (0')th and (m')th coordinate systems then

$$r(0') = r(0) + q \qquad (7)$$

$$r(m') = r(m) + q \qquad (8)$$

and

$$r(0') = U[u, \pm 2\pi/3] r(m') \qquad (9)$$

where q is the position vector of the origin of the (0)th (or (m)th) coordinate system w.r.t. the (0')th (or (m')th) and u is a unit vector (with direction cosines u_x, u_y and u_z in the (0)th system) lying along the symmetry axis.

If $\phi = \pm 2\pi/3$ then [35]:

$$U(u,\phi) = \begin{bmatrix} u_x^2(1-\cos\phi)+\cos\phi & u_xu_y(1-\cos\phi)-u_z\sin\phi & u_xu_z(1-\cos\phi)+u_y\sin\phi \\ u_yu_x(1-\cos\phi)+u_z\sin\phi & u_y^2(1-\cos\phi)+\cos\phi & u_yu_z(1-\cos\phi)-u_x\sin\phi \\ u_zu_x(1-\cos\phi)-u_y\sin\phi & u_zu_y(1-\cos\phi)+u_x\sin\phi & u_z^2(1-\cos\phi)+\cos\phi \end{bmatrix}$$

$$(10)$$

From (7), (8) and (9) we obtain

$$r(0) = U[u,+2\pi/3]r(m) + (U[u,2\pi/3]-I)q \qquad (11)$$

where I is the unit matrix. From (4) and (11) by comparison:

$$U = U[u,\pm2\pi/3] \qquad (12)$$

and

$$p=(U[u,2\pi/3]-I)q \qquad (13)$$

Thus U and p in (7) and (8) must, in a molecule with C3 symmetry, be given by (12) and (13). From (10), (12) and (13) we have

$$TrU = u = 1 + 2\cos(2\pi/3) \qquad (14)$$

and from (13)

$$U.p = 0 \qquad (15)$$

i.e. p must be perpendicular to the symmetry axis.

Starting from (14) and (15) we can now solve for the dependent variables τ_{m-1} and τ_m in terms of the independent variables τ_1, τ_2, ...τ_{m-2}. In order to emphasise that τ_{m-1} and τ_m are unknowns they shall be represented as ς and η respectively and similarly R_{m-1} and R_m will be replaced by X and Y. From (5)

$$U = AXBY \qquad (16)$$

and from (6)

$$p = a + AXb \qquad (17)$$

where

$$A = T(0)R(1)T(1)R(2).....T(m-3)R(m-2)T(m-2)$$

$$B = T(m-1)$$

$$a = p(0)+T(0)R(1)p(1)+....+T(0)R(1)T(1)R(2)..T(m-3)R(m-2)p(m-2)$$

$$b = p(m-1)$$

From (10)

$$\begin{pmatrix} u_{23} - u_{32} \\ u_{31} - u_{13} \\ u_{12} - u_{21} \end{pmatrix} = -2\sin\phi u \qquad (18)$$

and from (18), (15) becomes

$$p(1)(u_{23} - u_{32}) + p(2)(u_{31} - u_{13}) + p(3)(u_{12} - u_{21}) = 0 \qquad (19)$$

If we now introduce an additional matrix C such that C = AXB, U now equals CY. The matrix C and the vector p are functions only of the independent variables and of ONE unknown $\varsigma = \gamma_{m-1}$. (14) and (19) can now be written as:

$$\alpha'\cos\eta + \beta'\sin\eta + \gamma' = 0 \qquad (20)$$

and

$$\alpha''\cos\eta + \beta''\sin\eta + \gamma'' = 0 \qquad (21)$$

where

$$\alpha' = C(22) + C(33)$$

$$\beta' = C(23) - C(32) \qquad (22)$$

$$\gamma' = C(11) - 1 - 2\cos 2\pi/3$$

and

$$\alpha'' = p(1)(C(23) - C(32)) - p(2)C(13) + p(3)C(12)$$

$$\beta'' = -p(1)(C(22) + C(33)) + p(2)C(12) + p(3)C(13) \qquad (23)$$

$$\gamma'' = p(2)C(31) - p(3)C(21)$$

where $C(ij)$ is the (i,j)th element of C.

Solving (20) and (21) simultaneously gives

$$\cos\eta = (\beta'\gamma'' - \gamma'\beta'')/(\alpha'\beta'' - \beta'\alpha'')$$
$$\sin\eta = (\gamma'\alpha'' - \alpha'\gamma'')/(\alpha'\beta'' - \beta'\alpha'') \tag{24}$$

by squaring and adding each of (24) η can be eliminated and

$$(\beta'\gamma'' - \gamma'\beta'')^2 + (\alpha'\beta'' - \beta'\alpha'')^2$$
$$- (\alpha'\beta'' - \beta'\alpha'')^2 = 0 \tag{25}$$

which is an algebraic expression in a single unknown ς which can be solved numerically. Having solved for ς then η can be obtained from (24).

This method is not much use for larger than hexapeptides since for cyclohexaglycyl (if we assume fixed cis or trans amides) there are 12/3 variable torsion angles for C3 symmetry. Solving the above equations reduces the dimensionality from 4 (= 12/3) to 2 (for Ci or C2 the dimensionality is 3 which is still manageable). However for cyclononaglycyl the problem is four-dimensional even after solving the simultaneous equations and thus still intractable. The dimensionality continues to increase as the number of residues increases.

2.7 References

1. C.A. Bischoff, Chem. Ber, 23, 620 (1890); 24, 1074 (1891); 24, 1086 (1891); 26, 1452 (1891).

2. G.H. Christie and J. Kenner, J. Chem. Soc., LXXI, 614 (1922).

3. J.D. Kemp and K.S. Pitzer, J. Chem. Phys., 4, 749 (1936); J. Amer. Chem. Soc., 59, 276 (1937).

4. A. Baeyer, Chem. Ber., 18, 2269 (1885).

5. H. Sachse, Chem. Ber., 23, 1363 (1890); Z. Physik Chem. (Leipzig), 10, 203 (1892).

6. E. Mohr, J. Prakt. Chem., [2]98, 315 (1918); Chem. Ber., 55, 230 (1922).

7. O. Hassel, Tidsskr. Kjemi Bergresen Met., 3, 32 (1943); English translation in Topics in Stereochemistry, 6, 11 (1971).

8. For further discussion of ethane and butane see W.J. Orville-Thomas (ed) Internal Rotation in Molecules, Chap. 1, Wiley-Interscience, London, 1972.

9. G.N. Ramachandran and V. Sasiskharan, Adv. Prot. Chem., 23, 284 (1968).

10. P.N. Lewis, F.A. Momany and H.A. Scheraga, Isr. J. Chem, 11, 121 (1973).

11. J.F. Yan, F.A. Momany, R. Hoffmann and H.A. Scheraga, J. Phys. Chem., 74, 420 (1969).

12. D.N.J. White, Tetrahedron Letters, (1975) 2101.

13. D.N.J. White and M.J. Bovill, J. Chem. Soc., Perk. II, (1977) 1610.

14. D.N.J. White and M.J. Bovill, Tetrahedron Letters, (1975) 2239.

15. J. Dale, Acta. Chem. Scand., 27, 1115 (1973).

16. L. Hedberg and K. Hedberg, Abs. Papers, Nat. Meeting Amer. Cryst. Ass., 1964, Bozeman, Montana.

17. D.N.J. White and C. Morrow, Tetrahedron Letters, (1977) 3385.

18. N. Go and H.A. Scheraga, Macromolecules, 3, 188 (1970).

19. L. Levitt and A. Warshel, Nature (London), 253, 694, (1975).

20. S. Tanaka and H.A. Scheraga, Proc. Nat. Acad. Sci. U.S.A., 72, 3802 (1975).

21. P.Y. Chou and G.D. Fasman, Biochemistry, 13, 222 (1974).

22. I.D. Kuntz, G.M. Crippen, P.A. Kollmann and D. Kimelmann, J. Mol. Biol., 106, 983 (1976).

23. L. Pauling, The Nature of the Chemical Bond, p.282, Cornell, New York, 1960, (3rd. ed.).

24. Biochemistry, 9, 3471 (1970).

25. N. Go and H.A. Scheraga, Macromolecules, 6, 273 (1973).

26. N. Go and H.A. Scheraga, ibid, 6, 525 (1973).

27. C. Grathwol, A. Tun-kyi, A. Bundi, R. Schwyzer and K. Wuthrich, Helv. Chim. Acta., 58, 415 (1975).

28. J. Dale and K. Titlestad, J. Chem. Soc., Chem. Comms. (1972) 255.

29. N. GO and H.A. Scheraga, Macromolecules, 6, 178 (1973).

30. N. Go and H.A. Scheraga, ibid, 6, 283 (1973).

31. M. Levitt and S. Lifson, J. Mol. Biol, 46, 269 (1969).

32. H.A. Scheraga, Chem. Rev., 71, 195 (1971).

33. M. Levitt, J. Mol. Biol., 104, 59 (1976).

34. P.Y. Chou and G.D. Fasman, Biochemistry, 14, 2536 (1975).

35. H. Jeffreys and B. Jeffreys, Methods of Mathematical Physics, 3rd edition p.122, Cambridge University Press, Cambridge, 1956.

36. D.A. Brant and P.J. Flory, J. Amer. Chem. Soc., 87, 2792 (1965).

37. P.J. Flory, Statistical Mechanics of Chain Molecules, Interscience, New York, 1969.

CHAPTER THREE

Generation and Manipulation of Molecular Models

3.1 Introduction

In order to apply the molecular mechanics calculations described above it is necessary to derive a set of coordinates (internal or cartesian) corresponding to a reasonable approximation to the required geometry. This is unfortunately not a trivial problem for anything other than the simplest cases e.g. ethylene or benzene which are essentially only two-dimensional.

The most elementary method of generating a trial set of coordinates is by using 'standard' values for the internal coordinates (for instance the equilibrium bond lengths and angles of the corresponding force-field might be employed). Using simple three-dimensional geometry it is possible to derive cartesian coordinates from the internal coordinates. This process becomes very tedious and prone to error for large non-planar molecules.

Another simple method involves the construction of a mechanical (e.g. Dreiding) model of the structure and then measuring the coordinates either directly from the model or by by casting a shadow onto graph paper. In the latter case a second view (perpendicular to the first) must be measured to obtain the third coordinate. A more exotic version of this method (generally used for very large molecules) uses space-filling models which have metal indicators at the atom centres. X-ray photographs are taken of the finished model and the coordinates derived from these [1].

The most generally useful method involves generating the cartesian coordinates from internal coordinates using a computer. The following section describes an algorithm for doing this.

3.2 Cartesian Coordinates from Internal Coordinates

An algorithm for generating cartesian coordinates [2] is illustrated by figure 3.1.

Fig. 3.1
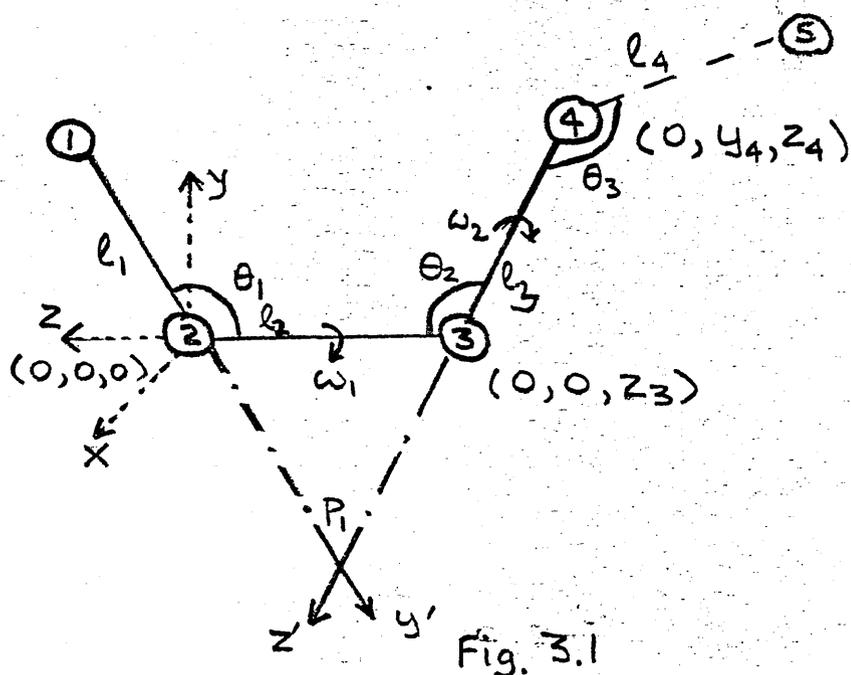
If 1, 2, 3 & 4 are assumed to be initially coplanar then their coordinates can be readily calculated from two-dimensional geometry. Atom 2 is taken as the origin and standard values used for the bond lengths (l_1, l_2 and l_3) and for the bond angles (θ_1 and θ_2). Atom 4 is then rotated around the 2-3 bond (the Z-axis) to give the required value of the torsion angle ω_1 by means of equation (1).

$$\begin{bmatrix} xn_4 \\ yn_4 \\ zn_4 \end{bmatrix} = \begin{bmatrix} \cos\omega_1 & -\sin\omega_1 & 0 \\ \sin\omega_1 & \cos\omega_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} xo_4 \\ yo_4 \\ zo_4 \end{bmatrix} \quad (1)$$

where xn, yn and zn are the new coordinates and xo, yo and zo are the old coordinates. The origin of coordinates is then shifted to atom 3 and a new orthogonal coordinate system defined in terms of the unit vectors $\underline{x}'$, $\underline{y}'$ and $\underline{z}'$ where $\underline{z}'$ is given by the direction cosines of the vector joining atoms 3 & 4; $\underline{y}'$ is given by the direction cosines of the vector connecting atom 2 and the point P; and $\underline{x}' = \underline{y}' \wedge \underline{z}'$. The partially constructed molecule is then rotated into the new coordinate system by:

$$\begin{bmatrix} xn(i) \\ yn(i) \\ zn(i) \end{bmatrix} = \begin{bmatrix} lx' & mx' & nx' \\ ly' & my' & ny' \\ lz' & mz' & nz' \end{bmatrix} \begin{bmatrix} xo(i) \\ yo(i) \\ zo(i) \end{bmatrix} \quad (2)$$

where i =1,2,...number of atoms; lx',ny' etc. are direction cosines; and xn,yo etc. have their previous significance. Atom 5 can now be added (initially coplanar with atoms 2, 3 & 4) and its coordinates found by use of equation (1) followed by equation (2). Further atoms, up to the required number, can be added and their coordinates found by repeating the process as above.

Numerous programs are available for carrying out these calculations but most suffer from the fact that the user cannot examine the structure until it is complete. Even when the structure has been completed it is difficult to tell from a list of coordinates (or even torsion angles) whether the correct conformer has been built. In addition these programs generally have very complex input formats in which atoms are referenced by their position in a list or by obscure code numbers. Great care, and often more than one attempt is required in the use of this type of program particularly when building cyclic molecules.

A method combining the advantages of letting a computer do the mathematics of transferring from internal to cartesian coordinates with the ability to examine the model as it is being constructed is the ideal solution to the problem of generating molecular models. Such a system was developed, and extensively used, in the course of the present work.

3.3 Chemical Graphics Systems

The use of computer-produced graphics to represent the structure of chemical species is logically divided into two distinct types of application: (a) static representation of single molecules or crystal packing and (b) diagrams which can be directly manipulated by the user.

The most widely used and most well-known program of the first type is the Oak Ridge Thermal Ellipsoid Plotting Program (ORTEP) [3] which utilises a pen plotter and in general runs in batch mode on a large mainframe computer. This program is limited to displaying already known structures (most commonly from X-ray experiments) although it does provide extensive

facilities for measuring the geometry of the structure. This program suffers from the complexity of the required input and the number of attempts required to produce a reasonable view of even a small molecule can be high.

The second type of program (commonly termed interactive) allows the user to communicate with and direct the operation of the computer. These programs require the use of a cathode-ray tube display since mechanical displays (such as pen plotters) take too long to draw the pictures. Interactive programs can be divided into two broad areas the first of which employs a storage tube display. In this type of program (e.g. ref. 4) a static picture is displayed and if it is required to alter this then the whole picture must be erased and redrawn in its updated form. This type of program can be regarded as an extension of the ORTEP-like programs since it essentially draws static pictures under more direct user control.

The second type of interactive system utilises display tubes which allow the selective alteration of parts of a picture without the necessity of redrawing the entire picture. Raster scan and refresh displays fall into this category and both have been used for real-time interactive systems (see for example ref. 5). The biggest advantage of these systems is that the user can see changes as they occur - for example in a static system it might require several attempts to obtain a satisfactory view of a molecule. In a dynamic system one simply watches as the view alters until a suitable one appears at which point the picture can be 'frozen'.

This section has described the differences between static and dynamic displays but one further consideration is important in chemistry (and in many other fields).

3.4 Representing Three-Dimensionality

A major difficulty in any method of drawing chemical structures is the fact that they are three-dimensional objects. Some method of representing the third dimension in the two dimensions available is required. Each of
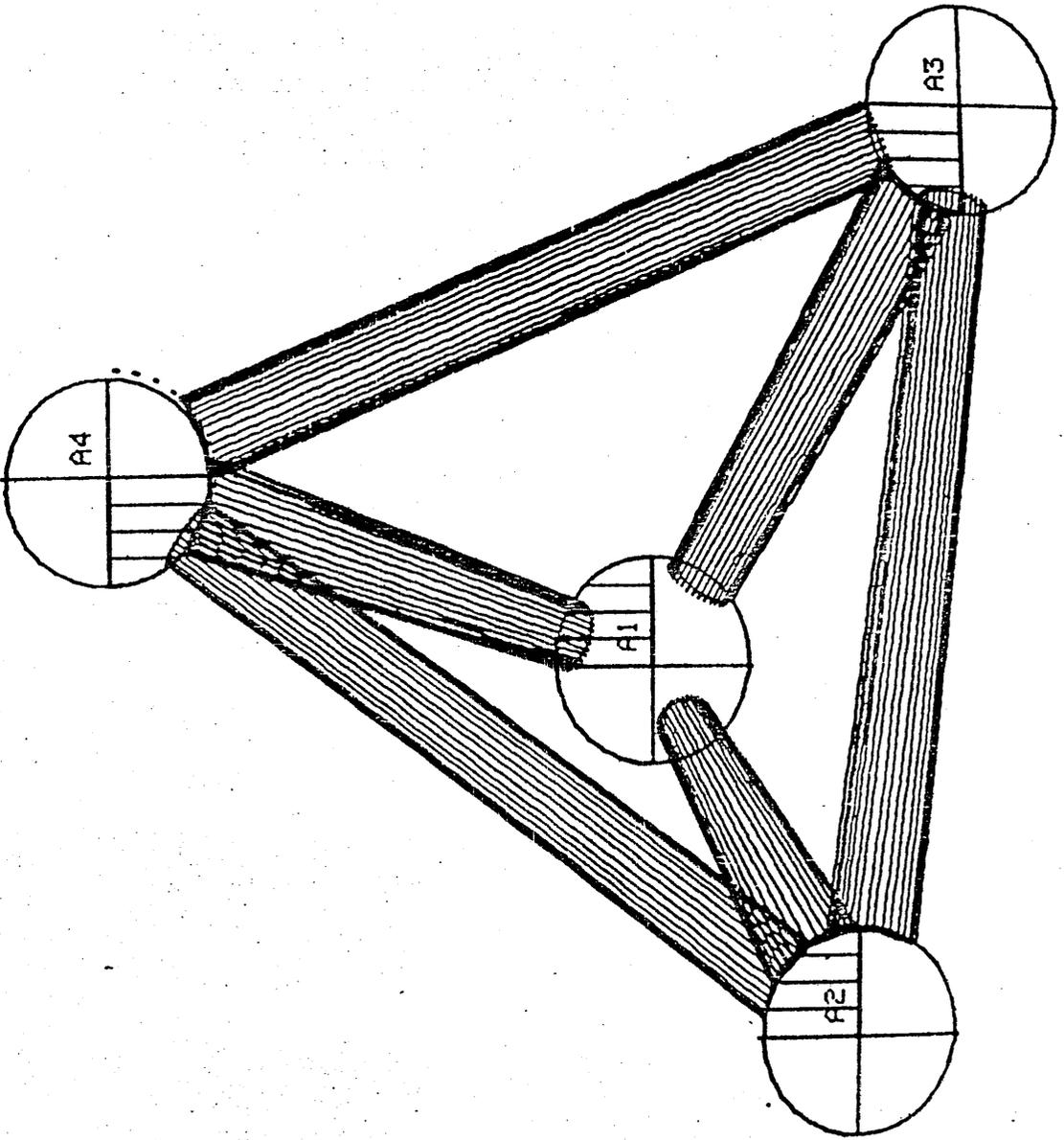
Figure 3.2 ORTEP Construction.

Figure 3.3 ORTEP Depth Anomaly.

the systems outlined in the preceding section is limited in some way by the hardware involved in producing the picture.

The ORTEP program uses tapered bonds to indicate depth, the end of the bond nearest to the viewer has a larger radius than the farther end. This construction is illustrated by figure 3.2. This is adequate for simple cases but becomes less so for more complicated structures. Figure 3.3 for example shows a conformation (Carbon atoms only) of cyclo-octa-1,5-diene in which it is easy to see the spatial relationship between bonded atoms but is C3 or C7 nearer the viewer? The situation is somewhat improved in a newer version of the program (ORTEP II) which eliminates those line segments which are hidden behind other lines. This is achieved at a considerable cost in computer resources (both in core requirements and execution time) limiting use of this version to mainframe computers.

A second method of displaying depth information available to static systems uses stereoscopic pairs of drawings. This involves drawing two views of the structure at approximately + and - 3 degrees to the required view and presenting, in some way, one only one view to each eye. The images should then overlap and give the appearance of three-dimensionality. In the ORTEP program this is achieved by drawing the views side-by-side (e.g. figure 3.4) and looking at them through a special viewer . A similar system has been developed for use on CRT displays where again the views are displayed side-by-side [6]. In this instance a more complicated viewer consisting of a system of lenses and prisms is required.

A more elegant version of this basic technique is to superimpose the stereo pair but draw each in a different colour. Meyer [7,8] has developed this method for use on a colour TV monitor; one view is drawn in red, the other in green and when the screen is observed through appropriately coloured filters (one for each eye) a yellow three-dimensional view results.

The major fault in each of these stereo pair methods is that they require some sort of optical device to achieve the effect of
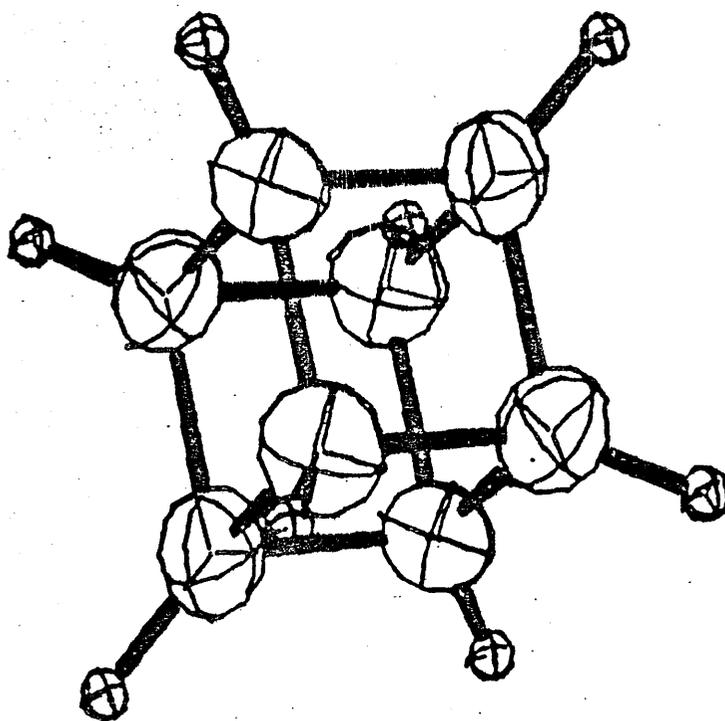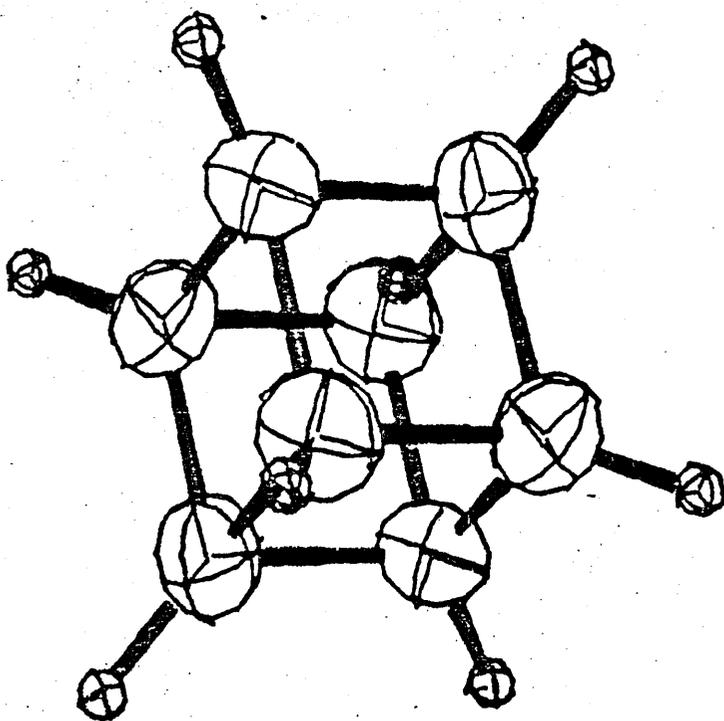
Figure 3.4 ORTEP Stereo Pair.

three-dimensionality which makes it difficult for more than one person to examine the structure at any one time. A further problem arises in that some people find it very difficult, if not impossible, to make the images overlap while extensive use is cumbersome and tiring.

The best static method of conveying depth information is an extension of the ORTEP system. The whole structure is 'depth-cued' in this method, not just each individual bond. This can only be done on a display which allows variable intensity drawing. Those parts of the structure closer to the viewer are drawn brighter than those further away as shown in figure 3.5. The intensity of any part of the display is proportional to the Z-coordinate (perpendicular to the plane of the screen, with X and Y in the plane of the screen) therefore front and back confusion does not arise. The obvious advantages of this system are that no optical aids are required and that depth assignments are unambiguous.

Perhaps the most elegant method of deriving depth information is only available on displays with a dynamic capability which is used to display a constantly rotating structure [9,10]. This is known as the 'kinetic depth effect' and the three-dimensional effect is obtained by the viewer from the fact that points with large |Z| coordinates (as defined above) appear to be moving faster than those with smaller |Z|.

Unfortunately this technique suffers from the Sinsteden illusion [11]. This arises from the fact that points of opposite sign of the Z-coordinate appear to move at the same speed — one cannot distinguish front from back. In order to overcome this defect it was necessary to use a combination of the 'kinetic depth' and 'depth-cueing' techniques in which the intensity of each point is varied as it is rotated producing a totally unambiguous picture of a structure.

To achieve this effect it is essential to use a refresh-type display since only they have the capabilities to perform both dynamic picture updating and have varying intensity display. This feature is embodied in the chemical graphics system developed in the course of the present work
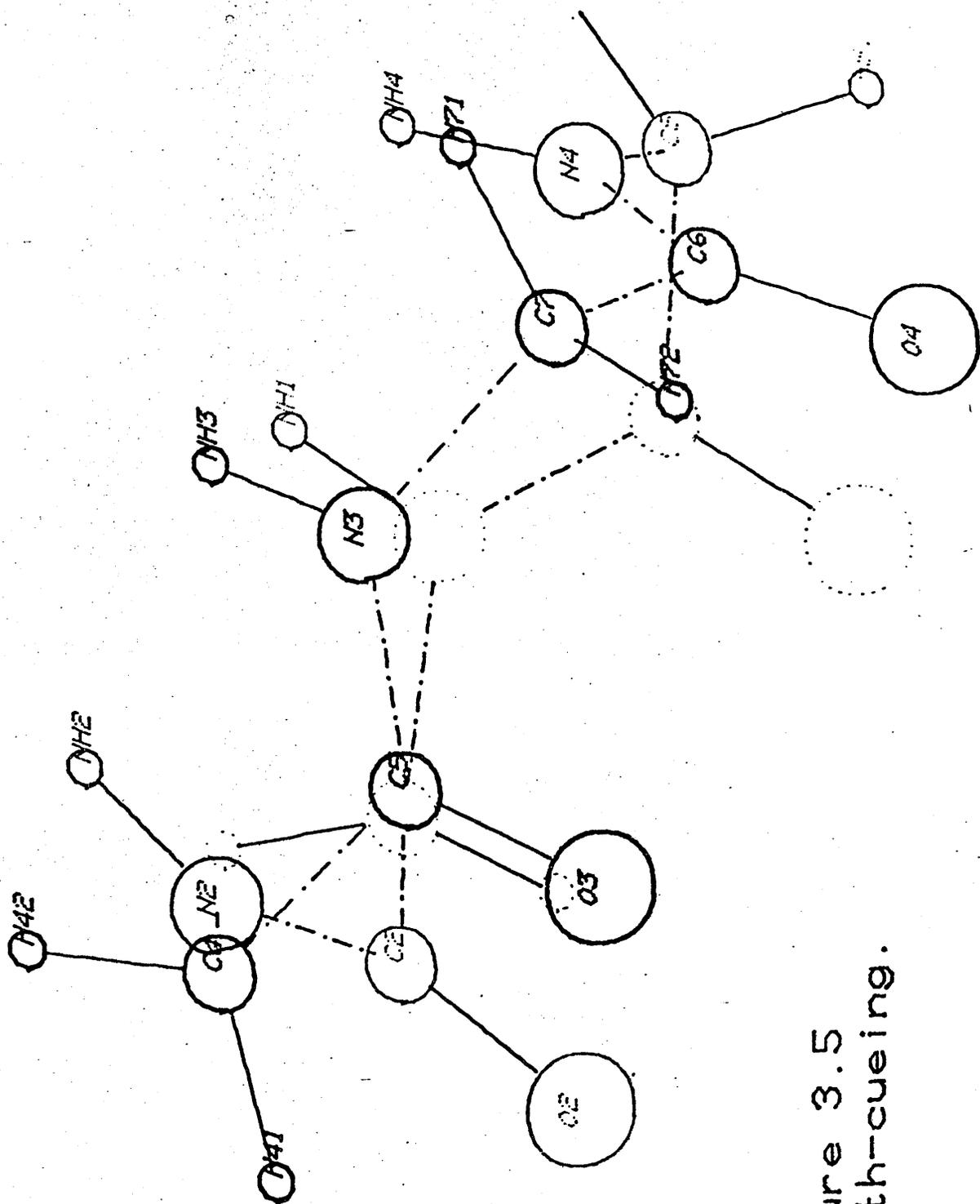
Figure 3.5
Depth-cueing.

and described in the following section.

3.5 A Description of the Graphics System Used

This section is logically divided into two parts: (i) hardware and (ii) software. 'Hardware' is a term used to encompass the physical parts of a computer system i.e. the electronics, switches, disc drives etc. which actually perform the various operations of the computer. Software, on the other hand, is the generic name for the instructions (i.e. programs) under whose control the hardware operates.

A detailed discussion of hardware is not appropriate to this thesis and therefore only a list of those items making up the facilities used will be given. This is most conveniently done in schematic form and figure 3.6 is a block diagram of the components of the system. As can be seen from the diagram the system is based on the ubiquitous PDP-11 central processor with the graphics capability provided by a VT-11 display processor with attached CRT display. This system provides a powerful general purpose computer with reasonable quality graphics at (relatively) modest cost.

The programs which now collectively form the Glasgow University Chemical Graphics System (GUCGS) were written during the course of the present work. They were written in FORTRAN to run under the Digital Equipment Co.'s RT-11 operating system. GUCGS is intended to allow the chemist (unaided by computer scientists or specialists) to build manipulate, and measure three-dimensional molecular models.

In order to facilitate the user interface with the system all dialogue with the various programs is in as plain English as possible while user input is preceded by an explanatory prompt. Erroneous input is recognised as such and the user invited to retype the line thus avoiding the usually fatal effects of bad typing. The various options available to the user are displayed in the form of a 'menu' and this is shown in figure 3.7 which also contains a view of a simple molecule. Options are selected by typing the capitalised letters in the option names which are arranged according to
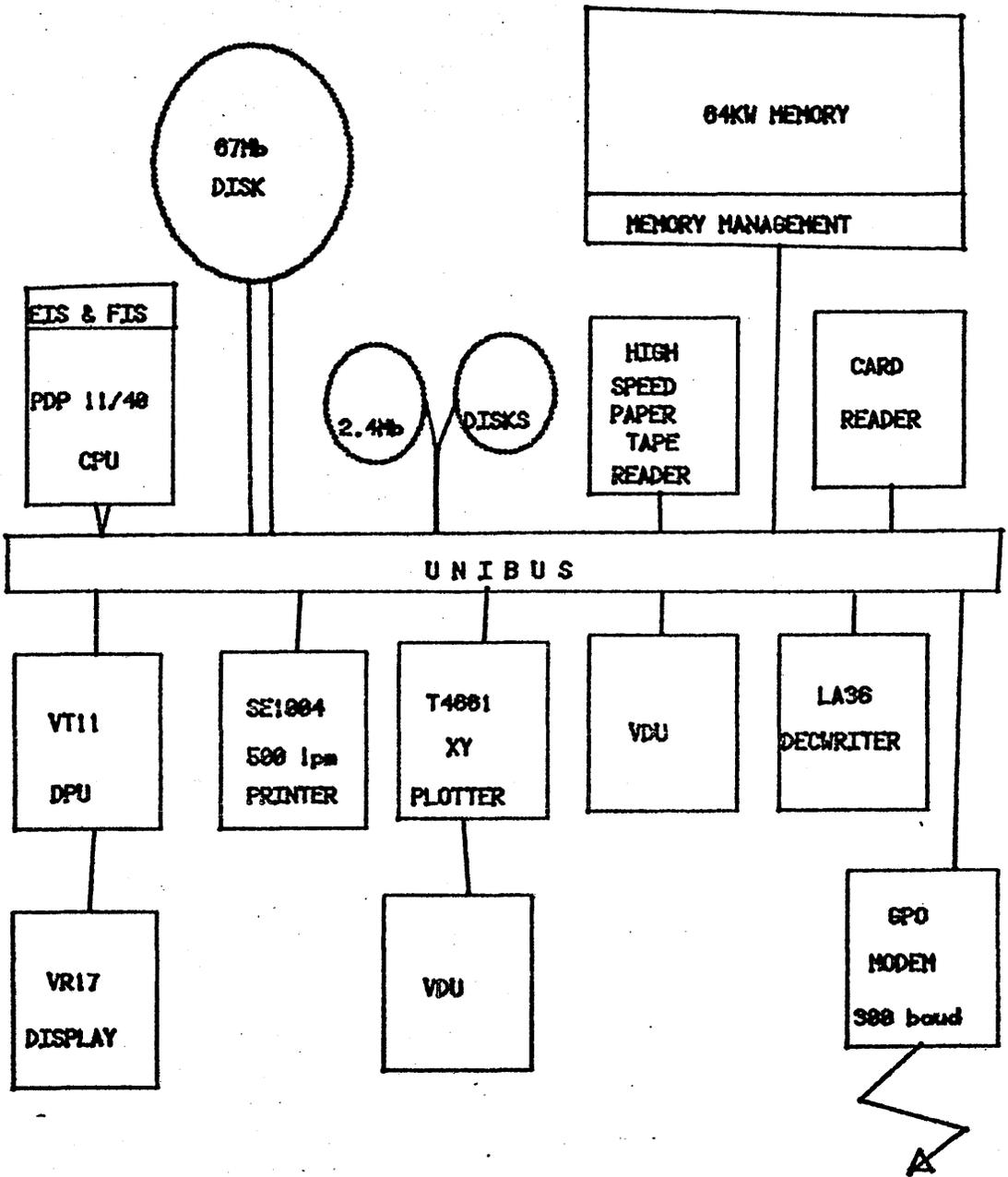
Figure 3.6 Computer Used.
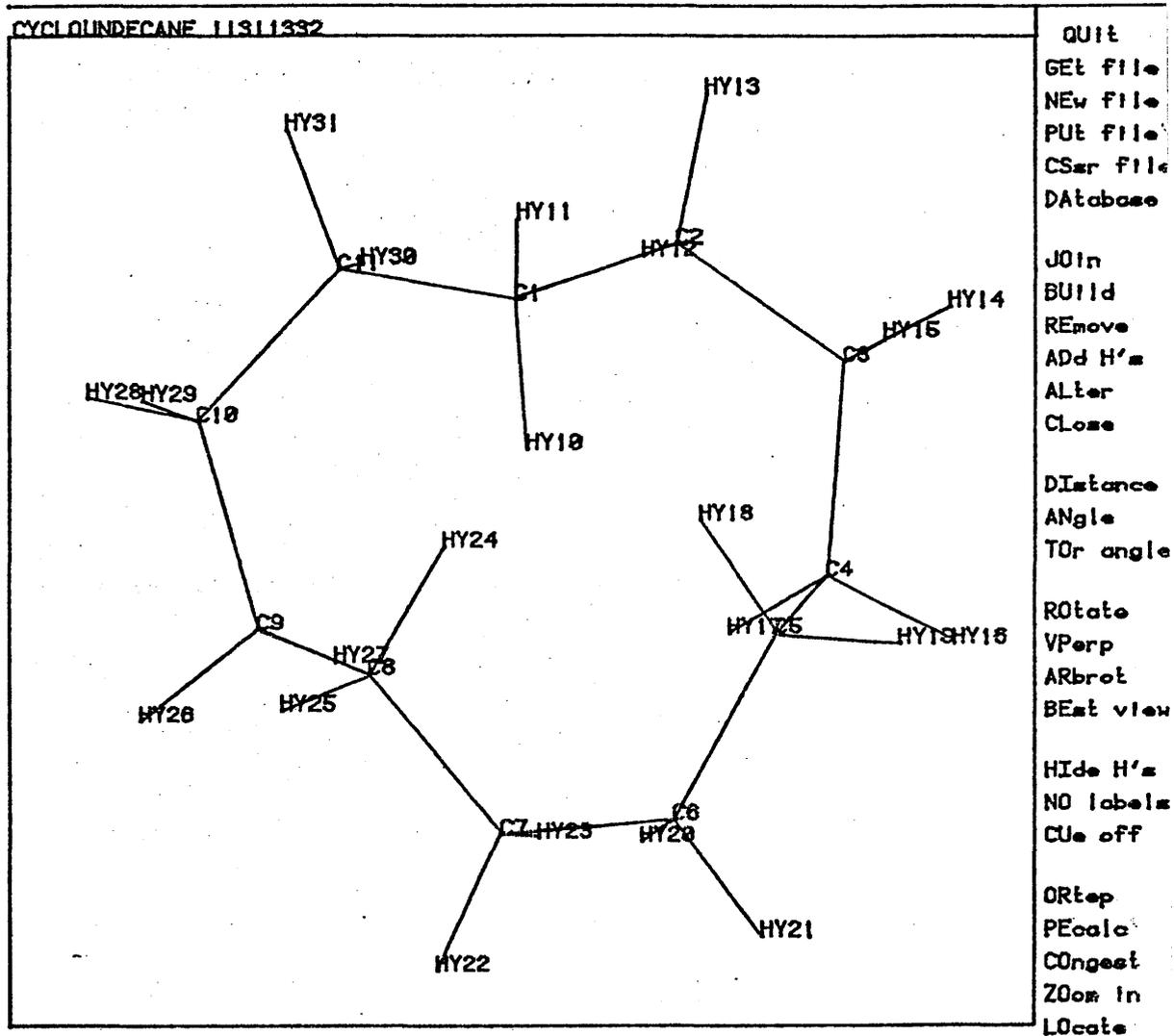
GLASGOW UNIVERSITY CHEMICAL GRAPHICS SYSTEM

FIGURE 3.7 Graphics System Menu.

function. The various options are described in the following subsections.

3.5.1 File Handling Options

This group of options allows the user to store and retrieve data in one of the standard formats.

GEt file fetches (GEts) the contents of a file which is in the standard system format. All details of the structure are obtained and a picture of the molecule displayed. Errors or omissions are reported to the user.

PUt file writes (PUts) a file containing all the information known to the system about the currently displayed structure in the format which can be read by GEt file. The logic and operation of the computer filestore are totally invisible to the user when using these commands – the user is merely aware that his data are stored somewhere.

NEw file allows the user to input the details of a structure from the console terminal. All the required input is prompted for and may be typed in free format. Errors are trapped and appropriate corrective action is taken or the user is requested to retype the line.

CSsr file reads files in the format used by the Crystal Structure Search & Retrieval (CSSR) system maintained by the Science Research Council. This allows access (via a computer in Edinburgh) to over 17,000 crystal structures.

DAtabase allows rapid retrieval of a structure from a locally-generated file. Two such files are currently maintained one of which contains the coordinates of all hydrocarbons studied by the molecular mechanics group at Glasgow. The other contains all of the cyclic tetrapeptide conformations generated during the present work.

3.5.2 Structure Alteration Options

This group of options allows the user to alter and manipulate the displayed structure. The various options are: JOin atoms together (i.e. 'form' bonds between them). Two modes are available (a) all atoms closer

than a specified distance are joined (except Hydrogens – H-H bonds are not allowed and the maximum X-H distance is 1.2Å); (b) two specified atoms are joined (or separated if already bonded).

BUild allows the user to construct a structure (or add to an existing one) by entering the internal coordinates of a new atom with respect to three existing atoms. Alternatively one can change the position (and connectivity) of an existing atom. The cartesian coordinates of the new atom are calculated using the algorithm described in Section 3.2.

REmove an atom from the structure. All trace of the specified atom is removed from the displayed structure.

ADd H's makes life easy for the user by building Hydrogen atoms onto medium atom backbones. This could be done by repeated use of the BUild option but would be tedious. The Hydrogen atoms are added according to medium atom type e.g. for methyl groups a torsion angle is requested and the Hydrogens disposed accordingly.

One can ALter a torsion angle without affecting any of the other internal coordinates (i.e. bond lengths and angles are maintained fixed). The alteration occurs in real time and the user selects the required setting of the torsion angle by striking any key on the console terminal. While the alteration is taking place counters display the current value of the torsion angle and also the distance between two specified atoms.

CLose performs a ring closing algorithm. This is a useful feature when building rings since one often finds that the ends of a ring are the wrong distance apart. CLose tries to rectify this by adjusting torsion angles by a maximum of |10| degrees, this small deviation ensures that the ring conformation is maintained but is sufficient in most instances.

3.5.3 Measurement Options

These three options enable the user to measure the various internal coordinates. Thus the DIstance separating any two atoms, the ANgle between any three atoms and the TOr(sion) ang(le) defined by any four atoms may be readily obtained.

A 'wild card' feature has been incorporated into these options such that replacing an atom name by an asterisk gets all reasonable atoms i.e. for angles specifying '*' instead of an atom name gets all bond angles.

3.5.4 View Alteration Options

These options allow the user to change the view of the structure without changing the structure itself in any way. The first two of these options are rotation functions which are performed in real time i.e. the rotation occurs dynamically and the desired view is 'frozen' by striking any key on the console device.

One can ROtate the structure about the X, Y or Z axis where the axial system is defined as in figure 3.8 with X and Y in the plane of the paper and +Z coming towards the viewer.
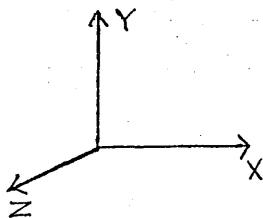
Fig. 3.8

The 'speed' of rotation can be chosen depending upon the accuracy with which a required view is needed. The molecule would first of all be set spinning fairly rapidly while a large number of possible views were examined; having selected a view the structure could be set to turn slowly while the exact orientation required was selected.

In addition to rotation about the coordinate axes, ARb(itrary) rot(ation) allows the user to specify a rotation axis defined by any two atoms. In this case the speed of rotation is fixed.

VPerp gets the V(iew) Perp(endicular) to a plane defined by three atoms. This feature is useful when examining trigonal systems. BEst view calculates and displays the view at right angles to the mean plane of the molecule. In general this corresponds to the view with least overlap of atoms and makes getting an uncluttered view of even large and complex

structures fairly straightforward.

3.5.5 Drawing Alteration Options
━━━━━ ━━━━━━━━ ━━━━━━━━━━ ━━━━━━━

These are essentially cosmetic in nature — being switches between alternative drawing modes. The three options are: (a) HIde/SHow H's which allows the hydrogen atoms to be temporarily removed from the drawing. This is useful in the case of a large or cluttered molecule since it can greatly simplify the drawing. The hydrogen atoms are still present and any rotations, for example, are performed on the hidden hydrogens so that they reappear in the correct positions when SHown. (b) Again in the case of a very crowded structure it may be desirable to omit the atom labelling. This is done by using NO labels. LAbels restores the atom names to view. (c) The displayed structure is depth-cued as described earlier (3.4,figure 3.5) and this feature can be switched on or off by using CUe on/off.

In the preceding three options only that facility which is actually available for use is shown as part of the 'menu'. Thus if the Hydrogen atoms are hidden then only SHow H's appears in the 'menu'.

3.5.6 Miscellaneous Options
━━━━━ ━━━━━━━━━━━━━ ━━━━━━━

ORtep produces an ORTEP diagram of the molecule in the current orientation. The required view is set up using, for example, ROtate and the ORTEP drawing produced either on the visual display unit or on a graph plotter. An added feature is that the program sets up a file containing standard ORTEP instructions which can be used to generate the drawing on any machine running ORTEP.

PEcalc is an implementation of the block diagonal potential energy minimisation routine described earlier (Chap. 1). A relatively crude structure can be generated and then used as input to the minimisation routine — the optimisation is performed and the refined structure displayed for further manipulation.

A ZOom feature has been included to facilitate close examination of particular areas of a molecule. One zooms onto a specified atom and all other atoms within a designated radius are displayed.

In a crowded structure it may be difficult to see a particular atom of interest. LOcate causes the required atom to blink on and off at the maximum brightness.

COngest can be used to predict the outcome of certain classes of stereoselective reactions. The algorithm of Wipke and Gund [12] is used to calculate the steric congestion above and below the plane of a trigonal atom or a double bond, the ratio of these congestions can be used to predict the relative amounts of the reaction products resulting from attack on the trigonal centre.

3.6 Conclusion

All of the operations described in the following chapters were carried out on the hardware described in Section 3.5 and represented by Figure 3.6 and used the programs making up the Chemical Graphics System also described in that section unless otherwise indicated. The FORTRAN coding involved in these programs runs to several hundred pages and is not, therefore, reproduced.

3.7 References

1. C. de Haen, E. Swanson and D.C. Teller, Biopolymers, 15, 1825 (1976).

2. D.N.J. White, Computers in Chemistry, 1, 225 (1977).

3. C.K. Johnston, Oak Ridge National Laboratory, Report ORNL-3794 (1965).

4. Program XRAY, R.J. Feldman, National Institutes of Health, Bethseda, Maryland, U.S.A.

5. M.D. Glick, T.J. Anderson, W.A. Butler, E.R. Corey and R.J. Srodawa, Computers in Chemistry, 1, 75 (1977).

6. C.D. Barry, R.A. Lewis, S. Graesser and G.R. Marshall, "Display and Manipulation in Three Dimensions" in "Pertinent Concepts in Computer Graphics" ed. by M. Faimau and J. Nievergelt, Univ. of Illinois Press, Urbana (1969).

7. E.F. Meyer, J. Appl. Cryst., 3, 392 (1970).

8. T.V. Willoughby, C.N. Morimoto, R.A. Sparks and E.F. Meyer, J. Appl. Cryst., 7, 430 (1974).

9. C. Levinthal, Scientific American, June 1965, p.42.

10. C. Levinthal, C.D. Barry, S.A. Ward and M. Zarick, "Computer Graphics and Macromolecular Chemistry" in "Merging Concepts in Computer Graphics" ed. by J. Nievergelt and D.W.A. Secrest, Benjamin, New York (1967).

11. see for example R.L. Gregory, Scientific American, November 1968, p.66.

12. W.T. Wipke and P.Gund, J. Amer. Chem. Soc., 98, 8107 (1976).

CHAPTER FOUR

An Algorithm to Locate the Potential Energy Minima

of Cyclic Molecules

4.1 Introduction

The concept of the Global Minimum Energy Conformation (GMEC) was introduced in Chapter 2 and methods for locating it were discussed in section 2.4. This chapter describes the development of the algorithm used in the location of the GMEC and other low energy conformers of cyclotetraglycyl. The methods evolved during the work on this compound have been subsequently refined into an automated single-program procedure.

The approach used here is an extension of a technique applied to Gramicidin S by de Santis and Liquori [1]. In this each of the amino acid residues was represented by the corresponding enantiomer of Alanine (i.e. Pro, Val, Orn and Leu were represented by L-Ala while the D-Phe residue was replaced by D-Ala). Preferred values of the (φ, ψ) torsion angle pairs could then be obtained from a single Ramachandran plot [2] of an Ala-Ala type dipeptide. The values for D-Ala were taken as the mirror images of the L-Ala set.

The Gramicidin S structure was then constructed by generating all the possible 2×4^5 permutations of the preferred torsional values. Only nine conformations resulted in an end-to-end distance of less than 10A (i.e. were approximately cyclic) and these were further examined for the presence of the known hydrogen bonding pattern. The two conformations containing this feature were subjected to potential energy minimisation, but only in (φ, ψ) space, in order to determine the global minimum.

4.2 Why Cyclic Peptides? [3]

Two criteria ought to be applied to any theoretical method of predicting the physical properties of molecules. Firstly, the scope of the problem must be such that the results should be generally useful; and secondly, the results, in the first instance at least, should be checkable by experimental methods. Molecular Mechanics meets both of these criteria in that molecules of a reasonable size (unlike ab-initio

calculations) can be dealt with and there are a multitude of experimental methods (X-ray and n.m.r. being prime examples) for the discovery of molecular structure.

Since resources of both human and computer time are not unlimited a rider must be applied to the first criterion – simply the problem must be tractable as well as interesting. The solution to this is "...to combine biological relevance with structural simplicity.." [3]. Cyclic peptides fulfil these conditions, especially if we choose to consider peptides composed entirely of glycine residues.

The large number of conformational states available to a linear peptide of even a few amino acid residues is substantially reduced by cyclisation. Thus the problem is reduced to manageable proportions without losing relevance since cyclic peptides are biologically active. They perform a variety of functions and it is known that their activity is intimately related to molecular conformation [4].

The small size of the glycine residue means that about 75% of the possible conformers of the dipeptide are sterically reasonable (see Figure 4.2) while for all other amino acids the set of available conformers is substantially reduced by the bulk of side chains and only about 25% of the total are reasonable [2]. It is therefore not unreasonable to suppose that the differences in ring conformation between cycloglycines and other cyclic peptides can be interpreted in terms of the side chains of the non-glycyl residues [5].

On the basis of these arguments it was decided to attempt the location of the Global Minimum Energy Conformation of cyclotetraglycyl, together with all of its low energy conformations. Cyclotetraglycyl is itself an interesting problem in that there is some doubt as to its conformation (this is discussed later). Further there are a number of cyclotetra-peptides and -depsipeptides whose structures have been determined.

The algorithm used in this work is shown as a block diagram in Figure 4.1 and each stage is described by the following sections.
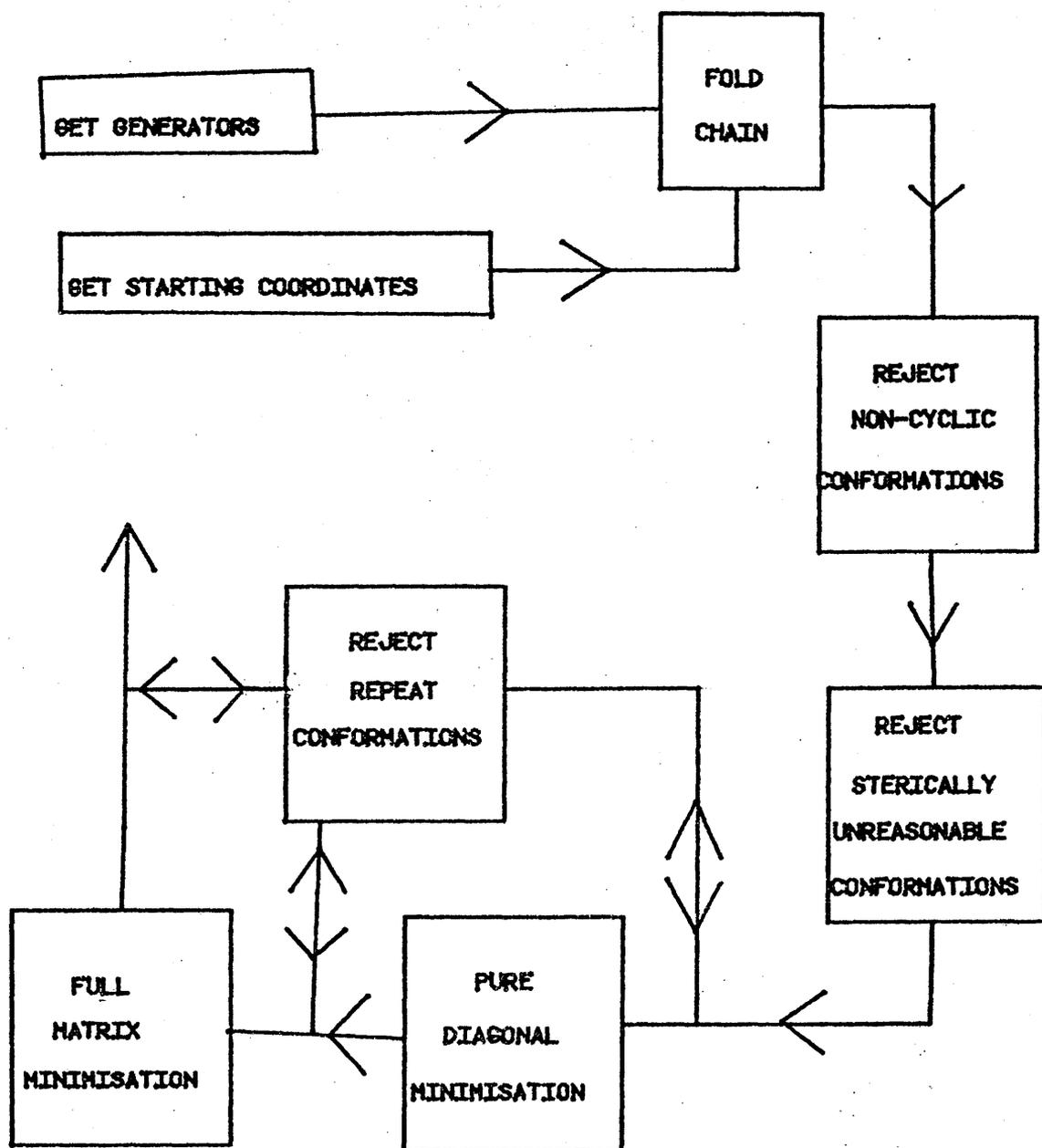
Figure 4.1 Block Diagram of Global

Minimisation Procedure.

4.3 Obtaining the Generators

In this context the term "Generators" is used to denote the pairs of (Φ, Ψ) torsion angles which are preferred by the amino acid residues making up the peptide under consideration. These preferred torsion angles can be obtained from a Ramachandran plot of the corresponding N-acetyl N'-methyl amide. Thus for cyclotetraglycyl only N-acetyl N'-methyl glycine (referred to hereafter as GlyGly) need be considered.

Rings composed of four amino acid residues are known to contain cis, as well as the more usual trans, amides. It is therefore necessary to calculate four Ramachandran plots — one each for t,t-GlyGly, c,t-GlyGly, t,c-GlyGly and c,c-GlyGly (where c = cis and t = trans). These maps are shown in Figures 4.2 — 4.5. In these diagrams the numbers represent the potential energy (in kcal/mole) above the minimum energy value. Values in excess of 7 kcal/mole above the minimum have been omitted thus the blank areas on the map represent regions of high potential energy. The energy was calculated at 10 degree intervals of Φ and Ψ.

These diagrams clearly show the large areas of conformational space available to the glycine residue. The calculated map for t,t-GlyGly is very similar to that obtained by Scheraga and coworkers [6] who located seven minima. In view of the size of the low energy areas and the slight differences between the four maps it was decided to use six arbitrary generator pairs which covered the available conformational space. This removes any possibility that the final results could have depended on the exact location of the minima used as generators which in turn may well be force-field or minimisation routine dependent. The chosen generators can be used to generate any of the cis/trans configurations whereas Scheraga's were only applicable to the all-trans configuration. The (Φ,Ψ) pairs (in degrees) used were: 1:(90,60), 2:(-120,60), 3:(120,180), 4:(-120,180), 5:(120,-60) and 6:(-90,-60).
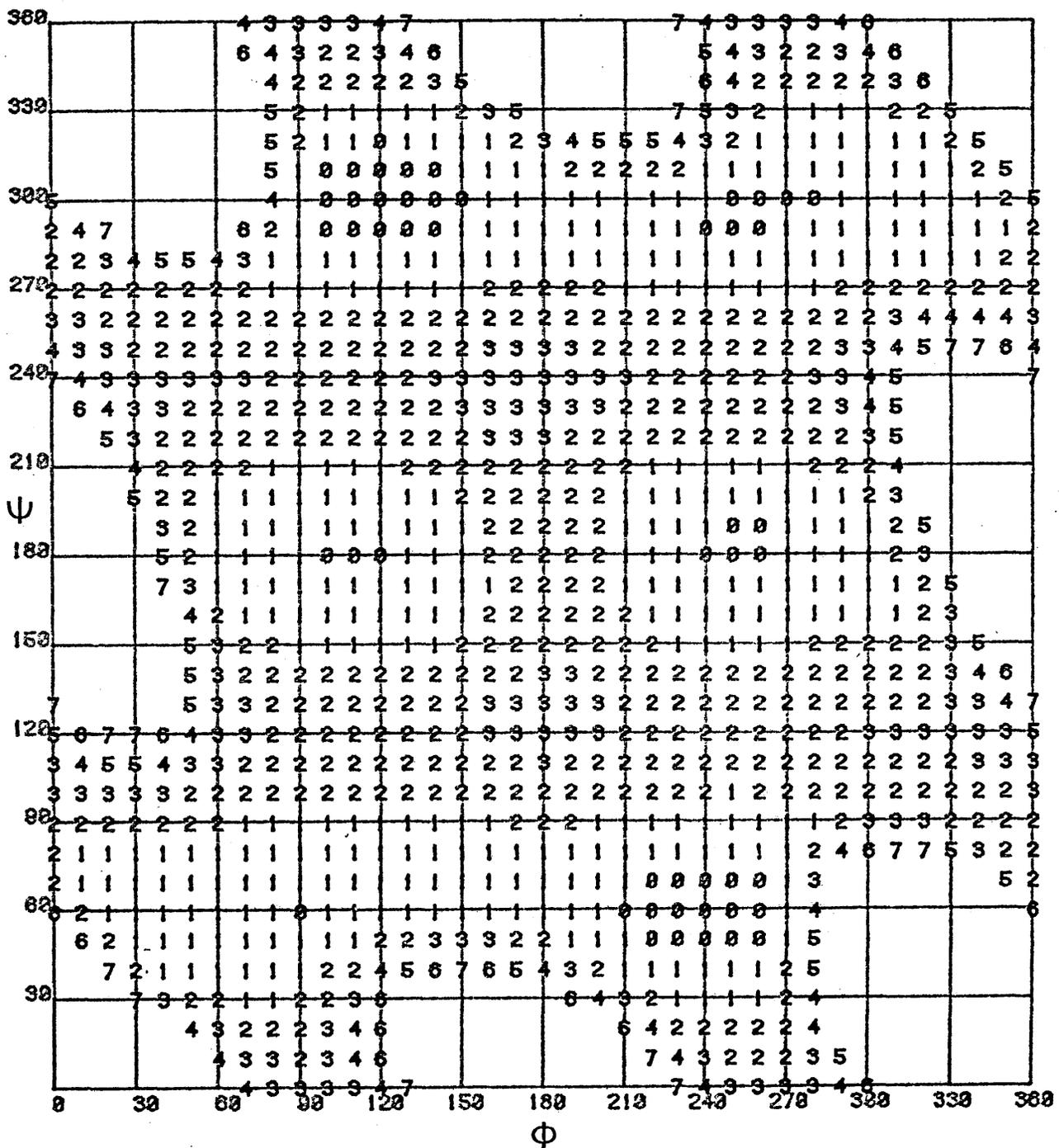
Figure 4.2   Ramachandran Plot of t,t-GlyGly.
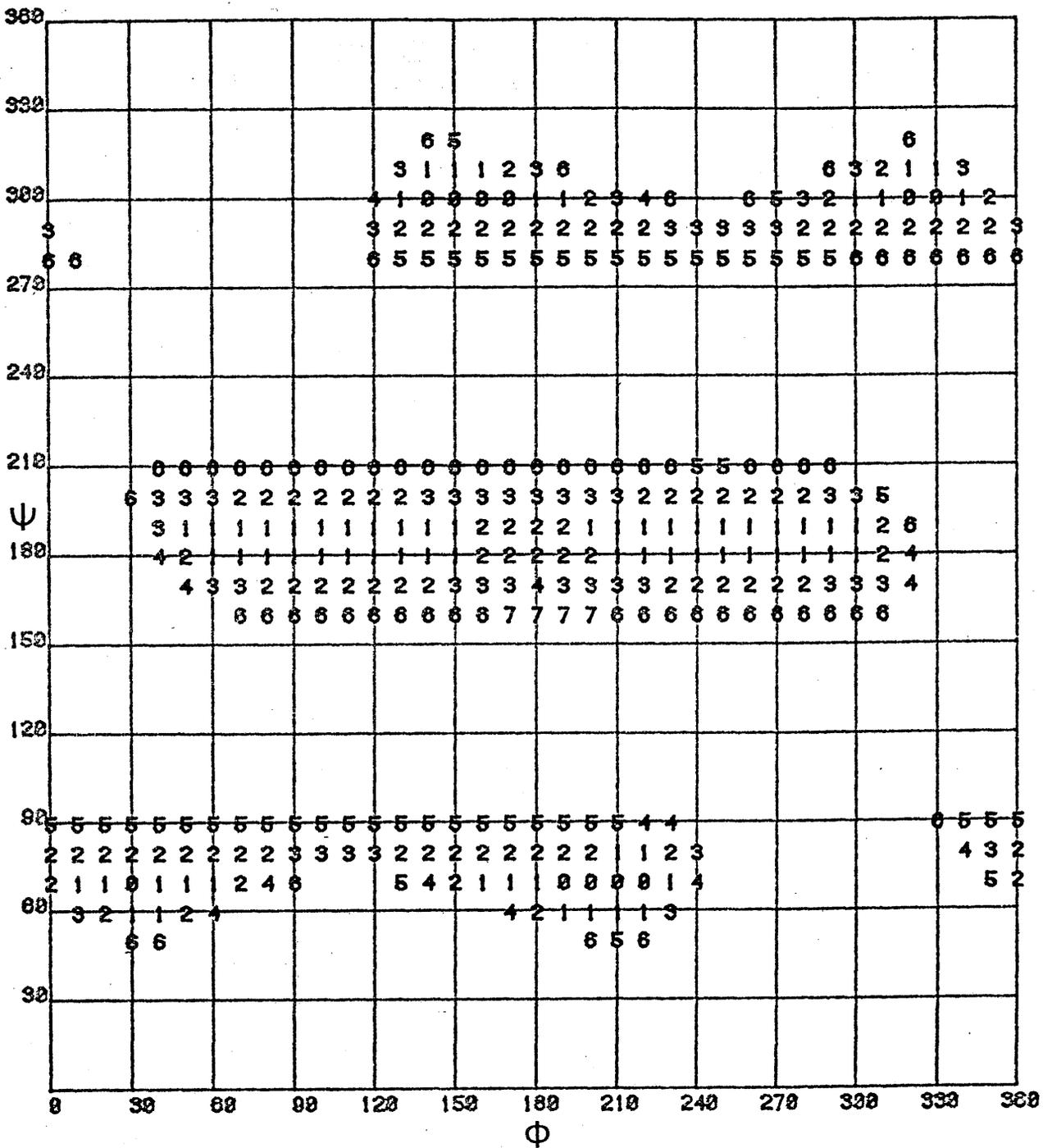
Figure 4.3   Ramachandran Plot of t,c-GlyGly.

Figure 4.4 Ramachandran Plot of c,t-GlyGly.
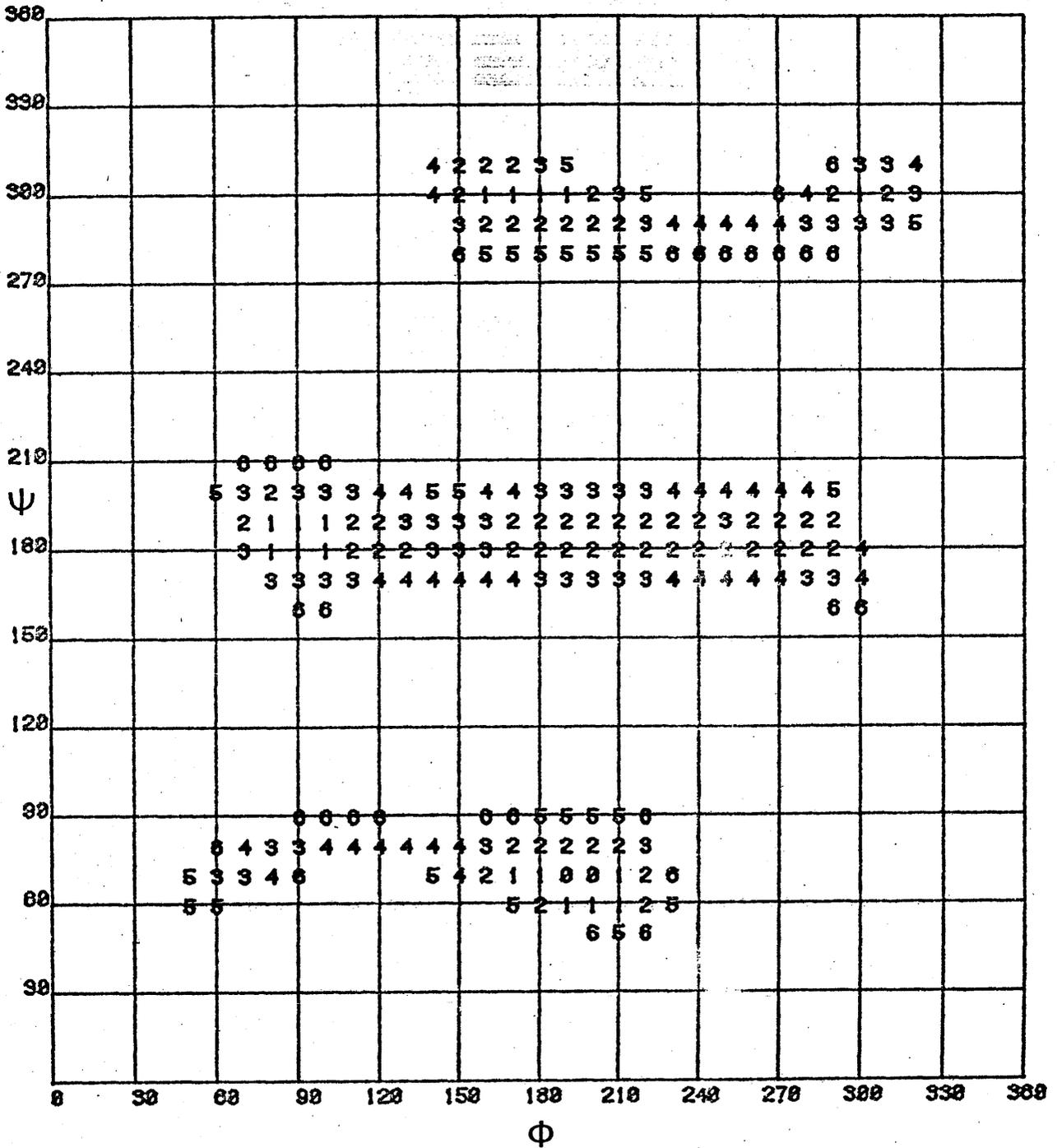
Figure 4.5 Ramachandran Plot of c,c-GlyGly.

4.4 Obtaining Starting Coordinates

The starting coordinates used must represent a reasonable approximation to the equilibrium bond lengths and angles in the target system since these internal coordinates, with the exception of α -carbon angles, are unchanged by the folding process. The values of the variable torsion angles (φ and ψ) are unimportant since these are set to the appropriate values during folding.

Two sources of starting coordinates are available: they may be obtained from crystal structures containing the required fragment; or the linear starting conformation may be constructed using one of the methods of Chapter 3. A combination of these two sources was used.

Initial coordinates were obtained from a crystal structure of cyclo-octasarcosyl [7] which contained amide units of varying geometry. The appropriate configuration of cis and trans amides was then constructed using the Chemical Graphics System. There are six unique cis/trans configurations of cyclotetraglycyl: TTTT, CTTT, CCTT, CTCT, CCCT and CCCC.

4.5 Chain Folding

The linear starting chain was folded up in all possible permutations of the generator torsion angle values. Cyclotetraglycyl has three independent and one dependent (φ, ψ) pairs (one pair lies across what will become the ring junction and is therefore undefined until the ring is closed). There are then 6^3 (for each cis/trans configuration) permutations of the independent pairs of torsion angles. In addition the remaining pair of torsion angles was set to each of the generator values after folding giving $6^4 = 1296$ possible conformations. Since there are six cis/trans configurations the total number of conformers generated was 7776.

Subsequent examination of the generated conformers has shown that the setting of the dependent torsion angles is in fact redundant since the six

'different' conformations generated by this almost invariably converged on the same conformer during potential energy minimisation. This feature has been omitted from the fully automated procedure described later.

4.6 Ring Closure

It should be apparent that of the large number of conformations generated only a proportion will even approximate to the required ring. Each of the generated conformers was subjected to a simple pattern search procedure aimed at optimising the distance between the ends of the chain. This procedure depended on geometric rather than energetic criteria and was aimed at getting the chain end atoms within bonding distance (an obvious necessity for cyclic molecules).

To effect closure the ring torsion angles (φ, ψ and ω) were allowed to be altered by up to 20 degrees. In addition the α-carbon valence angles were similarly allowed to vary by 20 degrees. Limiting the torsion angle variation to 20 degrees ensures that the generated conformation is approximately maintained. Valence angle variation to acheive ring closure is only required in small rings of less than 8 atoms, or 4 or fewer amino acid residues.

The target end-to-end distance was 1.5 Ångstroms but anything closing to within 0.3 Å of this value was taken as being acceptable. Fewer than one-sixth of the original conformations were in fact cyclic (or approximately so) and passed on to the following stages.

4.7 Sterically Unreasonable Conformations

The remaining conformations must be screened to ensure that they conform to the normal precepts of chemical sense. This is necessary since neither the folding nor ring closing algorithms pay any attention to the stereochemical outcome of their efforts. As a result of this many totally

unreasonable conformers are present at this stage i.e. conformations in which the ring is knotted or twisted back on itself.

The method employed in testing for unreasonable conformations was to calculate the number of interatomic distances of 1.9Å or less and comparing the result with the number of bonds that there ought to be. Since any excess 'bonds' can only be due to very short non-bonded contacts the degree of steric crowding can be readily assessed. Notice that only those contacts involving C, N or O atoms were included in the count of short contacts since most H...H non-bonded contacts can be handled by the minimisation routine. It was found that the exact value used as the minimum acceptable non-bonded distance was not critical since there was generally a clear division between the acceptable conformers and those to be rejected. Borderline cases were accepted subject to further examination.

The 20% of the conformations tested which turned out to be reasonable were all examined visually using the Chemical Graphics System. A number of conformers were rejected at this stage mainly due to very bad H...H contacts arising from overlapping $-CH_2-$ groups. Bad geometry at the ring junction caused the rejection of a few conformations.

4.8 Potential Energy Minimisation

All of the surviving conformers were optimised using the pure diagonal energy minimisation algorithm described in Chapter one. It is important to note that prior to optimisation in all 3N-6 coordinates it would have been possible to further reduce the number of conformers if ONLY the global minimum was required.

This could have been done by taking statistical weights into account since each generator can be assigned a statistical weight [8] and only those combinations of generators with high statistical weight products need be considered. Additionally a preliminary optimisation in a reduced

dimensionality such as (φ, ψ) space (keeping ω torsion angles and bond lengths and angles fixed) would have indicated the likely contenders for the global minimum.

In the present work the global minimum AND all other conformations within roughly 20 kcal/mole of it are required and therefore neither of the above procedures were adopted.

The two stage Newton-Raphson iterative optimisation procedure is described fully in Chapter one. An existing force-field [9] was employed and the parameters used are listed in Table I in which the various symbols have the significance assigned in Chapter one. The force constants have units of kcal/mole/A or kcal/mole/deg as appropriate while charges are in units of e.

The first (pure diagonal) stage of the minimisation was carried out on the PDP-11/40 minicomputer. The r.m.s. first derivatives (of energy w.r.t. the coordinates) were reduced to 0.1 kcal/mole/Å or less. Convergence of the pure diagonal method is rapid to this point typically requiring from 50 to 250 iterations of the optimisation. The number of iterations required was dependent on the starting geometry of the conformation.

50 iterations of optimisation take around 2.5 hours on the minicomputer compared to 5 minutes on an IBM 370 mainframe machine. However job scheduling priorities on the mainframe were such that the optimisations were run only at night which effectively gave a total elapsed time of 12 hours as compared to the run time of 5 minutes. Furthermore jobs requiring in excess of 15 minutes run time (i.e. more than 150 iterations) would not be run AT ALL on the mainframe. These scheduling arrangements, designed to handle a large number of small jobs, mean that for a single user (as in this case) a minicomputer provides just as much computing power as the largest of mainframes, as well as being much more convenient to use.

Some conformers failed to converge under the optimisation because the generated geometry was too far from a minimum. Where possible in these

STRETCH

Bond	l_0 (A)	$1/2k_l$
C(sp2)=O	1.24	632.0
C(sp2)-N	1.32	441.0
N-C(sp3)	1.47	379.0
N-H	1.00	430.0
C(sp2)-C(sp3)	1.51	372.0
C(sp3)-C(sp3)	1.535	326.0
C(sp3)-H	1.094	331.0

BEND

Angle	θ_0^n (°)			$1/2k_\theta$
n=	1	2	3	
C(sp2)-N-C(sp3)	0.0	124.0	121.5	.0166
C(sp3)-N-C(sp3)	0.0	0.0	117.0	.0166
C(sp2)-N-H	0.0	119.0	0.0	.0081
C(sp3)-N-H	0.0	117.0	0.0	.0096
O=C(sp2)-N	0.0	0.0	123.0	.0148
O=C(sp2)-C(sp3)	0.0	0.0	120.0	.0125
N-C(sp2)-C(sp3)	0.0	0.0	117.0	.0101
N-C(sp3)-C(sp2)	0.0	112.8	0.0	.0092
N-C(sp3)-C(sp3)	0.0	112.8	0.0	.0092
N-C(sp3)-H	107.8	109.5	0.0	.0092
C(sp2)-C(sp3)-C(sp3)	0.	112.8	0.0	.0067
C(sp2)-C(sp3)-H	109.5	109.5	0.0	.0082
C(sp3)-C(sp3)-H	0.0	109.5	0.0	.0082
H-C(sp3)-H	111.2	108.5	0.0	.0116

VAN der WAALS

Contact	r_1 (A)	r_2 (A)	ε
C(sp3)...C(sp3)	1.55	1.55	.116
C(sp3)...H	1.55	1.45	.084
C(sp2)...C(sp2)	1.85	1.85	.033
C(sp2)...C(sp3)	1.85	1.55	.062
C(sp2)...H	1.85	1.45	.044
O...O	1.54	1.54	.070
O...N	1.54	1.60	.068
O...C(sp2)	1.54	1.85	.048
O...C(sp3)	1.54	1.55	.090
O...H	1.54	1.45	.065

TABLE I. Force Field Parameters.

VAN DER WAALS (CONT'D.)

N...N	1.60	1.60	.066
N...C(sp2)	1.60	1.85	.047
N...C(sp3)	1.60	1.55	.088
N...H	1.60	1.45	.063
H...H	1.45	1.45	.060

TORSION

Angle	$1/2k_T$	Fold
C(sp3)-N-C(sp2)-O	4.487	-2
C(sp3)-N-C(sp2)-C(sp3)	3.228	-2
H-N-C(sp2)-O	0.900	1
H-N-C(sp2)-C(sp3)	3.228	-2
C(sp2)-N-C(sp3)-C(sp2)	0.083	-3
C(sp2)-N-C(sp3)-C(sp3)	0.083	-3
C(sp2)-N-C(sp3)-H	0.083	-3
C(sp3)-C(sp3)-N-C(sp3)	0.083	3
C(sp2)-C(sp3)-N-C(sp3)	0.083	3
C(sp3)-N-C(sp3)-H	0.083	3
H-N-C(sp3)-C(sp2)	0.083	3
H-N-C(sp3)-C(sp3)	0.083	3
H-N-C(sp3)-H	0.083	3
O-C(sp2)-C(sp3)-N	0.063	-3
O-C(sp2)-C(sp3)-H	0.063	-3
O-C(sp2)-C(sp3)-C(sp3)	0.063	-3
N-C(sp2)-C(sp3)-N	0.063	3
N-C(sp2)-C(sp3)-C(sp3)	0.063	3
N-C(sp2)-C(sp3)-H	0.063	3
N-C(sp3)-C(sp3)-C(sp2)	0.158	3
N-C(sp3)-C(sp3)-H	0.158	3
H-C(sp3)-C(sp3)-C(sp2)	0.158	3
H-C(sp3)-C(sp3)-H	0.158	3

OUT-OF-PLANE BENDING

Deformation Angle	$1/2k_\delta$
C(sp3)-O-C(sp2)-N	.0012
C(sp3)-H-N-C(sp2)	.0002

CHARGES[10]

Atom	Charge	Atom	Charge
O	-0.416	N	-0.305
C(sp2)	+0.449	H(-N)	+0.272

Dielectric Constant D = 4.

cases any obviously bad geometry was adjusted manually using the Chemical Graphics System, this was normally sufficient to induce convergence. After all of the conformers had undergone the first stage minimisation it was immediately apparent that some were converging on the same geometry. Where this was definitely happening only one of the pair (or more) passed on to the second stage (Full Matrix) optimisation. Where there was any doubt as to the final result no elimination was done.

Mirror images were also rejected in this (and later) comparisons, these arise since generators 1,2 & 3 are the mirror images of 6,5 & 4. This means that the generator sequence 1111 is the mirror image of 6666 in a symmetric configuration such as TTTT or CTCT.

The second stage optimisation was carried out on an IBM 370/168 at NUMAC since the amount of core available on the minicomputer (28K words) was insufficient. The PDP-11/40 now has 64K words of core and the Full Matrix minimisation is being implemented on it. Optimisation was regarded as complete when each of the first derivatives of energy w.r.t. the coordinates was less than 10^{-5} kcal/mole/Å.

On completion of the full matrix minimisation the remaining 66 conformers were examined using the Chemical Graphics System and it was noticed that in some cases amide groups had "flipped" from cis to trans or vice versa. This usually occurred near the ring junction and is therefore hardly surprising since the geometry of some of the ring junctions was far from ideal. A further comparison was therefore carried out between all of the remaining conformers rather than within configurations as previously done. This left a total of 40 unique conformers ranging up to about 20 kcal/mole above the global minimum.

4.9 Results

Table II shows the numbers of conformers remaining after each stage of the algorithm. The Global Minimum Energy Conformation is an all trans conformation with S4 symmetry. The most interesting feature of this

CONFIGURATION [C=cis, T=trans]	INITIAL NUMBER	AFTER RING CLOSURE	VISUAL EXAMINATION & BOND CHECK	STAGE I MINIMISATION	COMPARISON	FINAL MINIMISATION & COMPARISON
1. TTTT	1296	96	18	12	6	4
2. CTTT	1296	42	8	5	5	4
3. CCTT	1296	91	20	14	11	5
4. CTCT	1296	227	32	20	10	6
5. CCCT	1296	294	40	23	16	12
6. CCCC	1296	297	44	30	18	9
TOTAL	7776	1047	162	104	66	40

TABLE II. Numbers of Conformers.

CONFORMATION	CONFIGURATION & SYMMETRY		ENERGY	CONFORMATION	CONFIGURATION & SYMMETRY		ENERGY
1. GL1-5252	TTTT	S4	0.0000	21. GL5-1232	CCCT		8.0895
2. GL4-4532	CTCT	C1	1.2954	22. GL3-1622	CCCT		8.5317
3. GL4-1531	CTCT		2.1960	23. GL4-2122	CCCT		8.7843
4. GL1-5623	TTTT	C1	2.2106	24. GL5-2422	CCCC	C2	8.9170
5. GL4-5126	CTCT	C1	2.5037	25. GL5-2441	CCCT		9.1853
6. GL1-6252	TTTT		2.8738	26. GL6-1611	CCCC	S4	9.8127
7. GL1-5212	CTTT		4.0485	27. GL5-3522	CCCT		10.4246
8. GL4-2141	CTCT	C2	4.1620	28. GL5-6535	CCCC		10.4473
9. GL2-6251	CTTT		4.6682	29. GL5-2353	CCCC	C1	10.4538
10. GL4-6221	CTCT	C2	5.0444	30. GL5-1611	CCCT		10.5445
11. GL1-6214	CTTT		5.0708	31. GL5-1652	CCCC	S4	11.6331
12. GL4-2634	CTCT		5.0784	32. GL5-5251	CCCT		11.7272
13. GL1-1111	TTTT	C4	5.8381	33. GL3-1662	CCTT		12.0068
14. GL2-6664	CCTT		6.2128	34. GL5-3532	CCCT		12.3703
15. GL2-4113	CTTT		6.3104	35. GL6-1616	CCCC	S4	13.0554
16. GL5-4232	CCCT		6.5761	36. GL6-5235	CCCC		13.4853
17. GL3-3156	CCTT		6.6676	37. GL6-6541	CCCC		14.8631
18. GL6-2351	CCCT		7.1735	38. GL3-6152	CCTT		15.9037
19. GL5-4245	CCCT		7.3274	39. GL3-1255	CCTT		16.0581
20. GL5-5441	CCCT		8.0737	40. GL6-5254	CCCC	S4	20.0838

TABLE III. Energies (kcal/mole) above the Global Minimum.

conformation is the fact that all of the amide units are non-planar by 17 degrees, as was to be expected [10]. The energies, relative to this minimum, of the other low energy conformations are shown in Table III while Figures 4.6 — 4.45 show the geometries and ring torsion angles of these. ✳

The GLn codes in Table III and Figures 4.6 — 4.45 refer to the starting configuration of the conformation i.e. GL1 refers to a starting all—trans configuraton, and Table IIIA shows the whole list. The final configuration is shown as TTTT, CTCT etc. The numbers after the hyphen give the sequence of generators i.e. 1111 means that generator no. 1 was used four times.

The presence of conformations containing cis—amides in the low energy set is not surprising in view of the strain imposed on all—trans conformations by ring closure. Even the calculated minimum energy conformation has considerable Pitzer and Baeyer strain. In contrast the CTCT conformations are almost entirely free from strain in the region of the amide groups and this compensates for the presence of the energetically unfavourable cis amides.

✳ Facing these figures are tables of the orthogonal co-ordinates of the various conformers, together with the values of χ_C, χ_N, and Υ (as defined by Winkler and Dunitz [10]).

GL1	TTTT
GL2	CTTT
GL3	CCTT
GL4	CTCT
GL5	CCCT
GL6	CCCC

TABLE IIIA. Correspondence between GL codes and starting configurations.

ATOM	X	Y	Z
C'1	2.01167	-0.04681	-1.38374
O1	1.37280	-0.60695	-2.28732
N1	1.47357	0.27033	-0.21650
H1	2.01900	0.81312	0.42521
CA1	0.05320	0.09686	0.13639
H11	-0.45493	-0.57923	-0.55851
H12	-0.03052	-0.32353	1.14372
C'2	-0.57506	1.47203	0.09434
O2	-0.80611	2.08280	1.13883
N2	-0.61822	2.05330	-1.10430
H2	-0.31796	1.52556	-1.90132
CA2	-0.84748	3.49055	-1.33638
H21	-1.25875	3.98142	-0.44865
H22	-1.55633	3.62528	-2.15968
C'3	0.49522	4.08960	-1.69252
O3	0.75631	4.35359	-2.87594
N3	1.41708	4.07124	-0.74254
H3	1.15302	3.74748	0.16815
CA3	2.86194	4.26329	-0.96071
H31	3.06629	4.70270	-1.94220
H32	3.26703	4.93376	-0.19597
C'4	3.49974	2.89560	-0.85506
O4	4.10861	2.57927	0.17817
N4	3.15958	2.01566	-1.78382
H4	2.57828	2.32442	-2.53924
CA4	3.36427	0.55940	-1.68619
H41	4.07929	0.30504	-0.89738
H42	3.75194	0.17447	-2.63487

Orthogonal Coordinates of: GL1-5252

AMIDE	CHI(C)	CHI(N)	TAU
1	-12.42	11.31	174.97
2	12.54	-11.34	-174.99
3	-12.53	11.33	175.00
4	12.54	-11.35	-174.99

Planarity of the amide groups of: GL1-5252
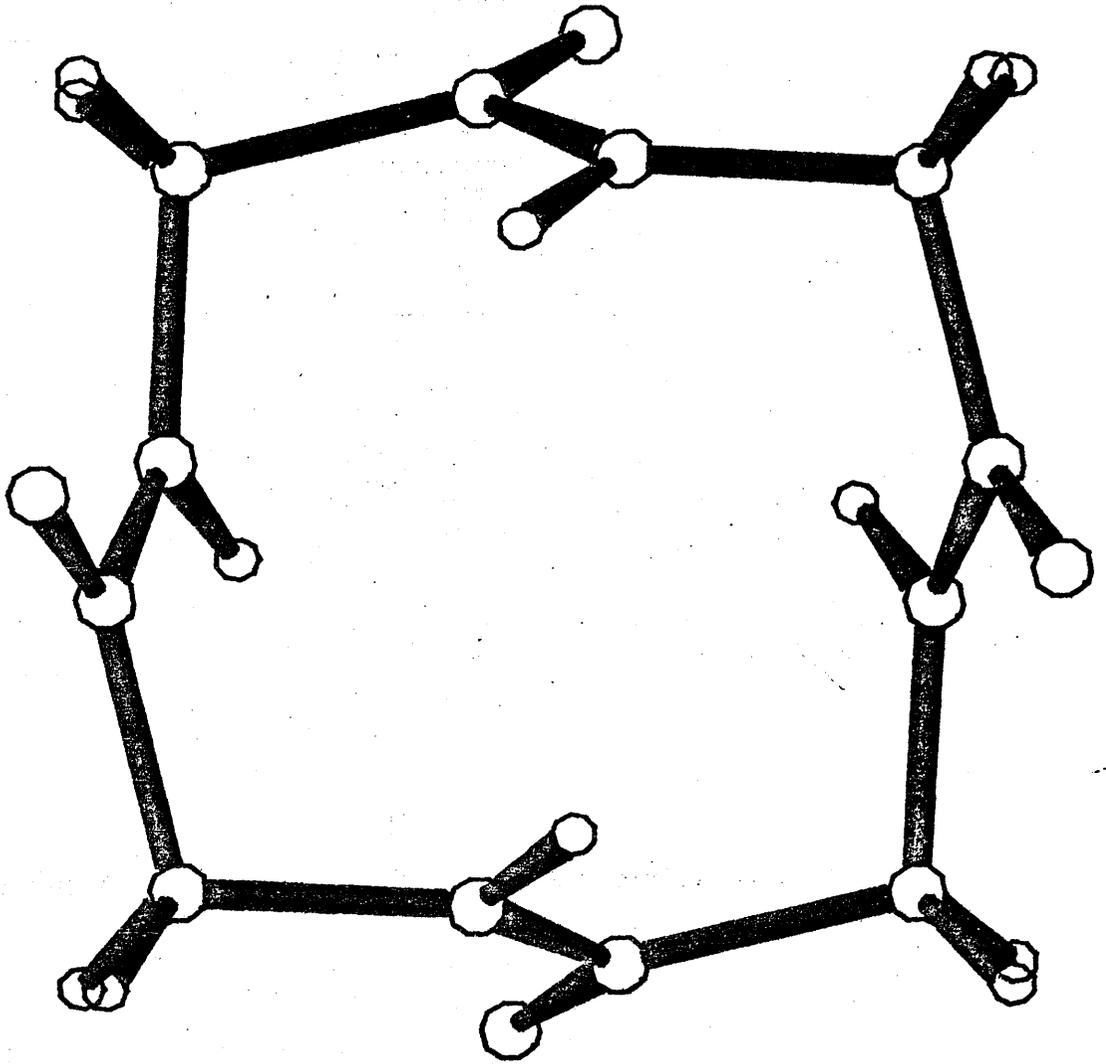
	ϕ_n	ψ_n	ω_n
n			
1	-104	65	163
2	104	-65	-163
3	-104	65	163
4	104	-65	-163

Figure 4.06 GL1-5252 TTTT S4

ATOM	X	Y	Z
C'1	0.89893	-0.33655	2.50748
O1	1.84298	-1.02982	2.09880
N1	0.02876	0.24473	1.69622
H1	-0.69038	0.81684	2.09460
CA1	0.04530	0.15612	0.22517
H11	0.84854	-0.49848	-0.12494
H12	-0.89701	-0.29921	-0.09622
C'2	0.16898	1.51904	-0.42341
O2	-0.78291	1.94341	-1.09677
N2	1.27149	2.23350	-0.24395
H2	1.29100	3.15929	-0.62497
CA2	2.46013	1.81881	0.52446
H21	3.33419	2.37222	0.16603
H22	2.66160	0.75899	0.34870
C'3	2.31297	2.06860	2.00904
O3	1.36877	2.76174	2.41772
N3	3.18909	1.48735	2.82033
H3	3.90219	0.91518	2.42197
CA3	3.16648	1.57591	4.29138
H31	4.10876	2.03127	4.61282
H32	2.36320	2.23046	4.64148
C'4	3.04284	0.21297	4.93993
O4	3.99475	-0.21138	5.61327
N4	1.94035	-0.50152	4.76048
H4	1.92086	-1.42731	5.14148
CA4	0.75169	-0.09684	3.99208
H41	0.55015	0.97295	4.16794
H42	-0.12234	-0.64035	4.35043

Orthogonal Coordinates of: GL4-4532

AMIDE	CHI(C)	CHI(N)	TAU
1	-2.52	1.90	179.89
2	1.07	3.14	2.70
3	2.55	-1.91	-179.90
4	-1.07	-3.14	-2.70

Planarity of the amide groups of: GL4-4532
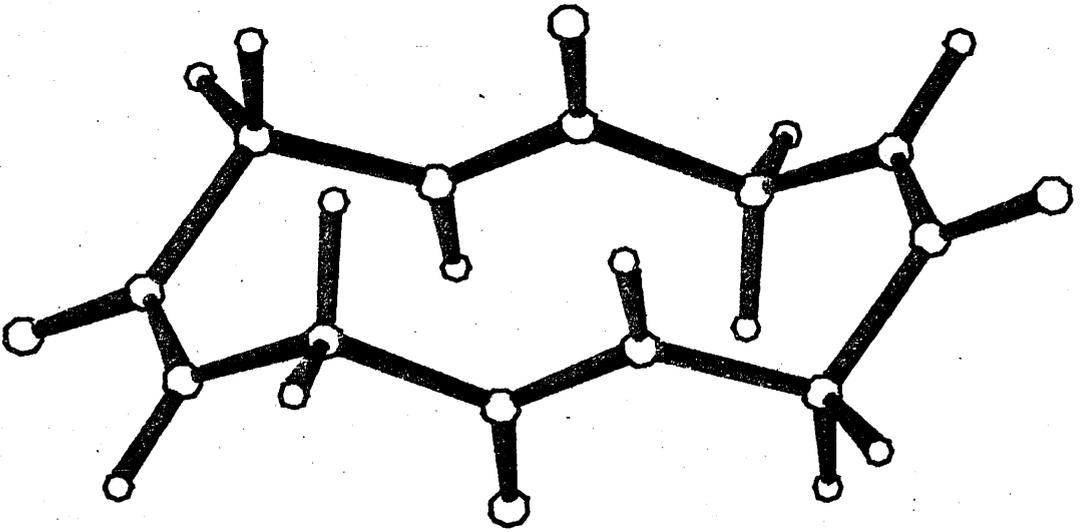
n	ϕ_n	ψ_n	ω_n
1	-121	65	178
2	-83	166	2
3	121	-65	-178
4	83	-166	-2

Figure 4.07 GL4-4532 CTCT Ci

ATOM	X	Y	Z
C'1	2.20140	-0.13998	0.85074
O1	2.16795	0.86434	1.57793
N1	1.16965	-0.52911	0.11868
H1	1.26382	-1.34647	-0.45240
CA1	-0.13167	0.16100	0.07769
H11	-0.17431	0.75034	-0.84375
H12	-0.94202	-0.57228	0.02873
C'2	-0.35555	1.08059	1.25913
O2	-0.37790	2.30593	1.06429
N2	-0.49971	0.56146	2.47113
H2	-0.59618	1.20297	3.23383
CA2	-0.55654	-0.87482	2.80872
H21	-1.25873	-1.37698	2.13714
H22	-0.96378	-0.98322	3.81942
C'3	0.76720	-1.60823	2.77205
O3	0.87034	-2.63183	2.07851
N3	1.76647	-1.14298	3.50570
H3	1.62410	-0.30434	4.03386
CA3	3.08779	-1.78607	3.63695
H31	3.23057	-2.03451	4.69362
H32	3.11867	-2.73176	3.08843
C'4	4.24373	-0.91171	3.19715
O4	5.05865	-0.53269	4.05269
N4	4.37249	-0.58784	1.91824
H4	5.10975	0.04305	1.67122
CA4	3.45621	-0.98134	0.83253
H41	3.19677	-2.03920	0.92677
H42	3.95241	-0.85409	-0.13509

Orthogonal Coordinates of: GL4-1531

AMIDE	CHI(C)	CHI(N)	TAU
1	-2.04	-0.03	178.59
2	-1.14	1.17	-3.24
3	-1.32	1.50	177.74
4	1.01	-5.57	-1.90

Planarity of the amide groups of: GL4-1531
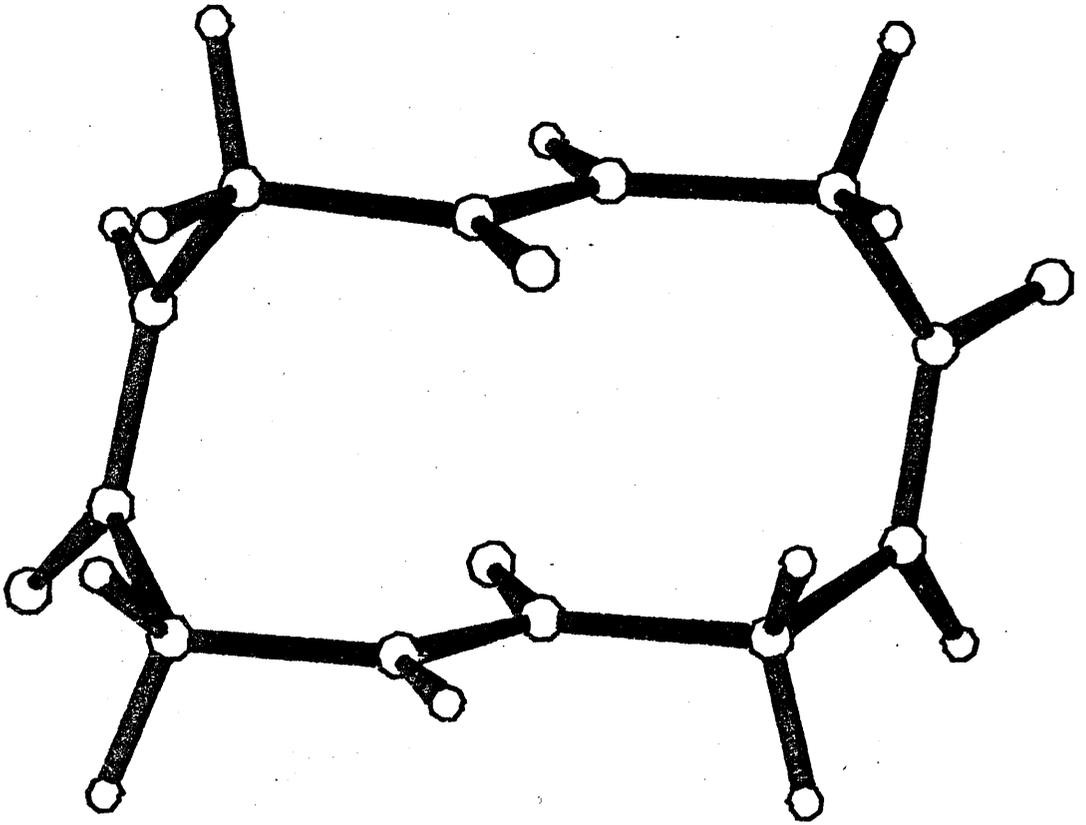
	ϕ_n	ψ_n	ω_n
n			
1	-19	-68	178
2	75	55	-4
3	121	-67	176
4	77	-167	1

Figure 4.08 GL4-1531 CTCT

ATOM	X	Y	Z
C'1	-1.14874	-0.15478	2.25376
O1	-1.99126	-0.82540	1.63858
N1	-0.00171	0.23385	1.71784
H1	0.58872	0.83359	2.26050
CA1	0.38289	0.03586	0.30819
H11	1.41311	-0.32757	0.23664
H12	-0.26979	-0.70159	-0.16936
C'2	0.23119	1.37269	-0.38543
O2	-0.80531	1.61571	-1.02181
N2	1.10989	2.31334	-0.07512
H2	1.87542	2.07535	0.52535
CA2	0.93264	3.75212	-0.34163
H21	0.14585	3.93173	-1.08116
H22	1.86533	4.17765	-0.72572
C'3	0.56473	4.38853	0.98070
O3	1.40784	5.05798	1.59637
N3	-0.58217	3.99975	1.51680
H3	-1.17250	3.39978	0.97428
CA3	-0.96641	4.19739	2.92661
H31	-0.31358	4.93472	3.40416
H32	-1.99658	4.56086	2.99848
C'4	-0.81463	2.86045	3.62001
O4	0.22218	2.61713	4.25575
N4	-1.69373	1.92002	3.31013
H4	-2.45950	2.15823	2.71005
CA4	-1.51655	0.48117	3.57633
H41	-2.44925	0.05558	3.96035
H42	-0.72973	0.30136	4.31577

Orthogonal Coordinates of: GL1-5623

AMIDE	CHI(C)	CHI(N)	TAU
1	11.48	-10.40	-173.81
2	12.49	-11.59	-174.61
3	-11.59	10.43	173.84
4	-12.49	11.60	174.61

Planarity of the amide groups of: GL1-5623
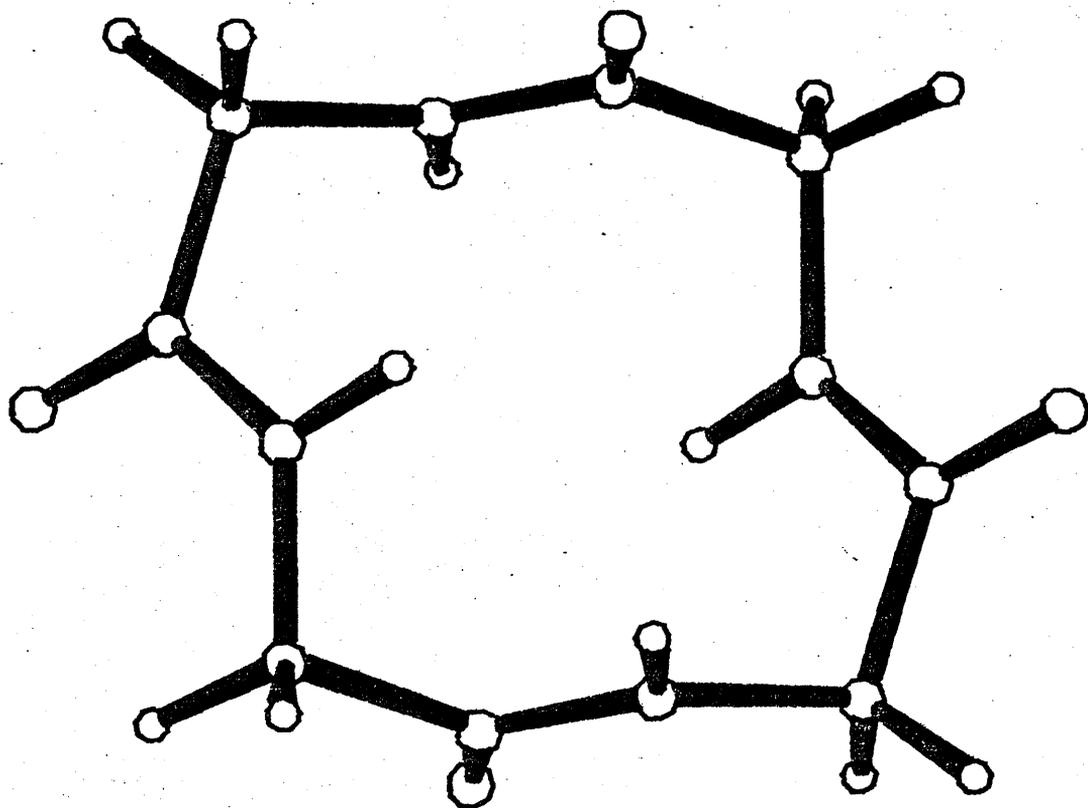
n	ϕ_n	ψ_n	ω_n
1	104	71	-163
2	103	-64	-163
3	-104	-71	163
4	-103	64	163

Figure 4.09 GL1-5623 TTTT Ci

ATOM	X	Y	Z
C'1	1.08240	0.99940	1.83962
O1	1.40298	1.95343	1.11439
N1	0.32583	-0.00208	1.41933
H1	0.11652	-0.74208	2.06076
CA1	-0.22050	-0.14268	0.05712
H11	0.34671	-0.93075	-0.44836
H12	-1.25853	-0.48328	0.11166
C'2	-0.14579	1.11257	-0.78644
O2	0.70959	1.17671	-1.68317
N2	-0.99569	2.10437	-0.55870
H2	-0.90227	2.92762	-1.12088
CA2	-2.07941	2.10824	0.44251
H21	-2.74681	1.26120	0.25802
H22	-2.68172	3.01478	0.32097
C'3	-1.58909	2.06389	1.87252
O3	-1.90970	1.10988	2.59774
N3	-0.83252	3.06538	2.29281
H3	-0.62322	3.80538	1.65139
CA3	-0.28622	3.20598	3.65503
H31	0.75181	3.54659	3.60052
H32	-0.85345	3.99403	4.16052
C'4	-0.36092	1.95072	4.49858
O4	-1.21630	1.88655	5.39530
N4	0.48899	0.95892	4.27083
H4	0.39558	0.13566	4.83300
CA4	1.57272	0.95507	3.26962
H41	2.17505	0.04855	3.39117
H42	2.24009	1.80214	3.45411

Orthogonal Coordinates of: GL4-5126

AMIDE	CHI(C)	CHI(N)	TAU
1	0.31	-1.50	-179.69
2	0.56	0.50	-2.00
3	-0.31	1.50	179.70
4	-0.56	-0.50	2.00

Planarity of the amide groups of: GL4-5126
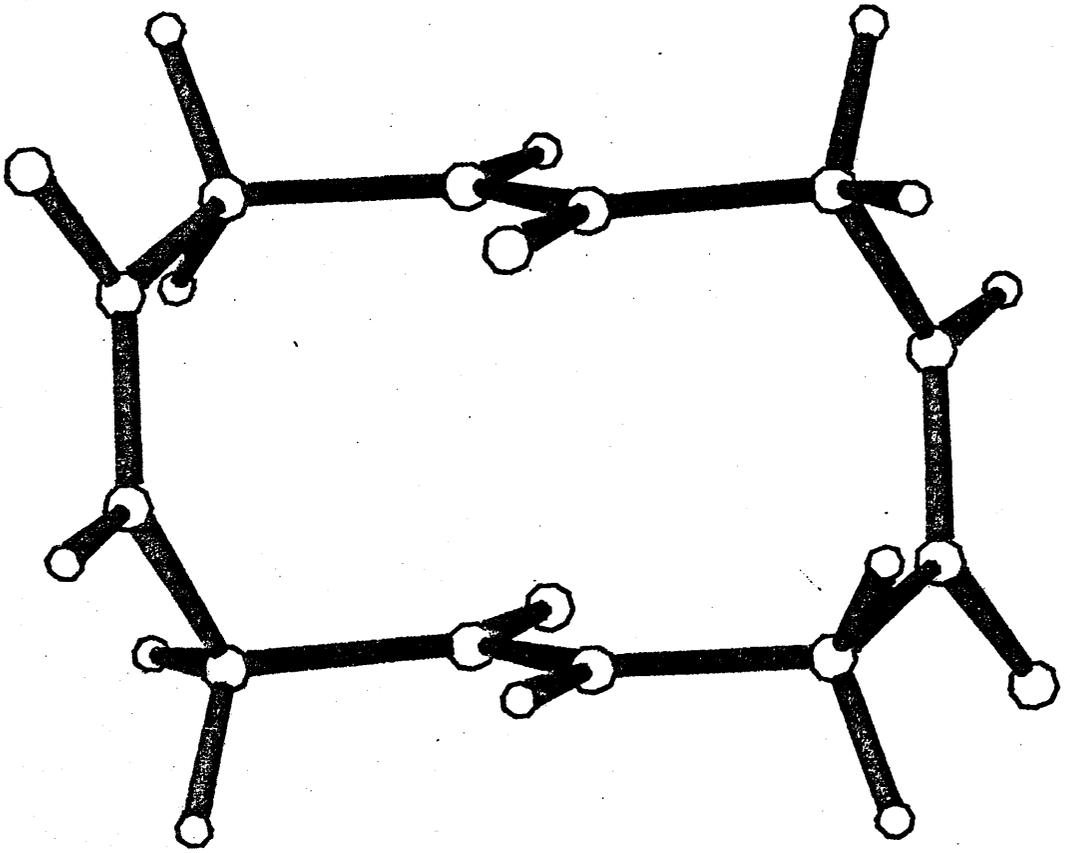
n	ϕ_n	ψ_n	ω_n
1	−14	−75	−179
2	66	61	−2
3	14	75	179
4	−66	−61	2

Figure 4.10 GL4−5126 CTCT Ci

ATOM	X	Y	Z
C'1	1.43327	0.48299	1.85147
O1	2.46981	0.19103	1.23652
N1	0.22449	0.36087	1.32554
H1	-0.55187	0.70250	1.85728
CA1	-0.03861	0.00326	-0.08035
H11	0.77558	-0.61036	-0.47871
H12	-0.96985	-0.56523	-0.17171
C'2	-0.11980	1.30765	-0.84316
O2	0.84380	1.67759	-1.53027
N2	-1.13055	2.10838	-0.54210
H2	-1.84344	1.76307	0.06922
CA2	-1.16301	3.55035	-0.84594
H21	-2.19223	3.89276	-0.99496
H22	-0.59974	3.76017	-1.76075
C'3	-0.52444	4.26681	0.32518
O3	0.60156	4.77158	0.20150
N3	-1.12046	4.13976	1.50098
H3	-1.99728	3.65924	1.54697
CA3	-0.50782	4.47480	2.79839
H31	0.32674	5.17297	2.67854
H32	-1.25221	4.94855	3.44654
C'4	-0.02750	3.17901	3.41534
O4	-0.74292	2.59631	4.24427
N4	1.05195	2.62433	2.88635
H4	1.53306	3.11791	2.15854
CA4	1.52275	1.25199	3.15092
H41	0.91804	0.76072	3.91971
H42	2.55957	1.28150	3.50085

Orthogonal Coordinates of: GL1-6252

AMIDE	CHI(C)	CHI(N)	TAU
1	-11.57	12.45	173.84
2	-12.34	11.46	172.29
3	-10.52	9.57	174.73
4	10.29	-8.36	-176.43

Planarity of the amide groups of: GL1-6252
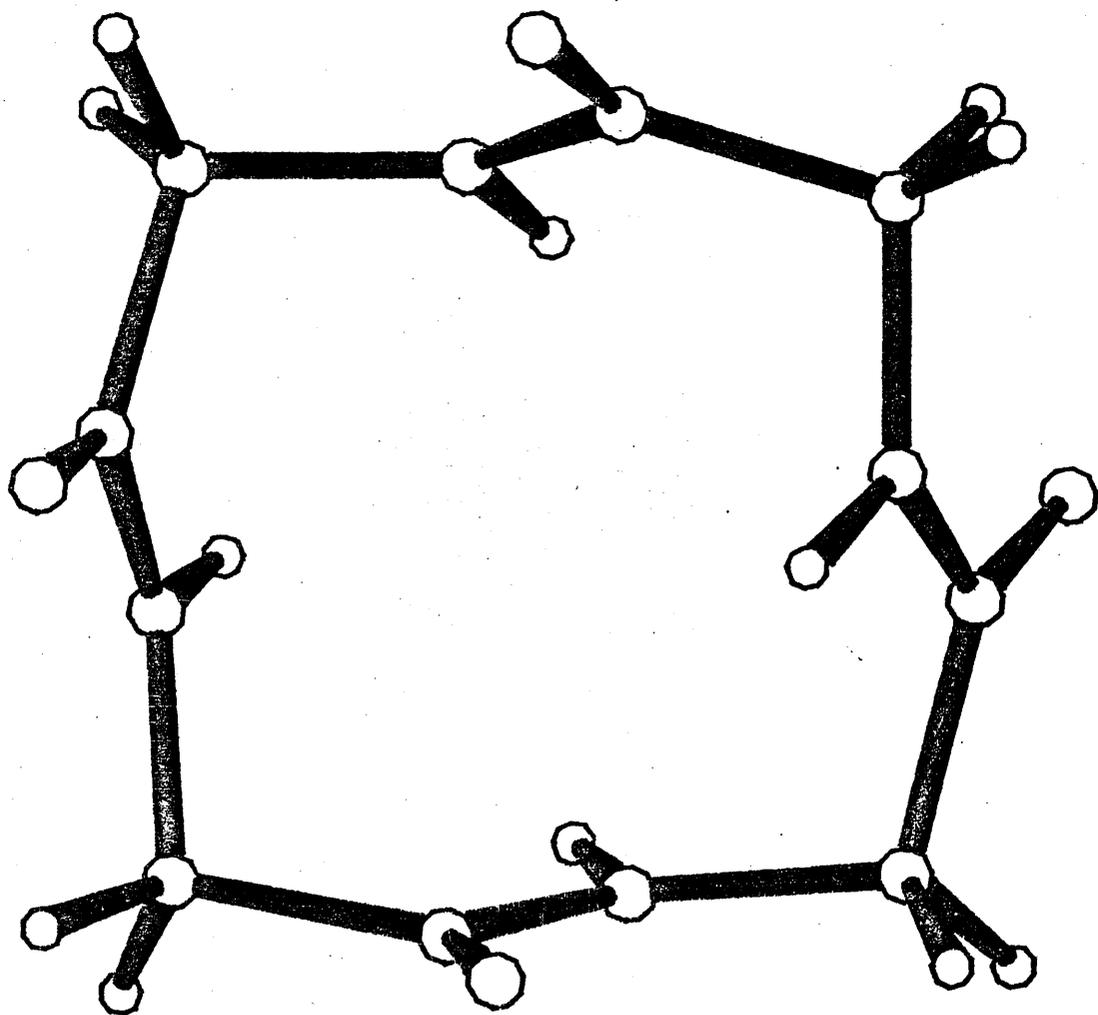
	ϕ_n	ψ_n	ω_n
n			
1	−91	−66	162
2	−88	−62	160
3	−100	73	165
4	115	−61	−167

Figure 4.11 GL1-6252 TTTT

ATOM	X	Y	Z
C'1	0.38272	-0.17575	2.63657
O1	-0.22864	-0.52212	3.65922
N1	-0.15145	-0.34305	1.43228
H1	-1.09837	-0.66733	1.39635
CA1	0.45435	0.01148	0.13479
H11	1.54027	-0.10369	0.18873
H12	0.09320	-0.66912	-0.64280
C'2	0.10072	1.43098	-0.24833
O2	-0.81690	1.99826	0.36285
N2	0.89601	2.07576	-1.08748
H2	1.68731	1.59312	-1.46741
CA2	0.87299	3.53710	-1.27718
H21	-0.14988	3.92686	-1.26018
H22	1.32429	3.80329	-2.23858
C'3	1.69134	4.09347	-0.13416
O3	2.91889	4.20273	-0.27405
N3	1.09886	4.15078	1.04843
H3	0.11420	3.98046	1.10774
CA3	1.80686	4.21203	2.33753
H31	2.87127	4.42006	2.19103
H32	1.38632	5.01260	2.95450
C'4	1.65199	2.87501	3.03242
O4	1.23161	2.82835	4.19824
N4	1.92420	1.78895	2.32243
H4	2.27613	1.90698	1.39382
CA4	1.77483	0.40255	2.80414
H41	2.02116	0.38428	3.87065
H42	2.50179	-0.24669	2.30599

Orthogonal Coordinates of: GL1-5212

AMIDE	CHI(C)	CHI(N)	TAU
1	-1.98	5.98	2.31
2	9.49	-13.97	-175.58
3	-15.37	11.46	171.62
4	3.07	1.32	-178.01

Planarity of the amide groups of: GL1-5212
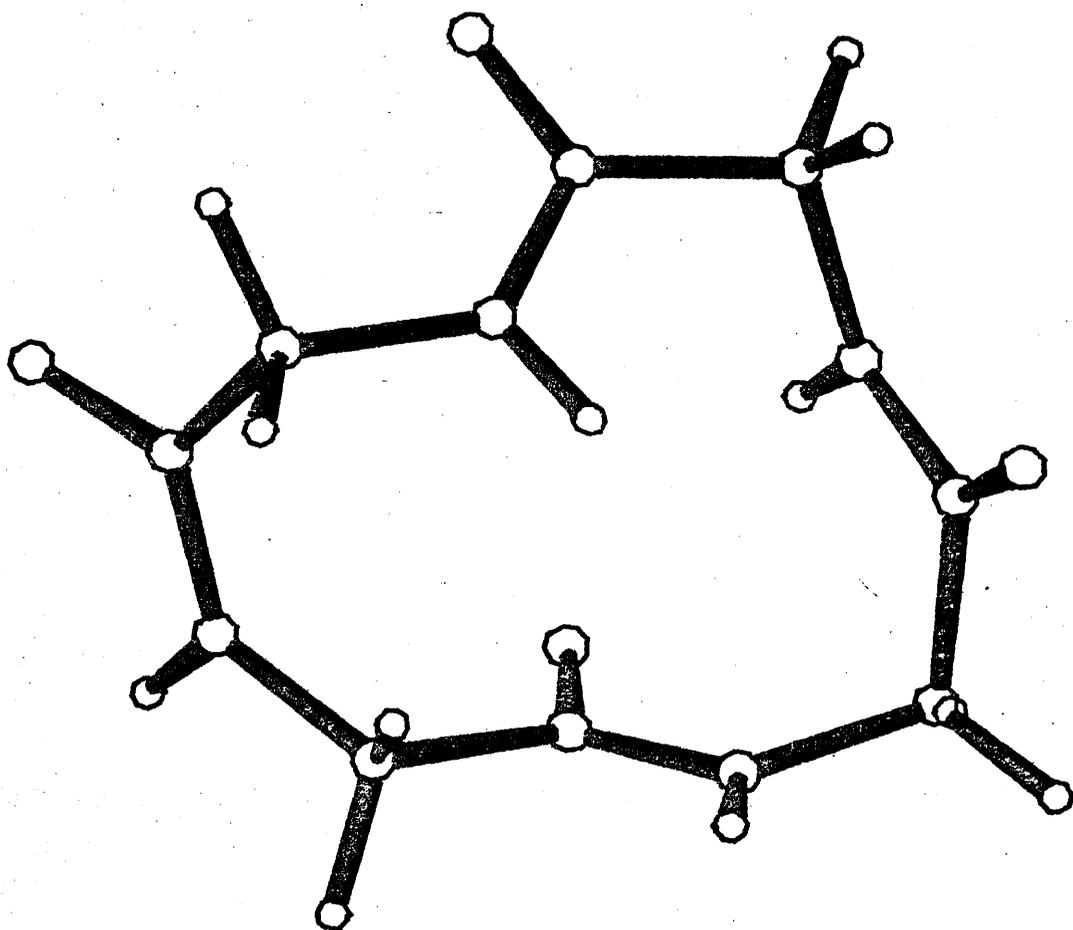
n	ϕ_n	ψ_n	ω_n
1	-89	156	-2
2	81	-78	-164
3	-108	49	158
4	86	63	-177

Figure 4.12 GL1-5212 CTTT

ATOM	X	Y	Z
C'1	1.85944	-1.34554	1.95487
O1	1.15366	-1.34554	1.95487
N1	1.34665	-0.24652	0.00578
H1	1.95471	0.04041	-0.73600
CA1	-0.08744	0.05085	-0.16414
H11	-0.21083	0.94970	-0.77703
H12	-0.54683	-0.77845	-0.71129
C'2	-0.83052	0.26116	1.13917
O2	-1.85467	-0.40628	1.35236
N2	-0.39692	1.17435	1.99890
H2	-0.86269	1.23051	2.88363
CA2	0.78884	2.03798	1.83646
H21	0.56760	3.03599	2.22767
H22	1.02838	2.15575	0.77654
C'3	1.98838	1.48440	2.57437
O3	1.80706	0.76347	3.56650
N3	3.19802	1.76977	2.11851
H3	3.29052	2.41393	1.35772
CA3	4.44900	1.22900	2.68111
H31	5.22011	1.19106	1.90475
H32	4.80304	1.91823	3.45426
C'4	4.30732	-0.15727	3.27541
O4	4.68842	-0.34452	4.44139
N4	3.82035	-1.14139	2.53014
H4	3.66475	-2.02342	2.97809
CA4	3.36275	-1.03721	1.13085
H41	3.65620	-1.93919	0.58464
H42	3.84796	-0.19176	0.63641

Orthogonal Coordinates of: GL4-2141

AMIDE	CHI(C)	CHI(N)	TAU
1	-4.69	0.96	-174.36
2	-2.05	6.17	3.23
3	1.28	1.80	-175.54
4	-2.05	6.17	3.23

Planarity of the amide groups of: GL4-2141
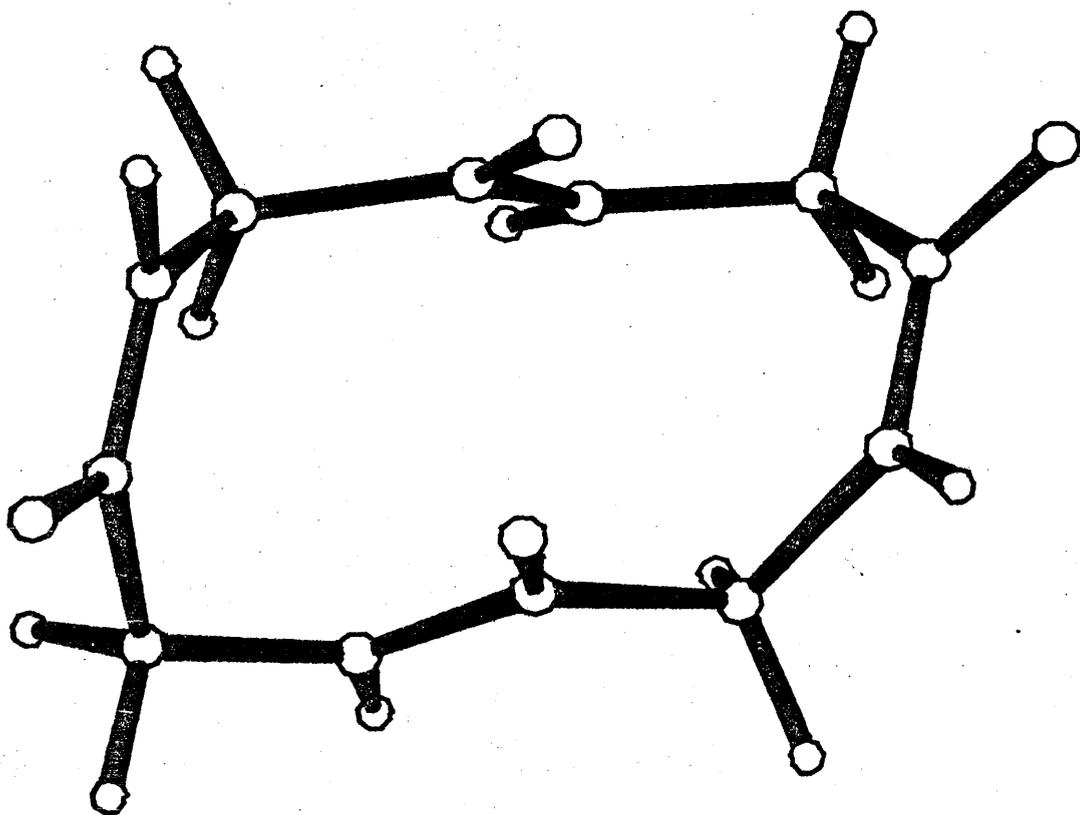
	ϕ_n	ψ_n	ω_n
n			
1	31	55	-176
2	-98	152	-1
3	31	55	-176
4	-98	152	-1

Figure 4.13 GL4-2141 CTCT C2

ATOM	X	Y	Z
C'1	1.69927	1.13054	-1.08289
O1	1.45417	0.66535	-2.20627
N1	1.05058	0.75703	0.01018
H1	1.34278	1.12950	0.89132
CA1	-0.06115	-0.20709	0.05035
H11	-0.75987	-0.01356	-0.76843
H12	0.35408	-1.20808	-0.10383
C'2	-0.78111	-0.18783	1.38377
O2	-0.47358	-1.05165	2.21949
N2	-1.65975	0.76393	1.67503
H2	-2.09550	0.71490	2.57555
CA2	-2.14191	1.83648	0.78199
H21	-2.99852	2.33331	1.24958
H22	-2.50670	1.39150	-0.14864
C'3	-1.11966	2.90378	0.45912
O3	-0.86202	3.14510	-0.72928
N3	-0.50005	3.51461	1.45804
H3	-0.70875	3.22934	2.39496
CA3	0.59847	4.48560	1.28678
H31	0.33593	5.23829	0.53652
H32	0.79145	5.00607	2.23064
C'4	1.82471	3.70530	0.87261
O4	2.49775	3.14681	1.75243
N4	1.93432	3.40016	-0.41031
H4	1.29909	3.81521	-1.06340
CA4	2.70116	2.25387	-0.92225
H41	3.49166	1.95721	-0.22565
H42	3.16429	2.50411	-1.88219

Orthogonal Coordinates of: GL2-6251

AMIDE	CHI(C)	CHI(N)	TAU
1	-4.79	-1.88	174.51
2	3.64	-4.78	0.72
3	-2.33	7.44	179.30
4	16.15	-15.57	-170.45

Planarity of the amide groups of: GL2-6251
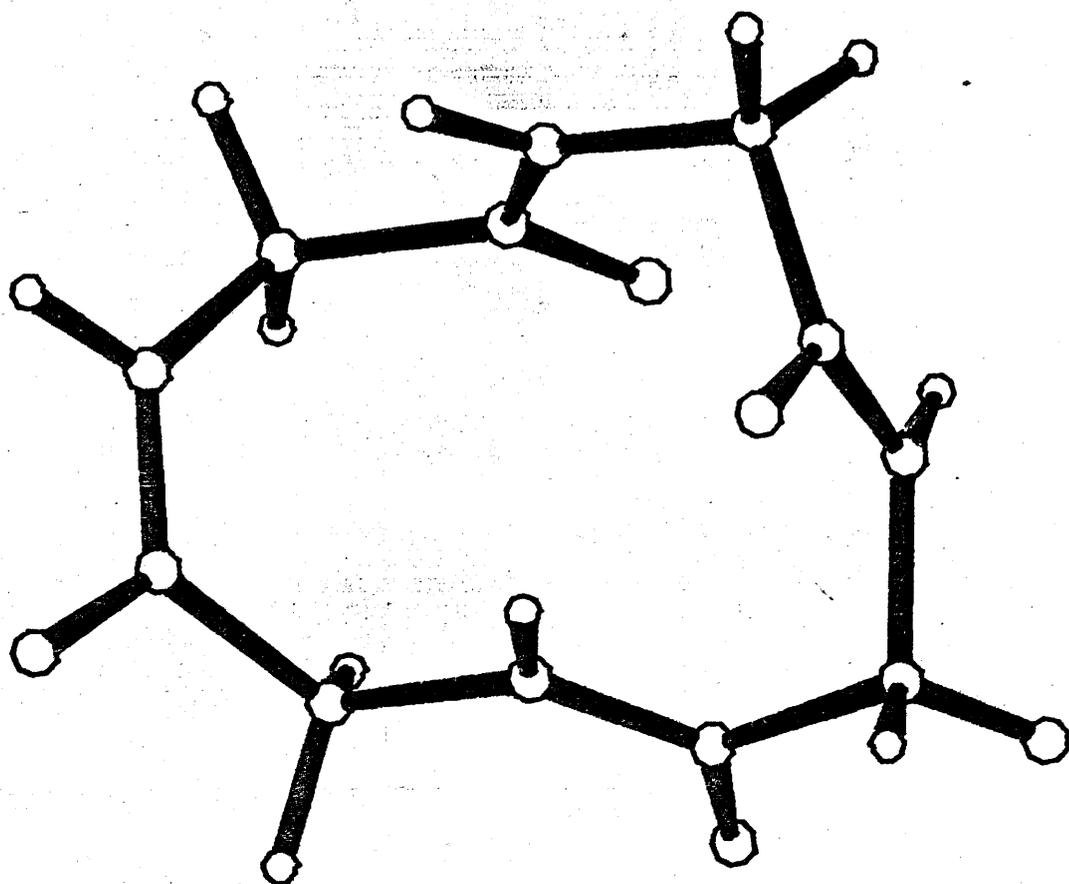
	Φ_n	Ψ_n	ω_n
n			
1	-167	79	173
2	-70	-53	5
3	-73	82	174
4	95	-59	-155

Figure 4.14 GL2-6251 CTTT

ATOM	X	Y	Z
C'1	2.16285	-0.49931	-1.25894
O1	1.47790	-1.30877	-1.90206
N1	1.64221	0.26521	-0.31180
H1	2.23868	0.87974	0.20536
CA1	0.22076	0.26747	0.07651
H11	-0.34812	-0.47290	-0.49269
H12	0.15765	-0.02603	1.12937
C'2	-0.41151	1.63338	-0.08195
O2	-0.68356	2.27830	0.94231
N2	-0.60424	2.14033	-1.29228
H2	-1.01646	3.05166	-1.34707
CA2	-0.31724	1.47933	-2.58120
H21	-0.88935	1.98354	-3.36687
H22	-0.66980	0.44463	-2.55329
C'3	1.13530	1.48353	-3.00833
O3	1.48959	0.73779	-3.93371
N3	1.98771	2.27939	-2.38172
H3	1.65165	2.88957	-1.66366
CA3	3.42807	2.36601	-2.68038
H31	3.70099	1.71442	-3.51506
H32	3.64685	3.39162	-2.99486
C'4	4.28130	2.03934	-1.47378
O4	4.88739	2.96550	-0.91351
N4	4.31842	0.80050	-1.00161
H4	4.88908	0.63435	-0.19526
CA4	3.63967	-0.38094	-1.57091
H41	4.12727	-1.28258	-1.18606
H42	3.76924	-0.39504	-2.65658

Orthogonal Coordinates of: GL4-6221

AMIDE	CHI(C)	CHI(N)	TAU
1	-0.16	-1.40	178.33
2	3.26	-2.34	0.21
3	-0.16	-1.39	178.33
4	3.26	-2.34	0.21

Planarity of the amide groups of: GL4-6221
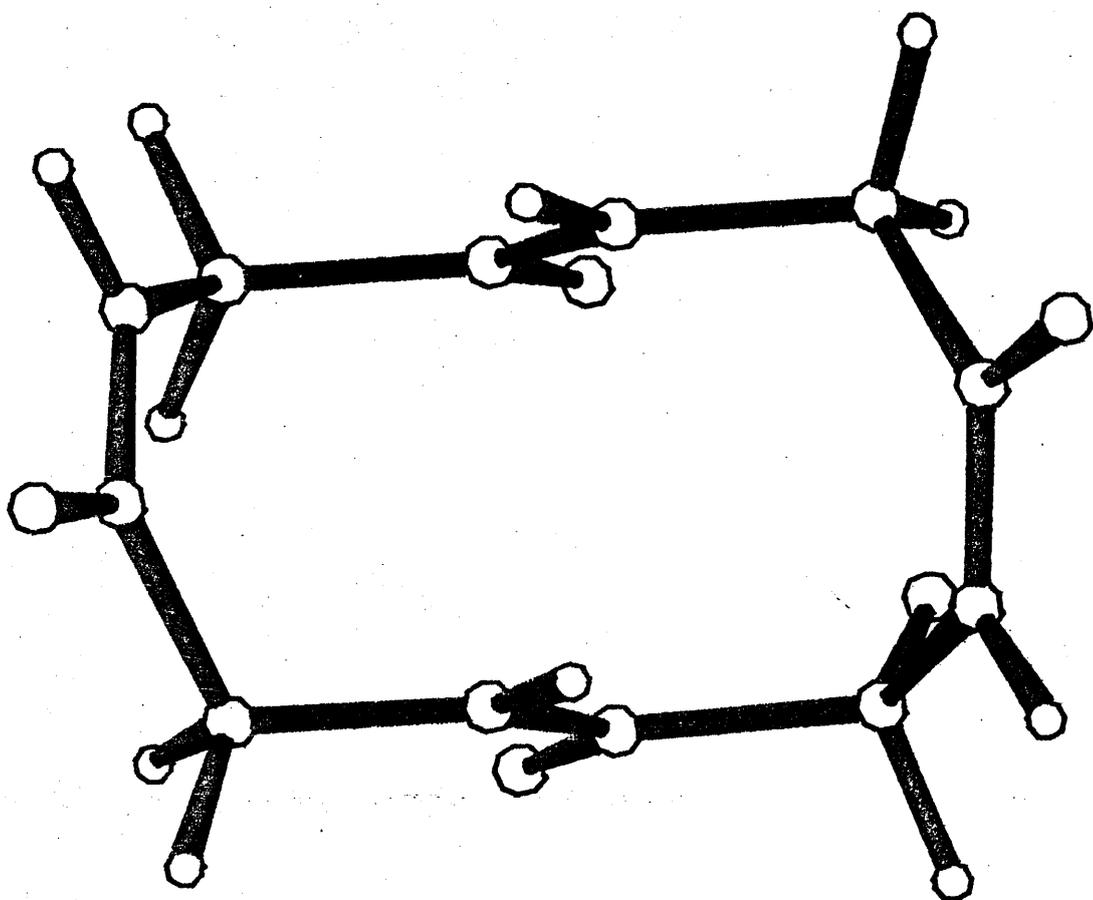
	ϕ_n	ψ_n	ω_n
n			
1	−121	69	179
2	−79	−14	3
3	−121	69	179
4	−79	−14	3

Figure 4.15 GL4-6221 CTCT C2

ATOM	X	Y	Z
C'1	0.71391	0.13768	2.42029
O1	0.37883	-0.10243	3.59039
N1	-0.10882	-0.12725	1.41138
H1	-0.98750	-0.54577	1.64771
CA1	0.14990	0.05379	-0.03138
H11	1.13792	-0.34711	-0.27549
H12	-0.57678	-0.53422	-0.60174
C'2	0.06318	1.48454	-0.51427
O2	1.03107	1.97224	-1.11665
N2	-1.02102	2.19394	-0.24197
H2	-1.77610	1.75969	0.25165
CA2	-1.14144	3.63193	-0.54894
H21	-2.10851	4.02427	-0.21715
H22	-1.06461	3.78433	-1.63016
C'3	-0.01857	4.35517	0.15999
O3	0.92209	4.83826	-0.48795
N3	0.00107	4.22926	1.47651
H3	-0.76549	3.76268	1.92107
CA3	1.16883	4.49270	2.33337
H31	1.98174	4.96898	1.77630
H32	0.88649	5.14824	3.16335
C'4	1.60762	3.13935	2.84529
O4	1.49456	2.86580	4.04958
N4	1.90593	2.23513	1.92517
H4	1.95590	2.52759	0.96996
CA4	2.06046	0.79509	2.19087
H41	2.66652	0.67203	3.09417
H42	2.60018	0.30652	1.37432

Orthogonal Coordinates of: GL1-6214

AMIDE	CHI(C)	CHI(N)	TAU
1	2.71	-3.73	1.71
2	-2.43	4.35	176.98
3	-11.48	12.04	173.06
4	10.39	-5.43	-173.70

Planarity of the amide groups of: GL1-6214
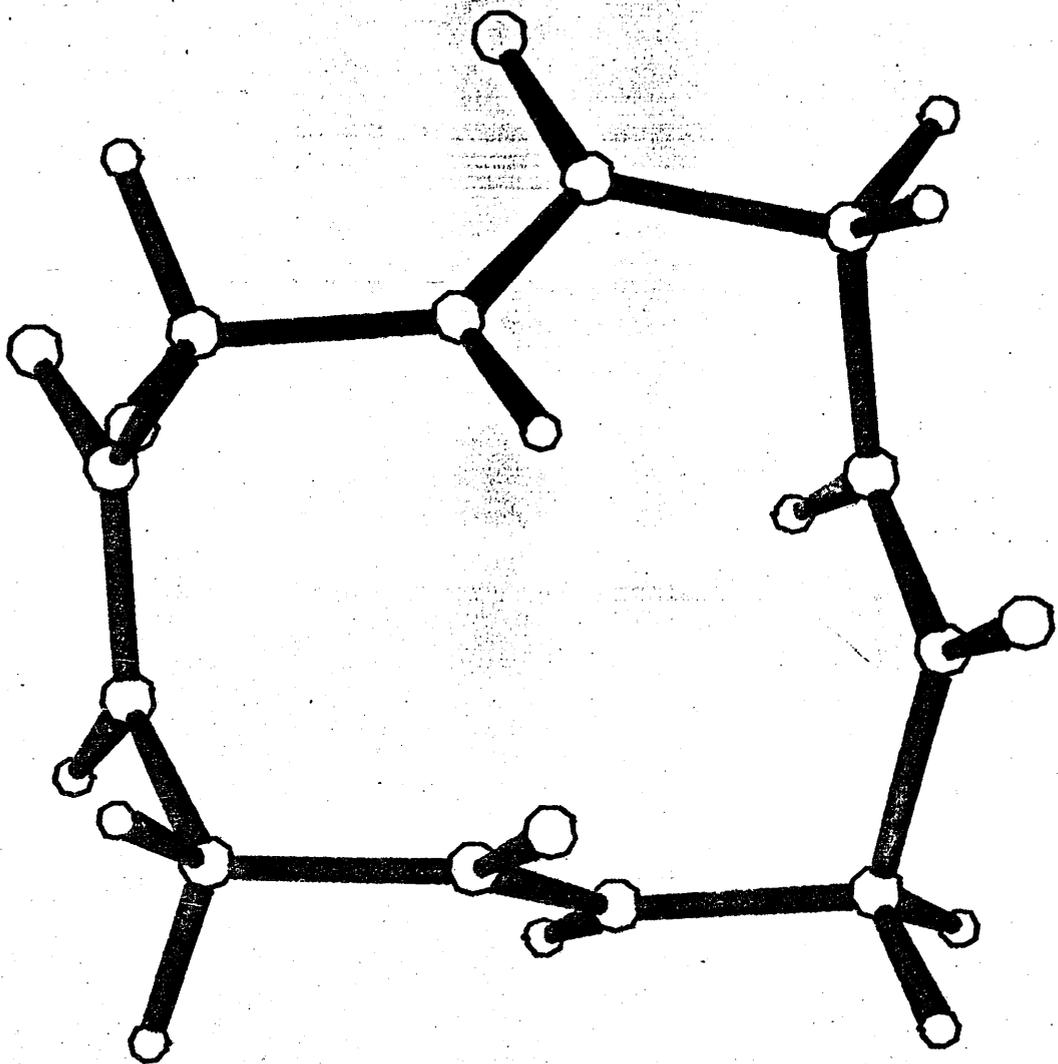
n	ϕ_n	ψ_n	ω_n
1	-77	-53	5
2	-58	-60	174
3	-110	57	161
4	74	76	-166

Figure 4.16 GL1-6214 CTTT

ATOM	X	Y	Z
C'1	2.26996	0.78402	-0.14406
O1	1.85344	0.77163	-1.31225
N1	1.51718	0.41251	0.88045
H1	1.92818	0.34816	1.79064
CA1	0.11308	-0.01817	0.77067
H11	-0.33141	0.33954	-0.16268
H12	0.09861	-1.11226	0.73859
C'2	-0.72901	0.43659	1.94291
O2	-1.06023	-0.40279	2.79427
N2	-1.04621	1.71707	2.06975
H2	-1.62918	1.97047	2.84385
CA2	-0.70077	2.80092	1.12980
H21	-1.17169	3.72913	1.46997
H22	-1.11930	2.57242	0.14487
C'3	0.78077	3.07602	0.99556
O3	1.24139	3.37144	-0.11685
N3	1.52817	3.01291	2.08684
H3	1.08100	2.80938	2.95911
CA3	2.97779	3.28070	2.14499
H31	3.12450	4.14057	2.80665
H32	3.49010	2.43537	2.61205
C'4	3.62098	3.61311	0.81390
O4	3.88704	4.80206	0.57791
N4	3.91072	2.65285	-0.05460
H4	4.24169	2.94093	-0.95488
CA4	3.70021	1.20277	0.11993
H41	4.02872	0.89475	1.11282
H42	4.33666	0.67162	-0.59552

Orthogonal Coordinates of: GL4-2634

AMIDE	CHI(C)	CHI(N)	TAU
1	1.66	-4.75	176.80
2	2.95	-4.45	-1.15
3	1.25	-3.01	179.34
4	1.34	-5.18	-4.54

Planarity of the amide groups of: GL4-2634
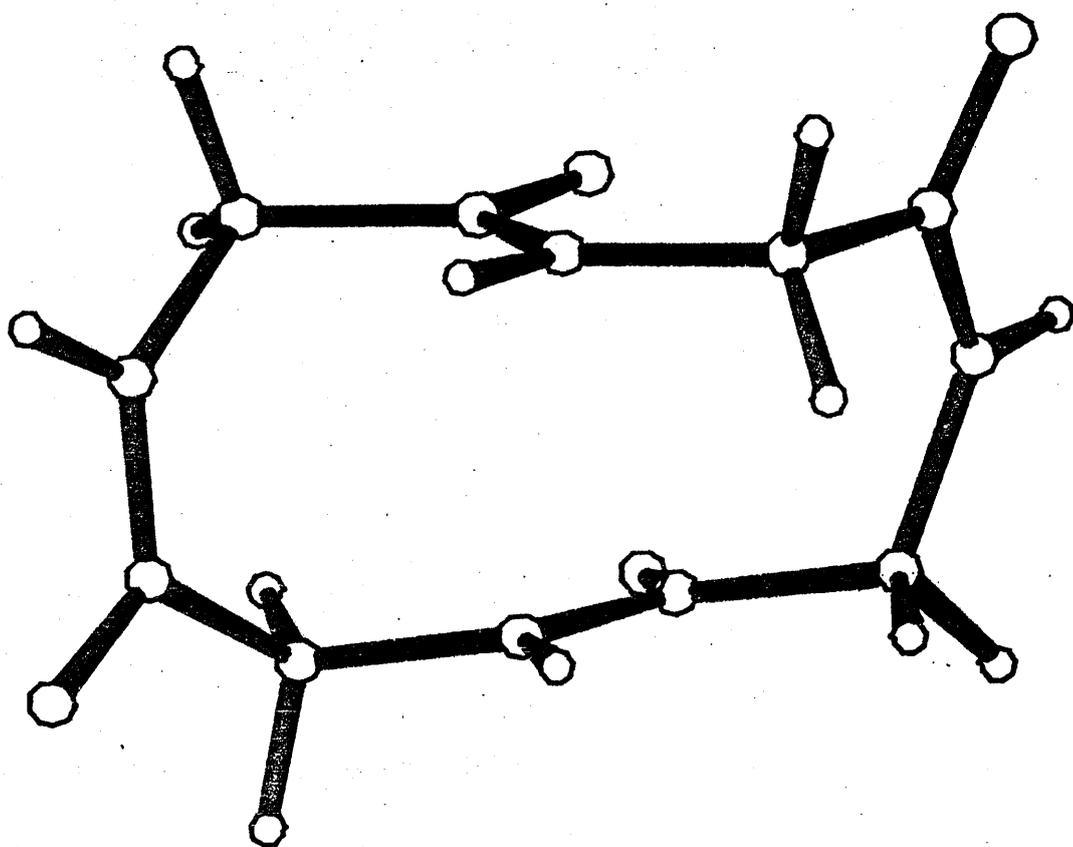
n	ϕ_n	ψ_n	ω_n
1	−144	72	180
2	−64	−39	3
3	−2	−78	−179
4	80	−109	−1

Figure 4.17 GL4-2634 CTCT

ATOM	X	Y	Z
C'1	1.91173	-1.34274	0.37780
O1	1.23712	-2.36077	0.16396
N1	1.42138	-0.12000	0.23910
H1	2.00879	0.65788	0.46299
CA1	0.00761	0.17926	-0.05323
H11	-0.08667	1.15063	-0.54963
H12	-0.40725	-0.58234	-0.72116
C'2	-0.75417	0.17946	1.25500
O2	-1.59523	-0.70347	1.47917
N2	-0.36055	1.04108	2.18100
H2	0.34922	1.70477	1.94434
CA2	-0.76563	0.98018	3.59733
H21	-1.79960	0.62984	3.67754
H22	-0.70892	1.97187	4.05797
C'3	0.15658	0.01417	4.31018
O3	-0.28949	-1.05827	4.74389
N3	1.45166	0.29196	4.29026
H3	1.74728	1.16055	3.89222
CA3	2.51194	-0.66656	4.65196
H31	3.41004	-0.13927	4.98993
H32	2.17245	-1.31172	5.46854
C'4	2.82216	-1.50842	3.43261
O4	2.54157	-2.71608	3.42922
N4	3.23397	-0.86890	2.34807
H4	3.40757	0.11405	2.41078
CA4	3.28563	-1.46674	1.00134
H41	4.03266	-0.95936	0.38218
H42	3.56541	-2.52283	1.06967

Orthogonal Coordinates of: GL1-1111

AMIDE	CHI(C)	CHI(N)	TAU
1	8.44	-8.50	-173.94
2	8.53	-8.52	-173.96
3	8.51	-8.52	-173.96
4	8.53	-8.53	-173.96

Planarity of the amide groups of: GL1-1111
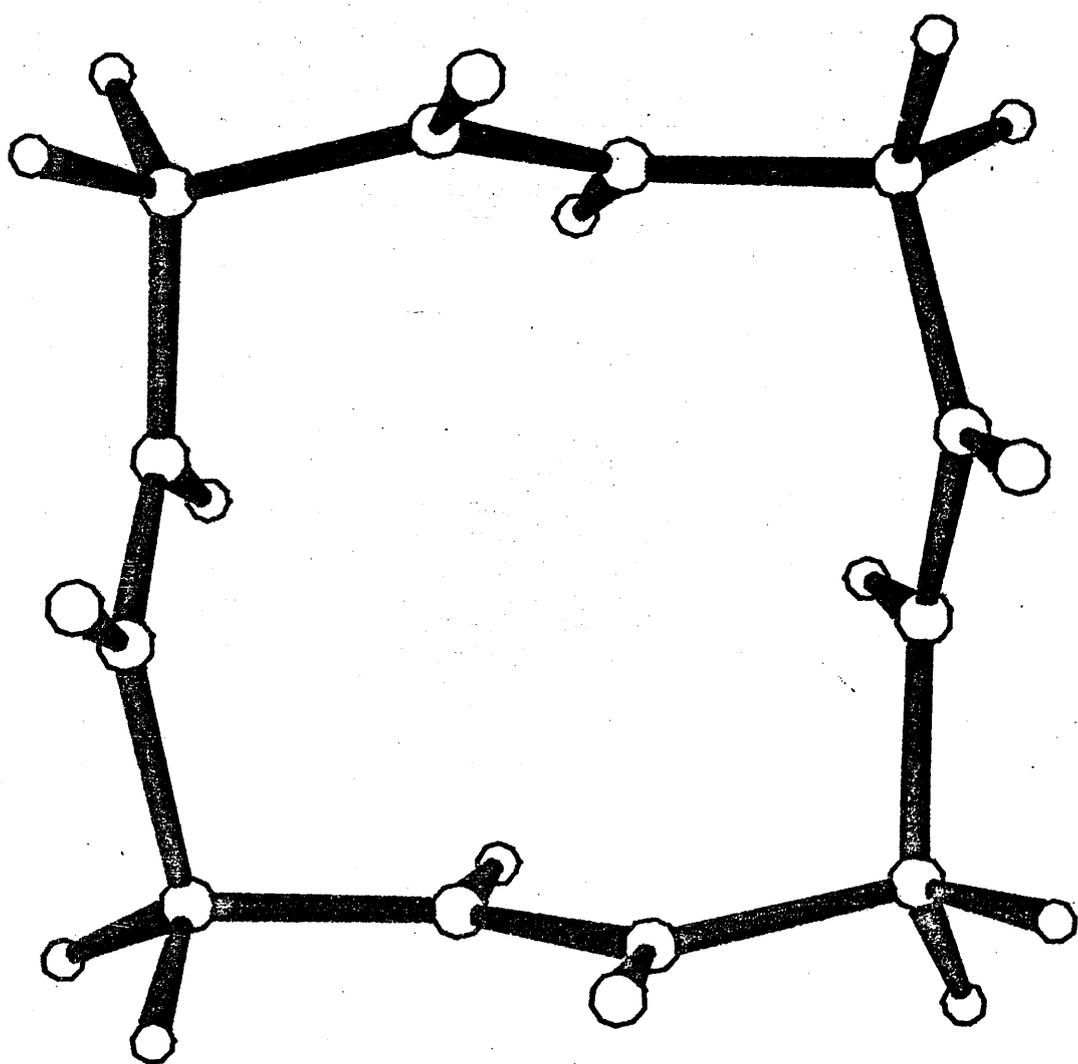
n	ϕ_n	ψ_n	ω_n
1	84	59	-165
2	84	59	-165
3	84	59	-165
4	84	59	-165

Figure 4.18 GL1-1111 TTTT C4

ATOM	X	Y	Z
C'1	2.55726	1.17401	0.00664
O1	3.77534	0.98793	0.15273
N1	1.69679	0.16517	0.07920
H1	2.07752	-0.74362	0.25928
CA1	0.22604	0.22608	-0.03584
H11	-0.08799	1.21556	-0.36480
H12	-0.09923	-0.48403	-0.80242
C'2	-0.43391	-0.15365	1.27484
O2	0.26647	-0.70163	2.13966
N2	-1.71620	0.10613	1.49384
H2	-2.10678	-0.20226	2.36317
CA2	-2.63599	0.80440	0.57651
H21	-3.64330	0.82857	1.00506
H22	-2.68971	0.25443	-0.36838
C'3	-2.19417	2.22588	0.31538
O3	-2.12223	2.63602	-0.85261
N3	-1.75711	2.93437	1.34439
H3	-1.79677	2.53893	2.26313
CA3	-1.04798	4.21581	1.19430
H31	-1.72750	4.96657	0.77869
H32	-0.68953	4.57788	2.16361
C'4	0.12677	3.99958	0.26684
O4	0.20123	4.62907	-0.79925
N4	0.99623	3.06044	0.60850
H4	0.91221	2.62152	1.50407
CA4	2.03703	2.58971	-0.26210
H41	2.93639	3.26865	-0.13629
H42	1.77183	2.65000	-1.30863

Orthogonal Coordinates of: GL2-6664

AMIDE	CHI(C)	CHI(N)	TAU
1	-0.23	1.95	-0.88
2	0.76	-0.29	-2.74
3	-7.61	8.27	175.08
4	-4.47	1.10	173.95

Planarity of the amide groups of: GL2-6664
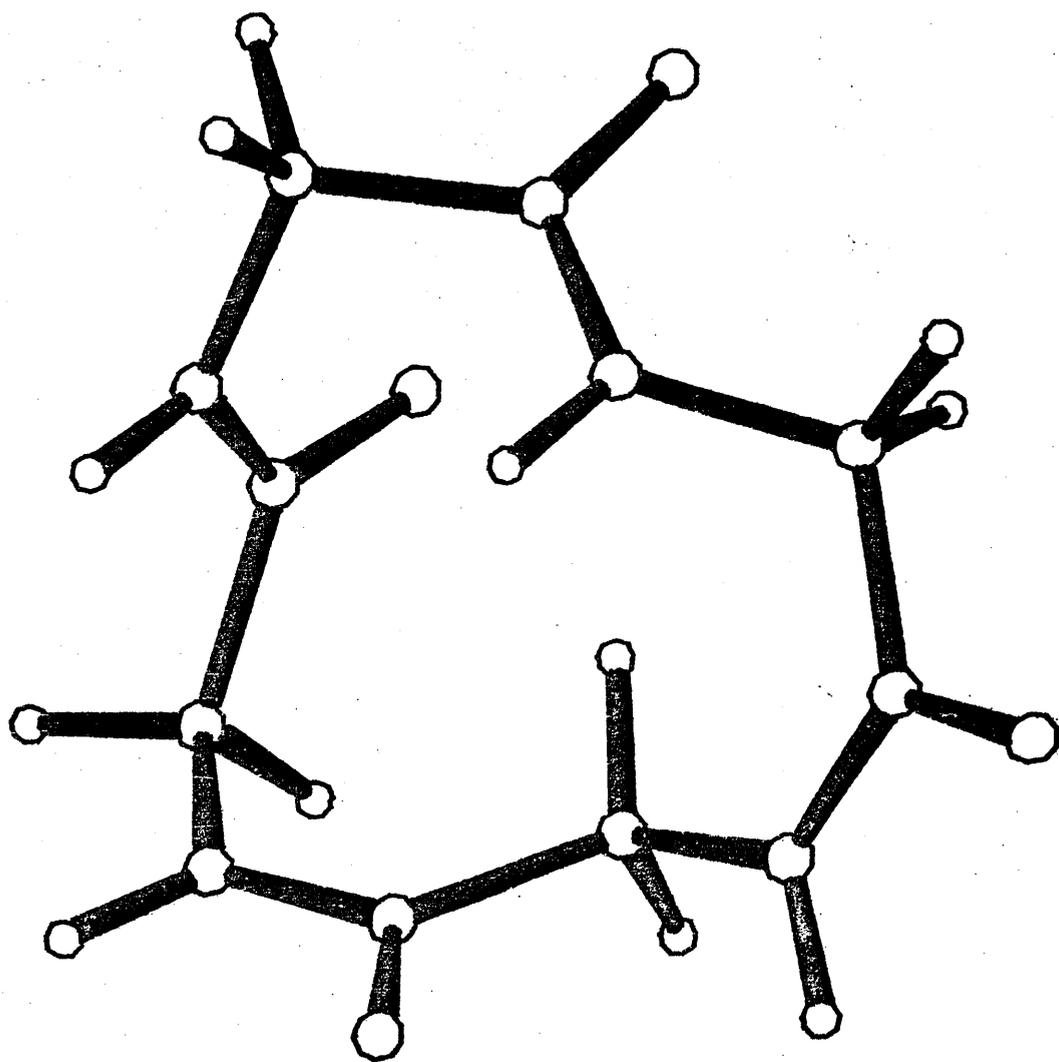
n	ϕ_n	ψ_n	ω_n
1	-114	165	-2
2	-62	-42	-2
3	-54	-56	167
4	-154	51	171

Figure 4.19 GL2-6664 CCTT

ATOM	X	Y	Z
C'1	2.03671	-0.57172	0.98120
O1	1.46708	-1.60356	1.36654
N1	1.44423	0.30896	0.18680
H1	1.98261	1.07100	-0.17280
CA1	0.05932	0.20234	-0.30843
H11	-0.18905	1.07978	-0.91448
H12	0.00687	-0.66643	-0.97239
C'2	-0.99653	0.05554	0.76859
O2	-1.75581	-0.92430	0.71213
N2	-1.14763	0.99194	1.69759
H2	-1.81059	0.80406	2.42430
CA2	-0.33275	2.21023	1.87287
H21	-0.95790	3.01637	2.27010
H22	0.07213	2.54704	0.91467
C'3	0.81008	1.94417	2.82434
O3	0.92637	0.80023	3.28651
N3	1.70262	2.89428	3.06436
H3	1.55873	3.81092	2.68794
CA3	2.96448	2.64135	3.78726
H31	3.72521	3.37807	3.50858
H32	2.79756	2.70496	4.86704
C'4	3.43505	1.25905	3.39260
O4	3.54621	0.36094	4.24074
N4	3.47387	1.04981	2.08590
H4	3.36011	1.84189	1.48601
CA4	3.45212	-0.27123	1.43554
H41	3.78565	-1.04521	2.13387
H42	4.12990	-0.27742	0.57591

Orthogonal Coordinates of: GL2-4113

AMIDE	CHI(C)	CHI(N)	TAU
1	0.93	5.18	-175.53
2	-3.89	7.97	2.75
3	4.99	-4.59	-173.83
4	11.82	-10.49	-169.77

Planarity of the amide groups of: GL2-4113
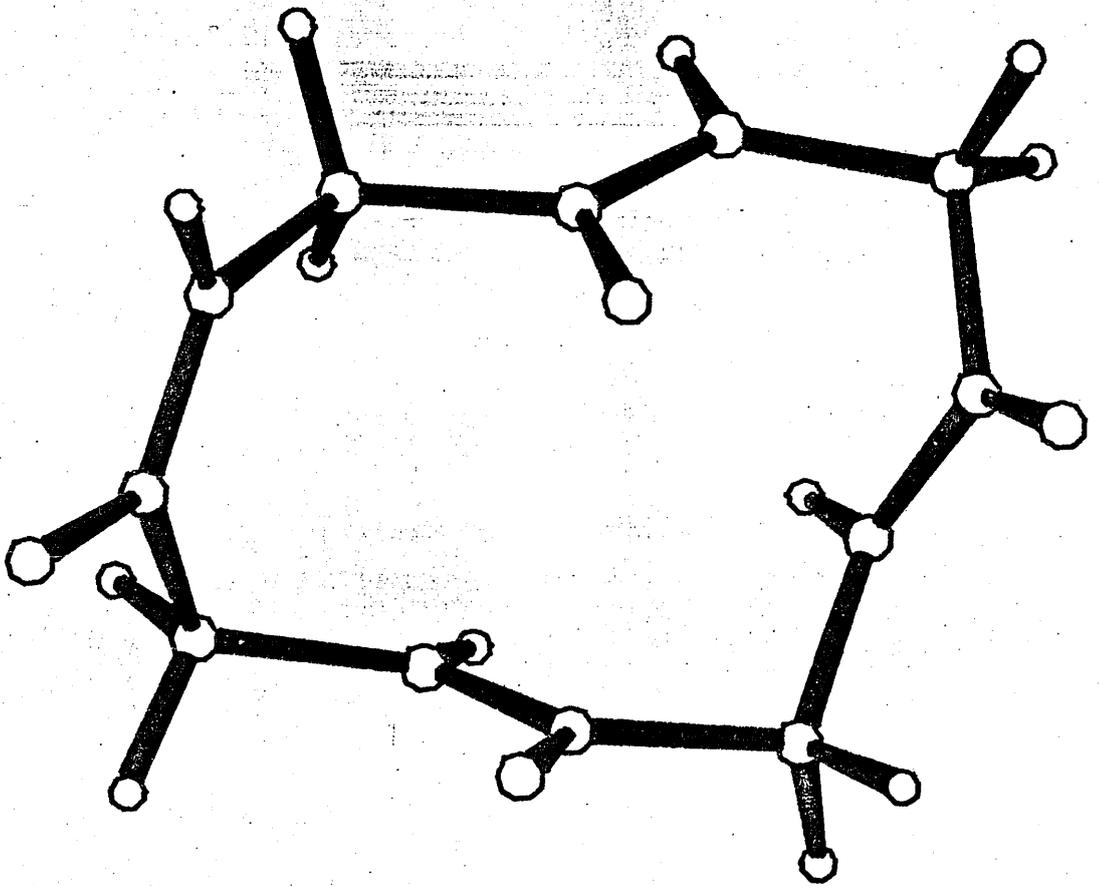
n	ϕ_n	ψ_n	ω_n
1	55	60	-178
2	-92	177	-3
3	36	51	-169
4	98	55	-159

Figure 4.20 GL2-4113 CTTT

ATOM	X	Y	Z
C'1	2.42923	-0.47047	-0.09548
O1	2.40279	-0.47612	-1.33539
N1	1.36823	-0.18504	0.64594
H1	1.44961	-0.22117	1.64337
CA1	0.04279	0.18851	0.11828
H11	0.05879	0.24699	-0.97438
H12	-0.66605	-0.60141	0.38663
C'2	-0.46212	1.50407	0.67478
O2	-1.58040	1.53451	1.21145
N2	0.30261	2.58359	0.58491
H2	-0.02588	3.42699	1.01327
CA2	1.64538	2.63772	-0.02163
H21	1.79578	3.61442	-0.49229
H22	1.69259	1.89213	-0.81412
C'3	2.72564	2.43343	1.02064
O3	2.40408	2.57382	2.21062
N3	3.96090	2.08303	0.68393
H3	4.64221	2.05252	1.41732
CA3	4.48148	1.81372	-0.66973
H31	3.69056	1.74246	-1.41697
H32	5.10569	2.66595	-0.95683
C'4	5.32420	0.55796	-0.77209
O4	6.39092	0.62744	-1.40210
N4	4.93680	-0.57634	-0.20115
H4	5.53498	-1.37079	-0.31951
CA4	3.72522	-0.79562	0.61330
H41	3.81422	-0.21745	1.53708
H42	3.67376	-1.84963	0.90547

Orthogonal Coordinates of: GL5-4232

AMIDE	CHI(C)	CHI(N)	TAU
1	-0.30	-0.12	178.34
2	0.52	4.09	1.67
3	1.82	-4.37	-4.47
4	0.33	1.08	-1.57

Planarity of the amide groups of: GL5-4232
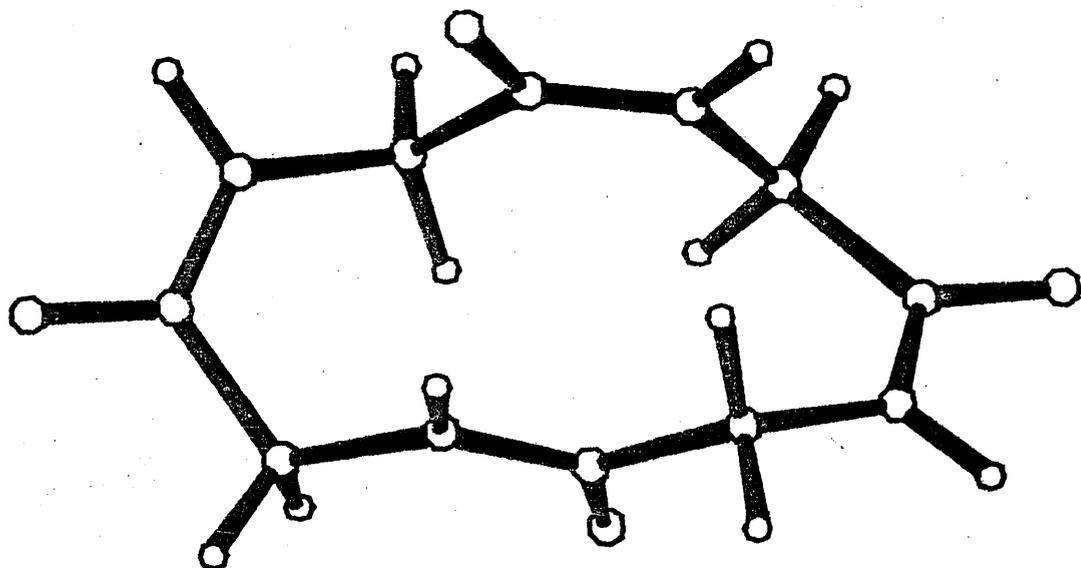
n	ϕ_n	ψ_n	ω_n
1	-125	53	178
2	-93	161	0
3	-134	45	-1
4	58	-157	-2

Figure 4.21 GL5-4232 CCCT

ATOM	X	Y	Z
C'1	1.46566	0.41324	1.59237
O1	2.31892	0.09592	0.75031
N1	0.16527	0.31343	1.36842
H1	-0.47318	0.57697	2.09329
CA1	-0.43855	-0.08894	0.08717
H11	-0.54580	-1.17824	0.08242
H12	-1.44363	0.33869	0.00765
C'2	0.34024	0.34934	-1.13585
O2	0.73076	-0.52689	-1.92309
N2	0.55720	1.63863	-1.37264
H2	1.03604	1.86287	-2.22304
CA2	0.15123	2.79152	-0.54043
H21	-0.15494	2.46057	0.44968
H22	-0.71946	3.26545	-1.00350
C'3	1.24700	3.83641	-0.44321
O3	2.01842	3.95039	-1.40805
N3	1.38245	4.61136	0.62721
H3	2.11014	5.29927	0.59835
CA3	0.57607	4.57860	1.86384
H31	0.93139	5.35125	2.55352
H32	-0.46436	4.81157	1.61575
C'4	0.63869	3.25031	2.58421
O4	-0.40479	2.76112	3.04308
N4	1.77450	2.57407	2.52250
H4	2.55248	2.98015	2.04051
CA4	1.89494	1.14309	2.84503
H41	1.26410	0.87144	3.69746
H42	2.93184	0.89703	3.09630

Orthogonal Coordinates of: GL3-3156

AMIDE	CHI(C)	CHI(N)	TAU
1	-9.60	3.76	172.67
2	2.17	1.33	1.42
3	0.32	0.17	1.97
4	10.03	-14.44	-175.22

Planarity of the amide groups of: GL3-3156
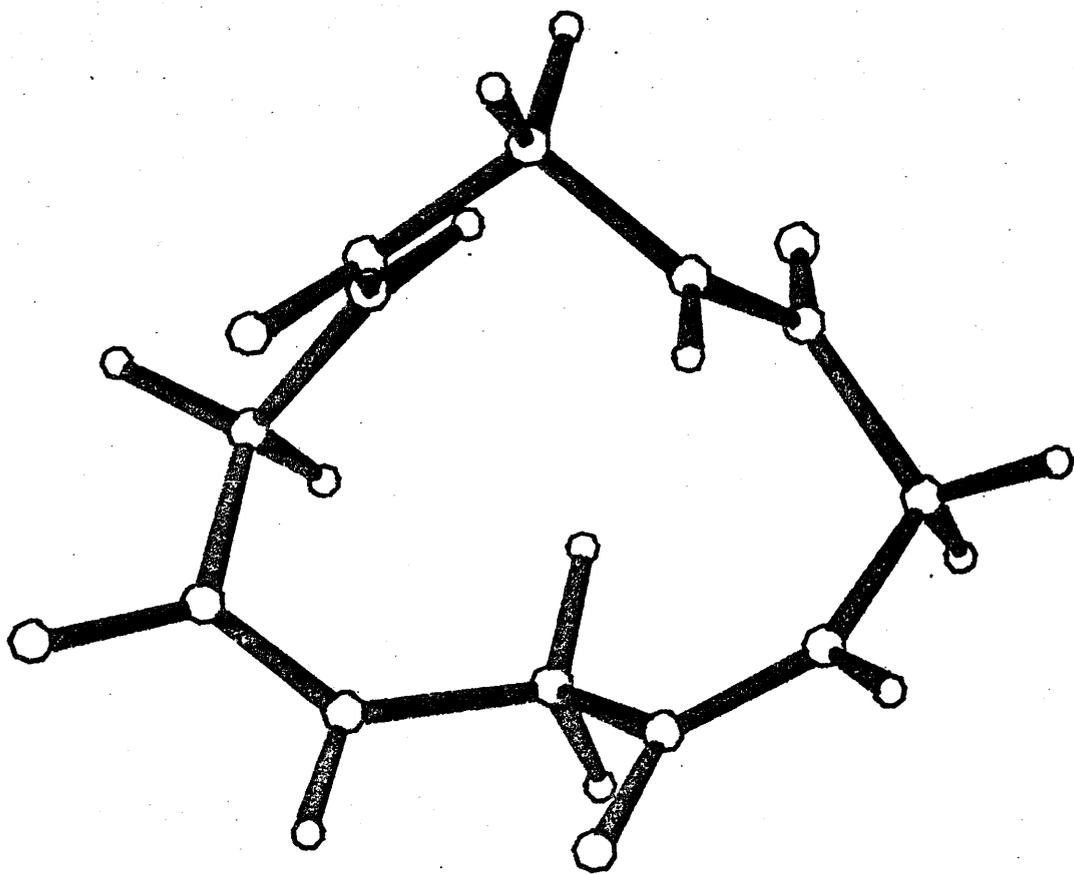
n	ϕ_n	ψ_n	ω_n
1	-33	-60	166
2	138	-150	2
3	60	33	2
4	82	-79	-163

Figure 4.22 GL3-3156 CCTT

ATOM	X	Y	Z
C'1	-1.06342	1.07954	1.71263
O1	-1.71345	1.84470	0.98460
N1	-0.27556	0.13289	1.22976
H1	0.20320	-0.47188	1.86806
CA1	-0.02259	-0.07501	-0.20643
H11	0.85029	-0.72058	-0.34945
H12	-0.88522	-0.59319	-0.63702
C'2	0.20758	1.22368	-0.95128
O2	-0.64395	1.59592	-1.77337
N2	1.30670	1.93016	-0.71764
H2	1.44140	2.76227	-1.25783
CA2	2.40997	1.57023	0.19463
H21	3.15726	1.02789	-0.39312
H22	2.07944	0.89668	0.98850
C'3	3.09135	2.76769	0.82638
O3	4.19197	3.12831	0.38159
N3	2.49781	3.40158	1.82794
H3	2.99848	4.14377	2.27655
CA3	1.19079	3.04569	2.40776
H31	0.73012	3.93200	2.85538
H32	0.54207	2.72698	1.59363
C'4	1.32090	1.97696	3.47264
O4	2.43004	1.81899	4.00507
N4	0.27914	1.22540	3.80376
H4	0.40748	0.56130	4.54249
CA4	-1.07030	1.28959	3.20995
H41	-1.51621	2.26178	3.44272
H42	-1.70621	0.51795	3.65624

Orthogonal Coordinates of: GL6-2351

AMIDE	CHI(C)	CHI(N)	TAU
1	4.97	-1.65	-174.90
2	-0.27	-4.64	-0.09
3	0.32	-4.59	-2.60
4	-1.01	1.49	2.44

Planarity of the amide groups of: GL6-2351
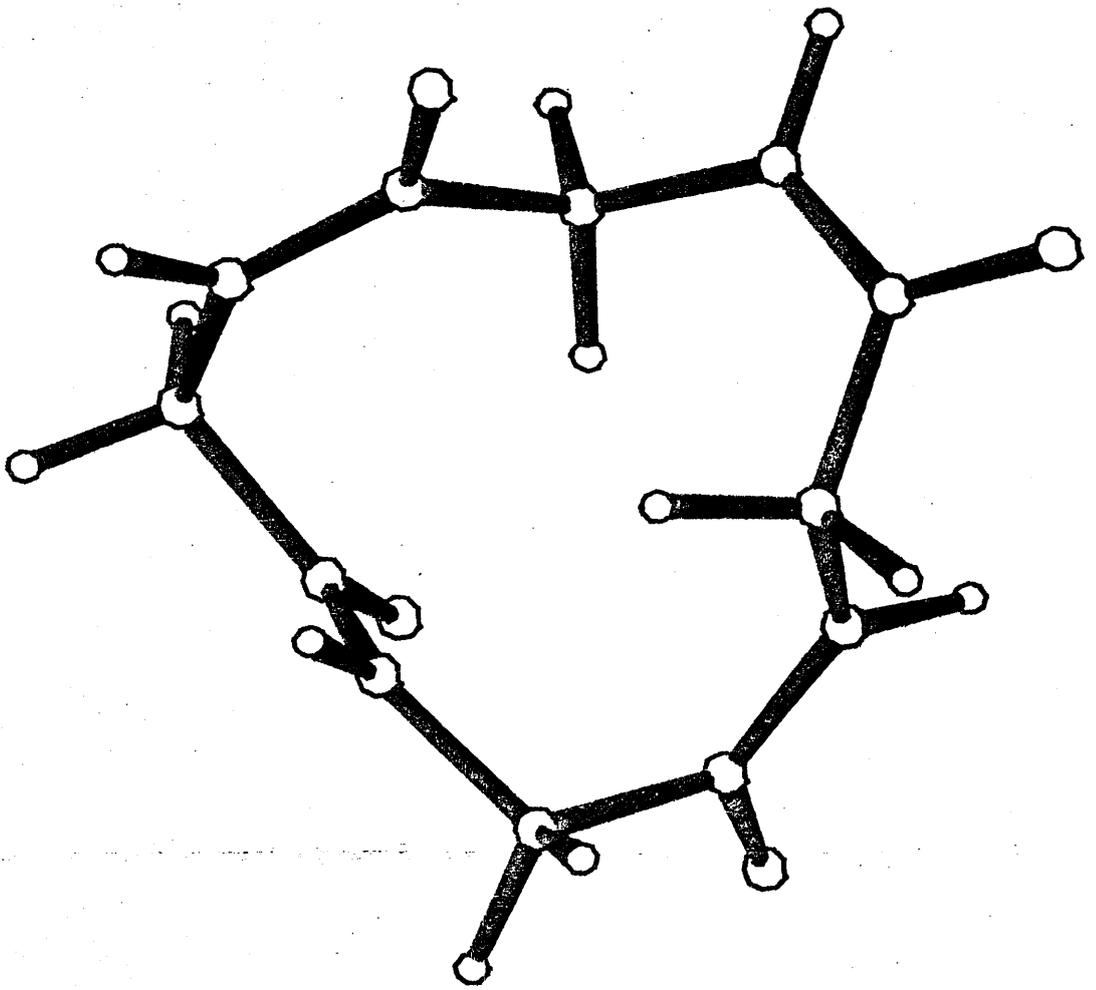
n	ϕ_n	ψ_n	ω_n
1	43	68	-172
2	-149	77	2
3	85	-158	0
4	57	42	1

Figure 4.23 GL6-2351 CCCT

ATOM	X	Y	Z
C'1	0.85378	-1.86790	1.89959
O1	-0.25070	-2.42261	2.00173
N1	1.05898	-0.81254	1.12636
H1	1.98400	-0.43707	1.04813
CA1	0.00516	-0.12157	0.36378
H11	0.21873	-0.24781	-0.70247
H12	-0.96604	-0.58778	0.55213
C'2	-0.11382	1.36024	0.66157
O2	-0.24028	2.13773	-0.29724
N2	-0.06966	1.80831	1.91066
H2	-0.18297	2.79373	2.04776
CA2	0.12973	1.00054	3.12961
H21	0.13665	1.65696	4.00597
H22	1.11978	0.54790	3.06438
C'3	-0.95565	-0.03194	3.34563
O3	-2.02149	0.11830	2.72888
N3	-0.77959	-1.04302	4.18779
H3	-1.54046	-1.68708	4.28523
CA3	0.40652	-1.33952	5.01613
H31	1.25091	-0.68104	4.80780
H32	0.12327	-1.15411	6.05716
C'4	0.84462	-2.78721	4.91046
O4	0.52415	-3.56149	5.82551
N4	1.55245	-3.20357	3.86817
H4	1.78155	-4.17749	3.82601
CA4	2.00628	-2.37618	2.73478
H41	2.66720	-2.96156	2.08722
H42	2.59214	-1.53702	3.12127

Orthogonal Coordinates of: GL5-4245

AMIDE	CHI(C)	CHI(N)	TAU
1	2.65	-0.67	-175.18
2	-0.70	0.04	2.25
3	-1.51	-0.40	0.70
4	-0.18	1.95	1.79

Planarity of the amide groups of: GL5-4245
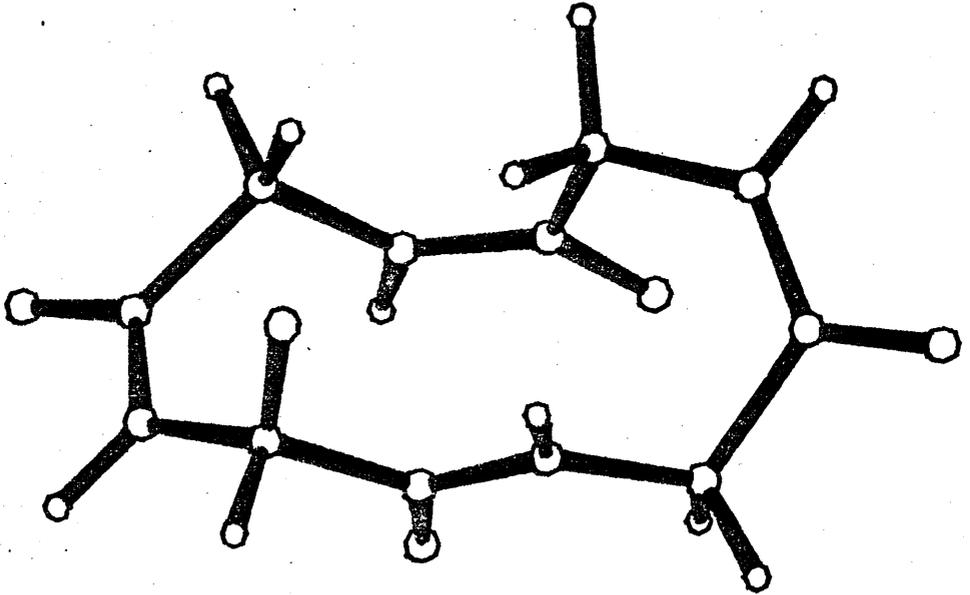
n	ϕ_n	ψ_n	ω_n
1	124	−44	−174
2	−61	164	2
3	−132	80	0
4	−66	160	0

Figure 4.24 GL5−4245 CCCT

ATOM	X	Y	Z
C'1	0.42814	-1.48281	1.63824
O1	-0.08170	-2.34630	0.90835
N1	0.61156	-0.23239	1.24423
H1	1.02824	0.42493	1.87422
CA1	0.15921	0.31915	-0.04356
H11	1.03993	0.63368	-0.61263
H12	-0.36569	-0.42999	-0.64415
C'2	-0.73658	1.51879	0.18330
O2	-0.25187	2.64662	0.00283
N2	-1.97752	1.36373	0.62879
H2	-2.53489	2.19072	0.71914
CA2	-2.68942	0.09414	0.87814
H21	-2.01005	-0.75080	1.00342
H22	-3.28926	-0.11840	-0.01182
C'3	-3.63310	0.16693	2.06171
O3	-4.82099	0.44748	1.83891
N3	-3.19214	-0.04147	3.29602
H3	-3.86986	-0.04208	4.03353
CA3	-1.81083	-0.37238	3.69263
H31	-1.12439	-0.00094	2.93406
H32	-1.56720	0.15621	4.61957
C'4	-1.64300	-1.85851	3.93891
O4	-2.60645	-2.47637	4.41747
N4	-0.48990	-2.46110	3.67877
H4	-0.44608	-3.45289	3.81112
CA4	0.70243	-1.82905	3.08535
H41	1.55408	-2.51577	3.12732
H42	0.96686	-0.94491	3.67388

Orthogonal Coordinates of: GLS-5441

AMIDE	CHI(C)	CHI(N)	TAU
1	7.30	-4.61	-174.04
2	-3.60	6.92	1.06
3	-1.16	4.81	2.19
4	-1.69	6.39	0.93

Planarity of the amide groups of: GLS-5441
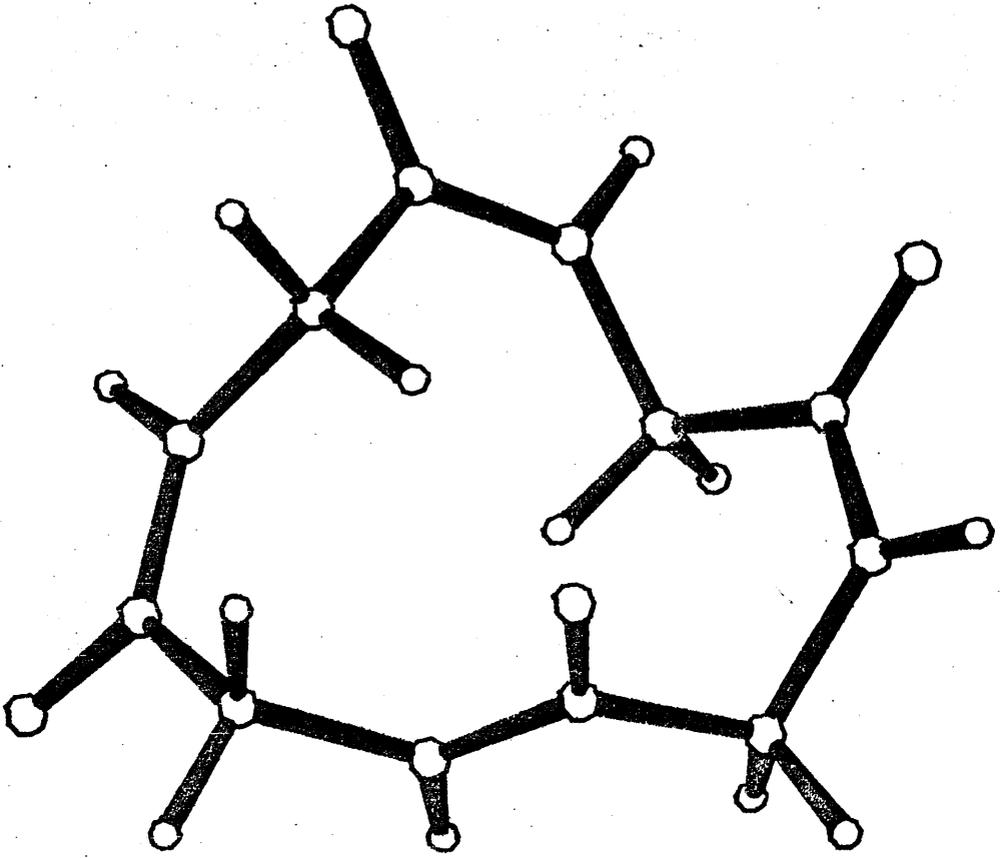
n	ϕ_n	ψ_n	ω_n
1	123	−74	−168
2	145	−85	−4
3	−99	148	−1
4	−69	125	−3

Figure 4.25 GL5-5441 CCCT

ATOM	X	Y	Z
C'1	1.15036	-0.04907	2.32998
O1	1.20124	1.18932	2.30426
N1	0.51802	-0.75867	1.40716
H1	0.57326	-1.75827	1.43068
CA1	-0.25226	-0.17367	0.29625
H11	0.37803	-0.17799	-0.59865
H12	-1.13332	-0.78461	0.07703
C'2	-0.69775	1.24216	0.58443
O2	-0.12140	2.17321	0.00040
N2	-1.66852	1.46241	1.46087
H2	-1.93710	2.41725	1.59860
CA2	-2.49584	0.45639	2.15966
H21	-3.27451	0.13997	1.45845
H22	-3.01036	0.94429	2.99413
C'3	-1.83917	-0.78936	2.71957
O3	-2.08029	-1.88192	2.18277
N3	-1.05589	-0.70394	3.78550
H3	-0.77093	-1.56019	4.21817
CA3	-0.69003	0.53559	4.49280
H31	-0.63096	1.37687	3.79839
H32	-1.49634	0.76167	5.19776
C'4	0.60540	0.46061	5.27778
O4	0.61148	0.94784	6.41892
N4	1.69074	-0.12126	4.77981
H4	2.49596	-0.15645	5.37438
CA4	1.84060	-0.77271	3.46334
H41	1.48780	-1.80491	3.54755
H42	2.90365	-0.81775	3.20520

Orthogonal Coordinates of: GL5-1232

AMIDE	CHI(C)	CHI(N)	TAU
1	0.02	-1.77	175.67
2	-1.13	6.72	-1.52
3	-1.97	-7.68	-5.55
4	1.09	2.50	0.24

Planarity of the amide groups of: GL5-1232
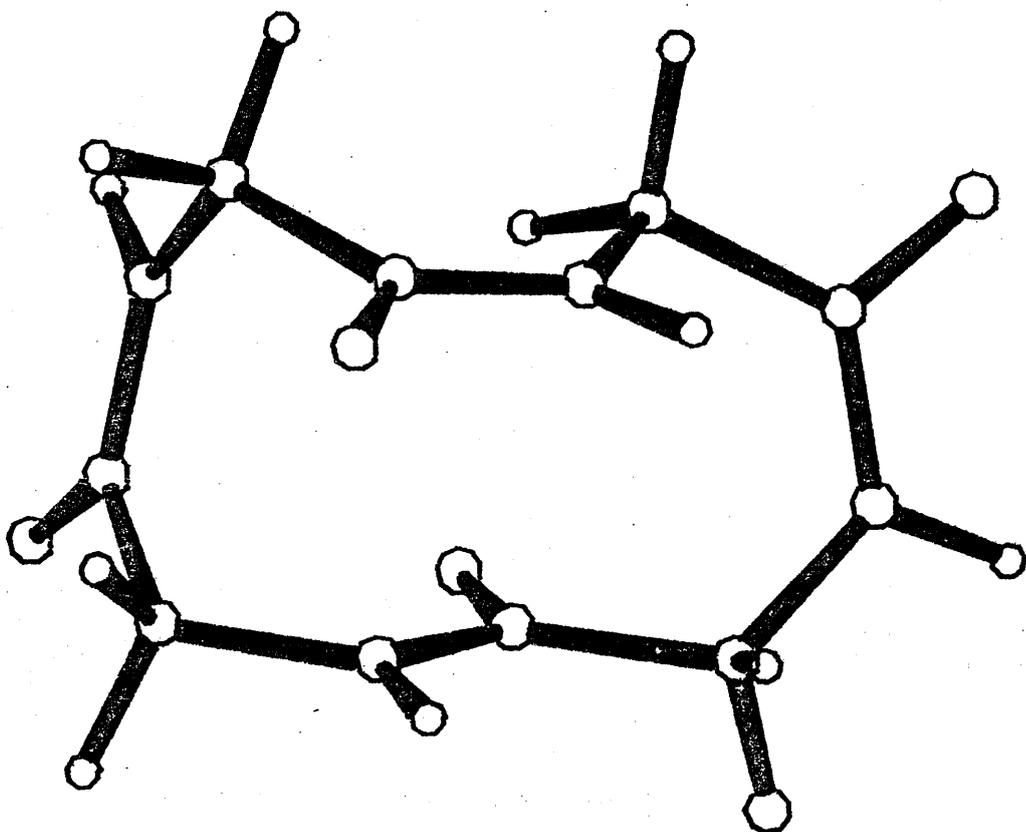
	ϕ_n	ψ_n	ω_n
n			
1	−22	−73	177
2	43	71	−5
3	−156	43	−3
4	40	−140	0

Figure 4.26 GL5−1232 CCCT

ATOM	X	Y	Z
C'1	2.36500	0.95067	-1.00128
O1	3.21813	0.63919	-1.84690
N1	1.46298	0.06854	-0.58586
H1	1.50025	-0.84597	-0.99217
CA1	0.43235	0.27265	0.45016
H11	-0.19851	-0.61910	0.52440
H12	0.95611	0.37393	1.39960
C'2	-0.47801	1.44961	0.18017
O2	-0.50307	1.90336	-0.97391
N2	-1.23236	1.96600	1.14260
H2	-1.82640	2.72836	0.87857
CA2	-1.38229	1.49683	2.53622
H21	-2.08512	0.65789	2.51865
H22	-1.85737	2.28955	3.12318
C'3	-0.15243	1.06366	3.30691
O3	-0.12330	-0.08743	3.76908
N3	0.85817	1.90370	3.48183
H3	1.64838	1.57124	3.99962
CA3	0.94469	3.28259	2.96681
H31	0.01965	3.81947	3.19761
H32	1.76168	3.81121	3.46869
C'4	1.19105	3.32717	1.47602
O4	0.50420	4.08511	0.77446
N4	2.11008	2.51663	0.97256
H4	2.67816	1.97791	1.59604
CA4	2.39569	2.36971	-0.46561
H41	3.39281	2.78216	-0.65039
H42	1.69380	2.96916	-1.05206

Orthogonal Coordinates of: GL3-1622

AMIDE	CHI(C)	CHI(N)	TAU
1	0.03	2.30	-1.76
2	0.76	4.89	-1.24
3	-0.32	1.09	-0.49
4	-2.17	-1.86	175.41

Planarity of the amide groups of: GL3-1622
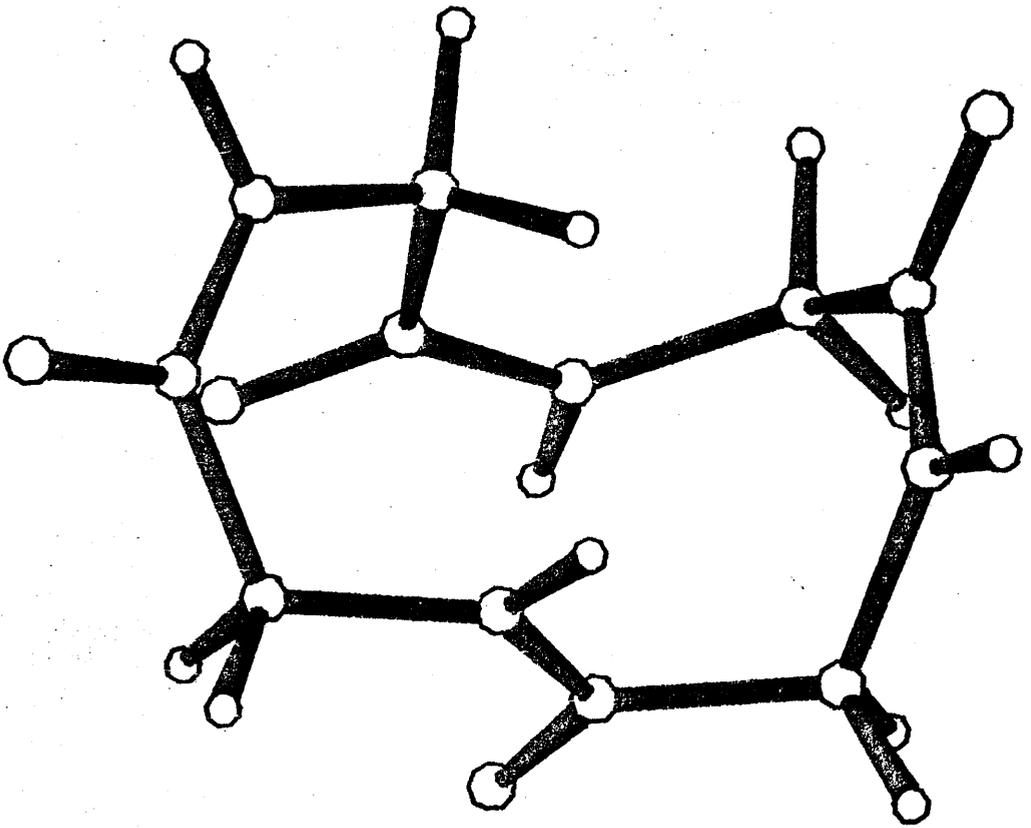
n	ϕ_n	ψ_n	ω_n
1	56	−164	−3
2	43	58	−3
3	−73	−45	−1
4	−127	39	175

Figure 4.27 GL3−1622 CCCT

ATOM	X	Y	Z
C'1	-0.37825	1.94471	-2.11521
O1	-0.62397	2.13133	-3.31713
N1	-0.20732	0.71264	-1.64633
H1	-0.23802	-0.02893	-2.31911
CA1	0.08160	0.28736	-0.26206
H11	-0.29925	1.00686	0.45967
H12	-0.45695	-0.64543	-0.06659
C'2	1.55550	0.00346	-0.06400
O2	2.18433	-0.48649	-1.01442
N2	2.14586	0.27312	1.09230
H2	3.11581	0.03867	1.17992
CA2	1.51547	0.90317	2.26805
H21	2.23040	0.92714	3.09704
H22	0.66052	0.30032	2.59061
C'3	1.07152	2.32331	1.99285
O3	-0.01472	2.72415	2.43793
N3	1.85272	3.06690	1.22554
H3	2.72587	2.68981	0.91398
CA3	1.54200	4.43263	0.76969
H31	0.52144	4.72362	1.03698
H32	2.21833	5.12257	1.28448
C'4	1.74298	4.58259	-0.72352
O4	2.72152	5.23587	-1.11748
N4	0.91792	3.98874	-1.57613
H4	1.07615	4.14791	-2.55229
CA4	-0.27647	3.18602	-1.24904
H41	-1.15418	3.80321	-1.46534
H42	-0.31905	2.95557	-0.18759

Orthogonal Coordinates of: GL4-2122

AMIDE	CHI(C)	CHI(N)	TAU
1	0.61	2.89	2.12
2	-0.02	0.73	-2.28
3	-3.03	0.39	177.10
4	2.43	-4.62	-1.08

Planarity of the amide groups of: GL4-2122
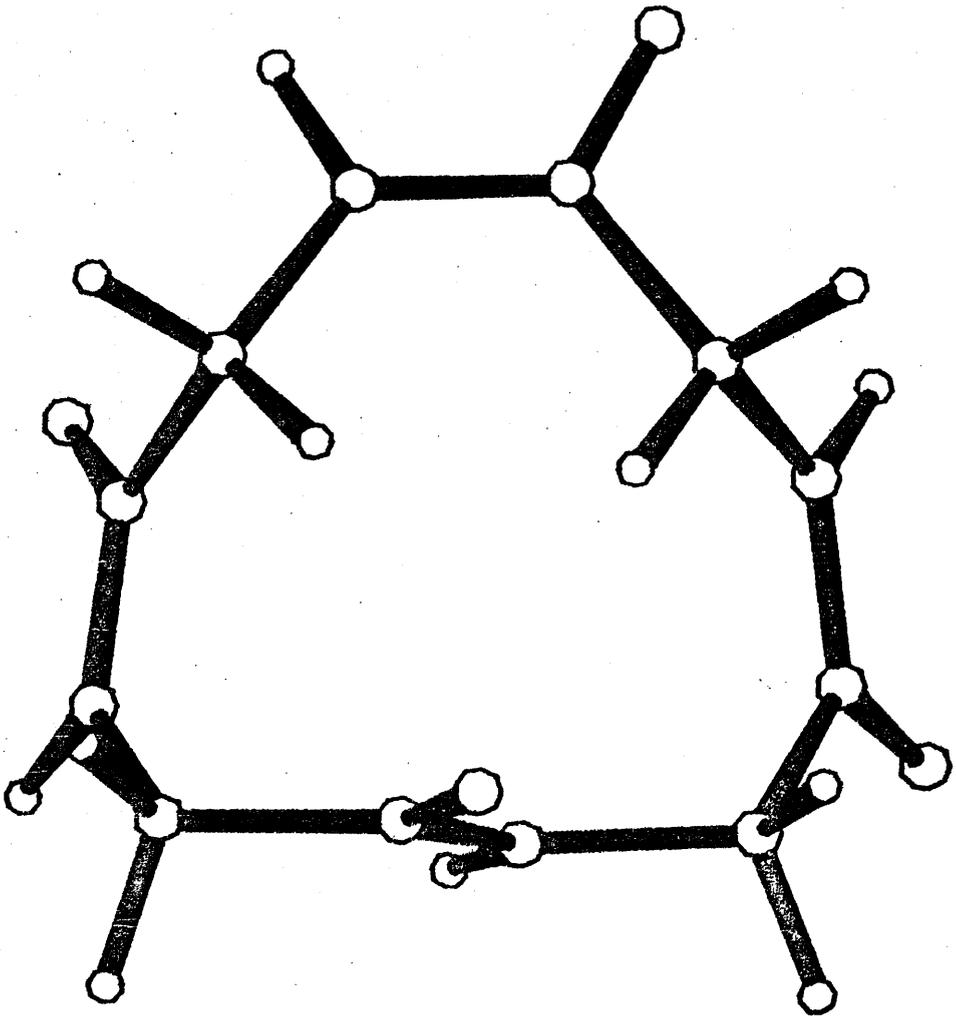
n	ϕ_n	ψ_n	ω_n
1	-99	145	1
2	-64	-37	-3
3	-130	70	175
4	-138	111	2

Figure 4.28 GL4-2122 CCCT

```
ATOM        X           Y           Z
 C'1     2.16065     -1.40586     0.66249
 O1      2.67355     -2.38782     1.22118
 N1      0.84524     -1.27682     0.54919
 H1      0.26946     -1.97034     0.98536
 CA1     0.14247     -0.15761    -0.10426
 H11     0.65250      0.04480    -1.04438
 H12    -0.88123     -0.45534    -0.35283
 C'2     0.09281      1.06027     0.79294
 O2      0.26038      0.87722     2.00719
 N2     -0.12624      2.27761     0.30607
 H2     -0.16418      3.02844     0.96822
 CA2    -0.33727      2.67334    -1.10056
 H21    -1.37002      3.02586    -1.18770
 H22    -0.23709      1.84103    -1.79935
 C'3     0.58043      3.80406    -1.52304
 O3      0.07255      4.89478    -1.82608
 N3      1.89306      3.61440    -1.55221
 H3      2.47314      4.40144    -1.76858
 CA3     2.58949      2.35759    -1.22224
 H31     2.05145      1.54728    -1.71098
 H32     3.60007      2.37060    -1.64290
 C'4     2.69608      2.15822     0.27464
 O4      2.55633      3.16223     0.99009
 N4      2.90603      0.96028     0.80710
 H4      2.97651      0.91588     1.80544
 CA4     3.07582     -0.33028     0.11041
 H41     4.10999     -0.65212     0.26965
 H42     2.94077     -0.25488    -0.96994
```

Orthogonal Coordinates of: GL4-2422


```
AMIDE    CHI(C)      CHI(N)        TAU
  1       0.48        3.51        2.54
  2       1.22       -1.22       -0.82
  3       0.49        3.51        2.54
  4       1.22       -1.22       -0.82
```

Planarity of the amide groups of: GL4-2422
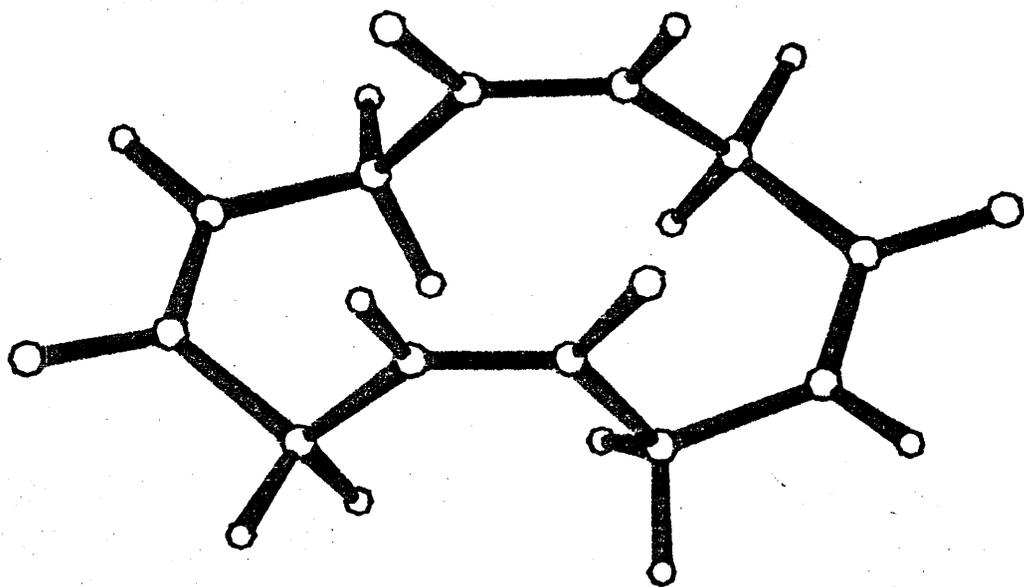
n	ϕ_n	ψ_n	ω_n
1	−79	159	1
2	−129	64	0
3	−79	159	1
4	−129	64	0

Figure 4.29 GL5-2422 CCCC   C2

ATOM	X	Y	Z
C'1	1.42437	0.49776	1.62516
O1	0.76722	1.47007	2.02564
N1	1.03247	-0.27642	0.62532
H1	1.61762	-1.04139	0.35166
CA1	-0.20961	-0.10017	-0.14906
H11	0.02468	-0.07111	-1.21809
H12	-0.83662	-0.98193	0.01812
C'2	-1.01674	1.13077	0.20440
O2	-2.11596	0.97332	0.75861
N2	-0.55543	2.33669	-0.09951
H2	-1.13631	3.11961	0.13264
CA2	0.71835	2.66231	-0.77132
H21	0.48056	2.91470	-1.80945
H22	1.38811	1.80482	-0.81249
C'3	1.36608	3.88852	-0.15574
O3	0.75781	4.95858	-0.31456
N3	2.50949	3.86798	0.52299
H3	2.81908	4.74721	0.88988
CA3	3.40570	2.73206	0.82048
H31	3.07927	1.80528	0.35424
H32	4.38443	2.96002	0.38663
C'4	3.60925	2.56436	2.31353
O4	4.07422	3.52713	2.94314
N4	3.31871	1.41808	2.91462
H4	3.45134	1.36778	3.90592
CA4	2.76561	0.21628	2.26461
H41	2.63175	-0.58153	3.00244
H42	3.48457	-0.14345	1.52147

Orthogonal Coordinates of: GL5-2441

AMIDE	CHI(C)	CHI(N)	TAU
1	2.81	-0.74	-178.61
2	-1.20	-1.28	-0.34
3	0.20	-0.12	0.92
4	-1.13	2.51	0.73

Planarity of the amide groups of: GL5-2441
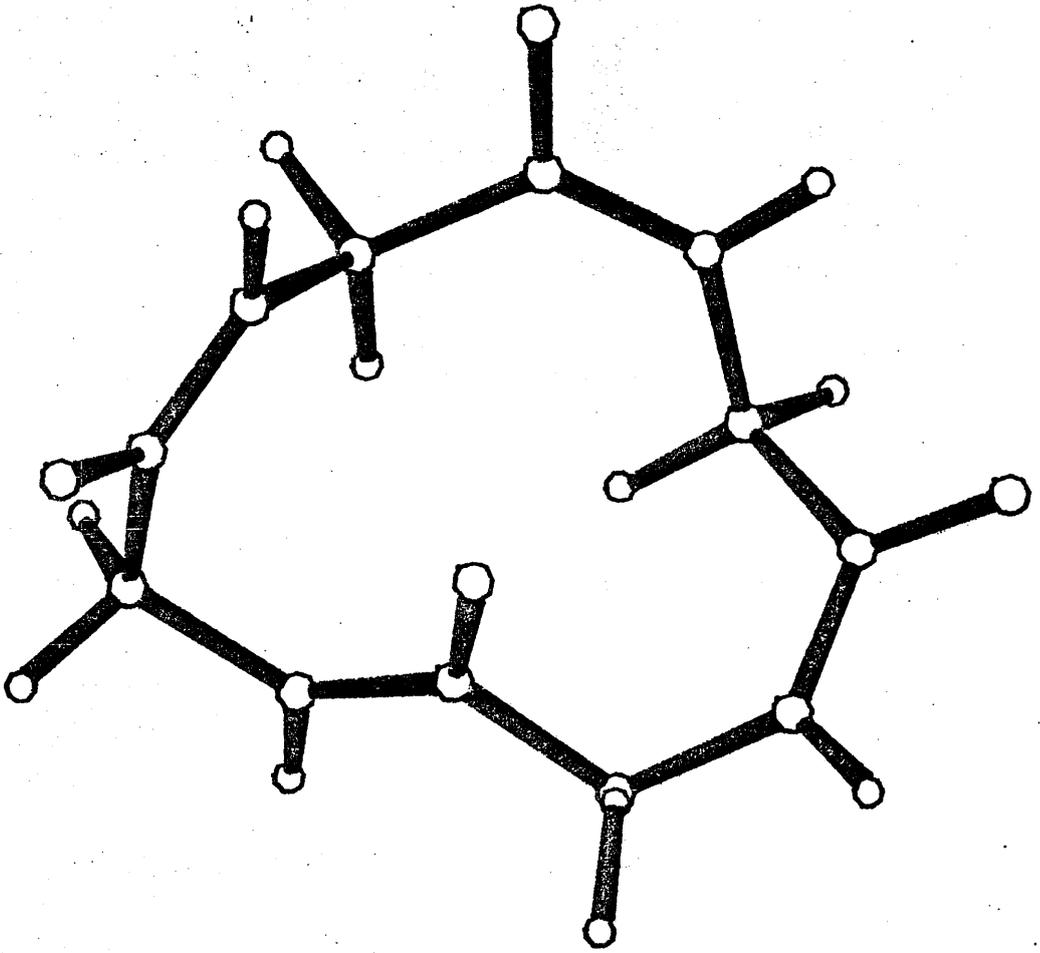
n	ϕ_n	ψ_n	ω_n
1	4	69	-177
2	-140	113	0
3	-123	123	1
4	-60	156	-1

Figure 4.30 GL5-2441 CCCT

ATOM	X	Y	Z
C'1	-1.00721	1.19653	2.21728
O1	-1.75049	1.32851	3.20230
N1	-1.13390	0.16143	1.39714
H1	-1.80017	-0.53968	1.65912
CA1	-0.32814	-0.15209	0.20016
H11	0.52558	-0.75030	0.53400
H12	-0.91244	-0.78937	-0.47130
C'2	0.19890	0.99838	-0.62889
O2	1.42020	1.21773	-0.60726
N2	-0.61680	1.73670	-1.37027
H2	-0.18786	2.41943	-1.96523
CA2	-2.07923	1.59996	-1.52160
H21	-2.25398	0.90094	-2.34561
H22	-2.50118	2.56101	-1.83263
C'3	-2.88114	1.13013	-0.32791
O3	-3.37835	-0.00603	-0.37350
N3	-3.06272	1.91571	0.72566
H3	-3.68155	1.57641	1.43701
CA3	-2.55983	3.28867	0.93151
H31	-3.30896	3.97158	0.51846
H32	-2.50640	3.49616	2.00504
C'4	-1.21491	3.65370	0.34276
O4	-1.19578	4.43858	-0.61836
N4	-0.09097	3.16493	0.85067
H4	0.76519	3.52260	0.47231
CA4	0.06283	2.24223	1.99316
H41	0.13298	2.85657	2.89638
H42	1.01565	1.71098	1.90211

Orthogonal Coordinates of: GL6-1611

AMIDE	CHI(C)	CHI(N)	TAU
1	-0.97	-5.99	-2.65
2	0.98	5.98	2.65
3	-0.98	-5.98	-2.65
4	0.98	5.98	2.65

Planarity of the amide groups of: GL6-1611
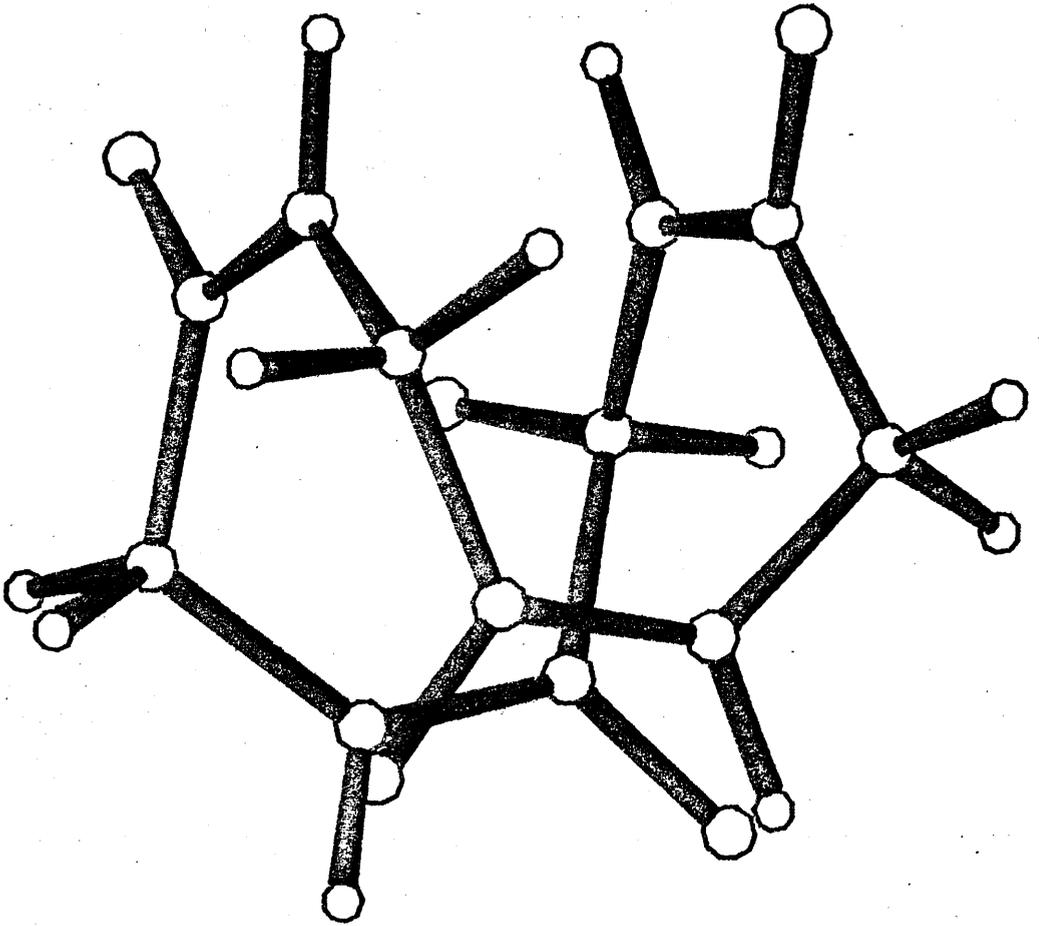
n	ϕ_n	ψ_n	ω_n
1	-34	-71	0
2	34	71	0
3	-34	-71	0
4	34	71	0

Figure 4.31 GL6-1611 CCCC S4

ATOM	X	Y	Z
C'1	2.08142	0.27540	-0.41167
O1	1.77116	1.37291	-0.89732
N1	1.21262	-0.48723	0.23185
H1	1.49906	-1.39742	0.53406
CA1	-0.18276	-0.10990	0.52613
H11	-0.83208	-0.61822	-0.19376
H12	-0.45959	-0.47842	1.52106
C'2	-0.47680	1.37663	0.49047
O2	-1.43781	1.77075	-0.18862
N2	0.25521	2.21255	1.21649
H2	0.05949	3.19070	1.13145
CA2	1.41519	1.85808	2.05713
H21	1.42799	0.77962	2.21500
H22	1.30583	2.32649	3.04014
C'3	2.70929	2.34215	1.43356
O3	2.66777	3.37556	0.74834
N3	3.84063	1.66415	1.58378
H3	4.65792	2.04334	1.14625
CA3	4.05881	0.46933	2.42113
H31	4.97220	0.65485	2.99593
H32	3.26573	0.36894	3.16552
C'4	4.24494	-0.89323	1.75869
O4	4.64579	-1.79643	2.49682
N4	3.97472	-1.13977	0.48322
H4	4.14774	-2.07861	0.17894
CA4	3.50239	-0.21803	-0.56880
H41	3.54342	-0.74160	-1.52971
H42	4.18638	0.63053	-0.65129

Orthogonal Coordinates of: GLS-3522

AMIDE	CHI(C)	CHI(N)	TAU
1	0.51	-1.23	175.67
2	2.68	-5.63	-1.43
3	-1.91	5.66	-1.73
4	1.96	-2.67	0.20

Planarity of the amide groups of: GLS-3522
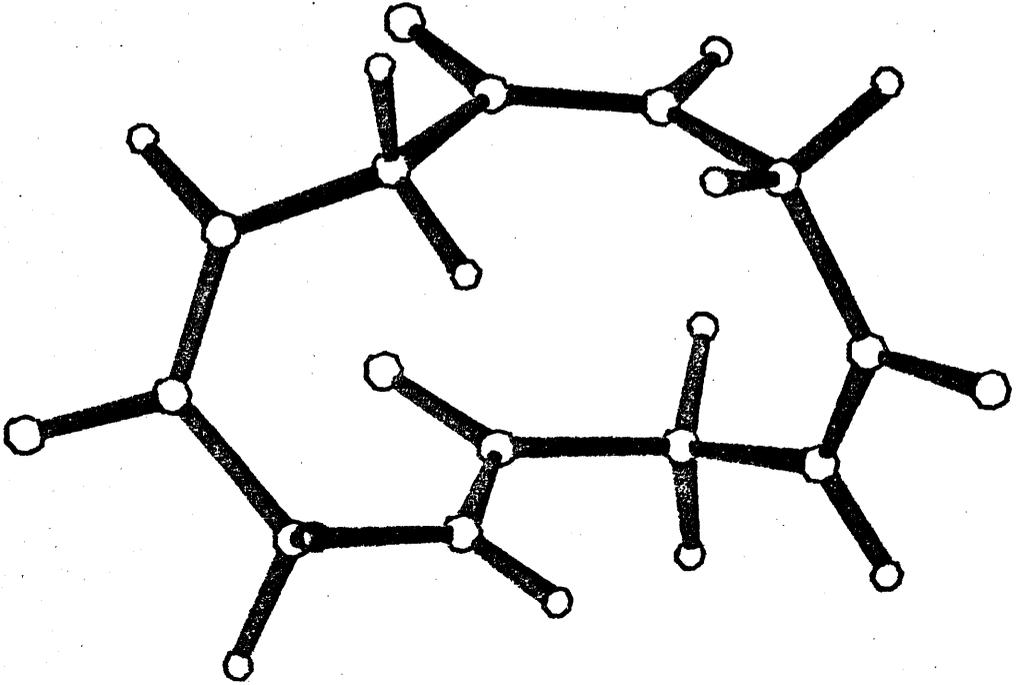
n	ϕ_n	ψ_n	ω_n
1	−20	−54	177
2	108	−147	3
3	108	−14	−5
4	−70	−27	3

Figure 4.32 GL5−3522 CCCT

ATOM	X	Y	Z
C'1	-1.95404	-0.00120	1.13803
O1	-2.74297	-0.53078	1.93601
N1	-0.67164	-0.34819	1.12560
H1	-0.41893	-1.10401	1.73299
CA1	0.43594	0.16131	0.28845
H11	1.35735	0.13275	0.87898
H12	0.56920	-0.55151	-0.53154
C'2	0.34875	1.54593	-0.31536
O2	0.16904	1.63873	-1.53981
N2	0.45014	2.62327	0.45245
H2	0.43634	3.51143	-0.01041
CA2	0.69422	2.65835	1.90789
H21	0.58815	3.68667	2.26920
H22	1.73832	2.37525	2.07454
C'3	-0.18577	1.78253	2.77416
O3	0.34464	0.83241	3.37067
N3	-1.47919	2.04649	2.90769
H3	-2.00008	1.47373	3.54346
CA3	-2.23797	3.14136	2.27312
H31	-2.38044	3.92529	3.02363
H32	-1.69071	3.59320	1.44274
C'4	-3.60281	2.69094	1.79032
O4	-4.61375	3.21832	2.27918
N4	-3.68095	1.73788	0.87190
H4	-4.58872	1.40289	0.61407
CA4	-2.52708	1.07395	0.24053
H41	-1.80739	1.85315	0.01180
H42	-2.82461	0.61969	-0.70986

Orthogonal Coordinates of: GL5-6535

AMIDE	CHI(C)	CHI(N)	TAU
1	-2.29	4.02	2.02
2	0.64	-5.51	-0.90
3	1.40	3.36	1.84
4	-0.75	-3.33	-2.51

Planarity of the amide groups of: GL5-6535
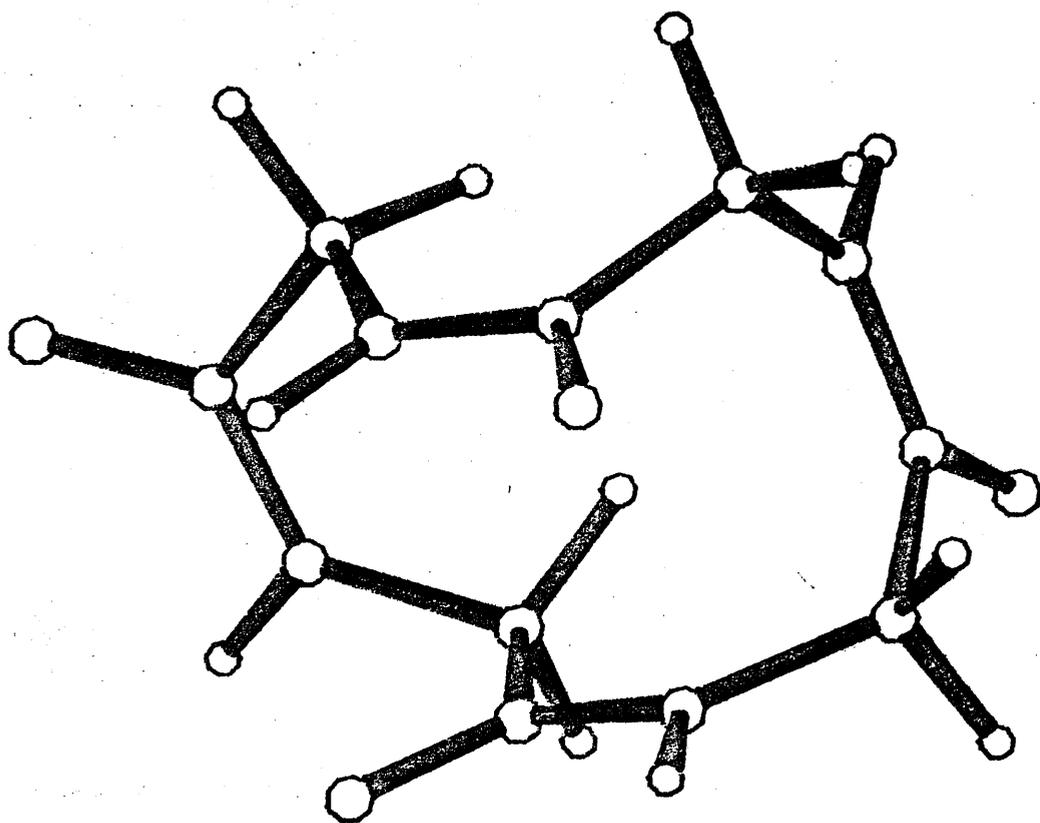
	ϕ_n	ψ_n	ω_n
n			
1	24	71	−1
2	−50	−68	2
3	139	−61	1
4	80	−150	−1

Figure 4.33 GL5-6535 CCCC

ATOM	X	Y	Z
C'1	-0.34278	0.82525	2.63007
O1	-1.02309	0.78937	3.66713
N1	-0.74954	0.22393	1.51865
H1	-1.65735	-0.19868	1.52868
CA1	-0.01377	0.14808	0.24290
H11	1.03257	-0.02861	0.47906
H12	-0.36006	-0.72074	-0.32611
C'2	-0.18549	1.37789	-0.62333
O2	-1.14527	2.12512	-0.38000
N2	0.68329	1.65475	-1.59015
H2	0.51344	2.47619	-2.13722
CA2	1.86138	0.87782	-1.96950
H21	1.72561	0.52860	-2.99535
H22	2.01045	-0.01852	-1.36109
C'3	3.18740	1.64881	-1.93564
O3	3.86776	1.68461	-2.97267
N3	3.59419	2.25007	-0.82420
H3	4.50201	2.67269	-0.83423
CA3	2.85942	2.32594	0.45155
H31	3.20472	3.19477	1.02055
H32	1.81208	2.50263	0.21538
C'4	3.03013	1.09614	1.31779
O4	3.98991	0.34890	1.07448
N4	2.16133	0.81930	2.28460
H4	2.33118	-0.00214	2.83168
CA4	0.96325	1.59623	2.66393
H41	0.83419	2.49256	2.05552
H42	1.11901	1.94546	3.68978

Orthogonal Coordinates of: GLS-2353

AMIDE	CHI(C)	CHI(N)	TAU
1	0.88	2.06	4.02
2	1.47	-0.63	-1.06
3	-0.89	-2.05	-4.02
4	-1.47	0.63	1.06

Planarity of the amide groups of: GLS-2353
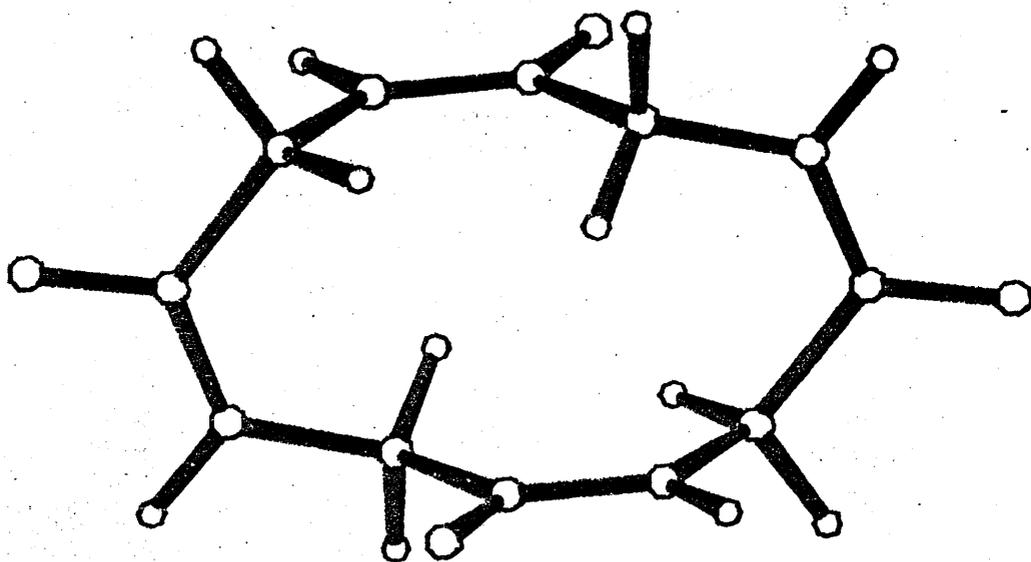
	ϕ_n	ψ_n	ω_n
n			
1	-83	160	3
2	-123	57	0
3	83	-160	-3
4	123	-57	0

Figure 4.34 GL5-2353 CCCC Ci

ATOM	X	Y	Z
C'1	-0.75503	1.18344	2.07640
O1	-1.61811	0.34796	2.38567
N1	-0.03302	1.08598	0.97471
H1	0.72864	1.72265	0.84160
CA1	-0.21218	0.05622	-0.06180
H11	0.60320	-0.66743	0.03712
H12	-1.14783	-0.49558	0.05911
C'2	-0.13055	0.68272	-1.43466
O2	0.99327	0.75037	-1.95645
N2	-1.18862	1.21927	-2.02890
H2	-1.03145	1.58093	-2.94969
CA2	-2.60348	1.23123	-1.59881
H21	-3.09804	0.41389	-2.13347
H22	-3.05436	2.15784	-1.96849
C'3	-3.00707	1.12387	-0.14206
O3	-3.68490	0.14567	0.20894
N3	-2.71216	2.10389	0.69983
H3	-3.13743	2.07892	1.60634
CA3	-1.90489	3.29313	0.38261
H31	-0.99340	3.01553	-0.15063
H32	-2.49317	3.91127	-0.30286
C'4	-1.53567	4.15877	1.57076
O4	-1.83189	5.36091	1.48808
N4	-0.91428	3.70292	2.65508
H4	-0.72363	4.38931	3.35974
CA4	-0.47148	2.33647	3.01210
H41	-0.94185	2.09836	3.96929
H42	0.60828	2.36389	3.19154

Orthogonal Coordinates of: GL5-1611

AMIDE	CHI(C)	CHI(N)	TAU
1	-2.30	3.43	-174.84
2	-4.51	7.62	0.33
3	-4.64	-5.68	-5.90
4	-0.83	1.16	-0.79

Planarity of the amide groups of: GL5-1611
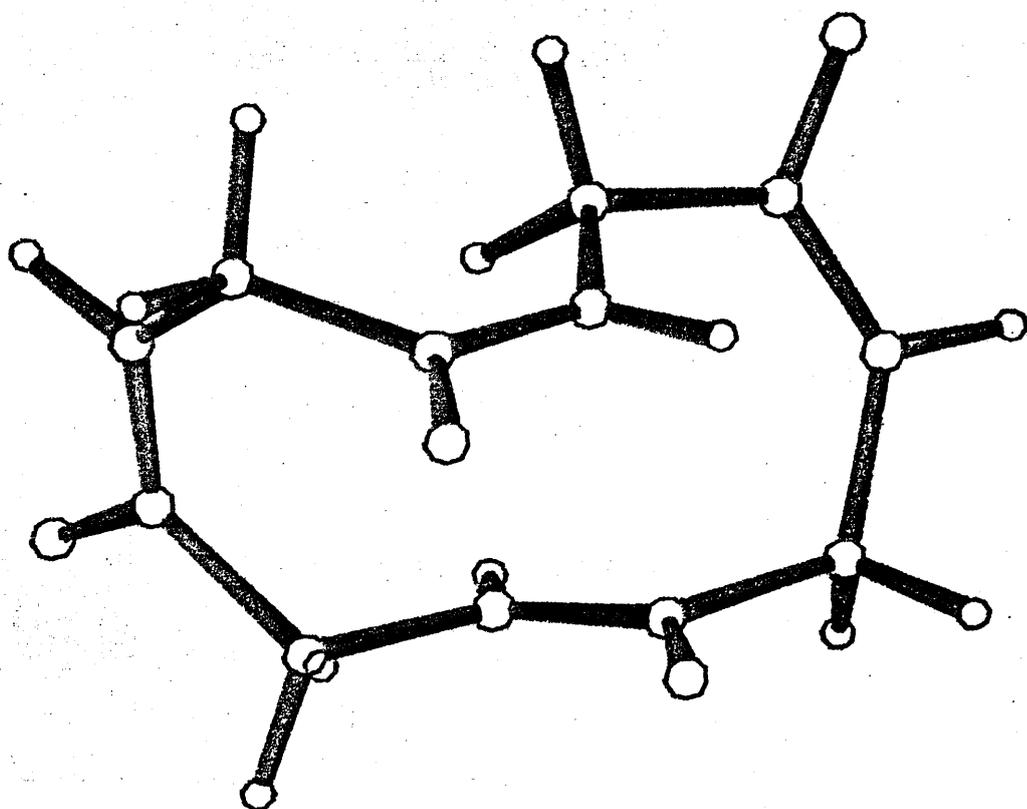
	ϕ_n	ψ_n	ω_n
n			
1	137	−86	−178
2	26	67	−6
3	−171	55	−6
4	−2	75	−2

Figure 4.35 GL5−1611 CCCT

ATOM	X	Y	Z
C'1	0.41246	1.71855	-1.94146
O1	0.82729	2.09097	-3.05029
N1	0.43648	0.43713	-1.59900
H1	0.70987	-0.20949	-2.31410
CA1	-0.03443	-0.17847	-0.34330
H11	-1.12749	-0.21769	-0.38896
H12	0.31834	-1.21436	-0.30594
C'2	0.36043	0.45226	0.97385
O2	-0.55196	0.83356	1.72381
N2	1.63243	0.57248	1.33084
H2	1.80411	0.89000	2.26568
CA2	2.83198	0.15383	0.58030
H21	2.91436	-0.93351	0.67688
H22	3.71612	0.58256	1.06341
C'3	2.93703	0.50644	-0.88710
O3	3.05276	-0.42851	-1.69503
N3	2.93226	1.76763	-1.29880
H3	3.12559	1.91917	-2.27032
CA3	2.82937	2.99162	-0.48116
H31	3.80565	3.15472	-0.01353
H32	2.64995	3.84361	-1.14511
C'4	1.77144	3.06405	0.59780
O4	2.15339	3.24526	1.76467
N4	0.48027	2.96415	0.31009
H4	-0.15815	3.14159	1.06186
CA4	-0.14552	2.77437	-1.01274
H41	-0.11108	3.73787	-1.53129
H42	-1.20307	2.52957	-0.81928

Orthogonal Coordinates of: GLS-1652

AMIDE	CHI(C)	CHI(N)	TAU
1	-1.31	-7.63	-3.27
2	1.32	7.62	3.27
3	-1.32	-7.62	-3.27
4	1.32	7.62	3.27

Planarity of the amide groups of: GLS-1652
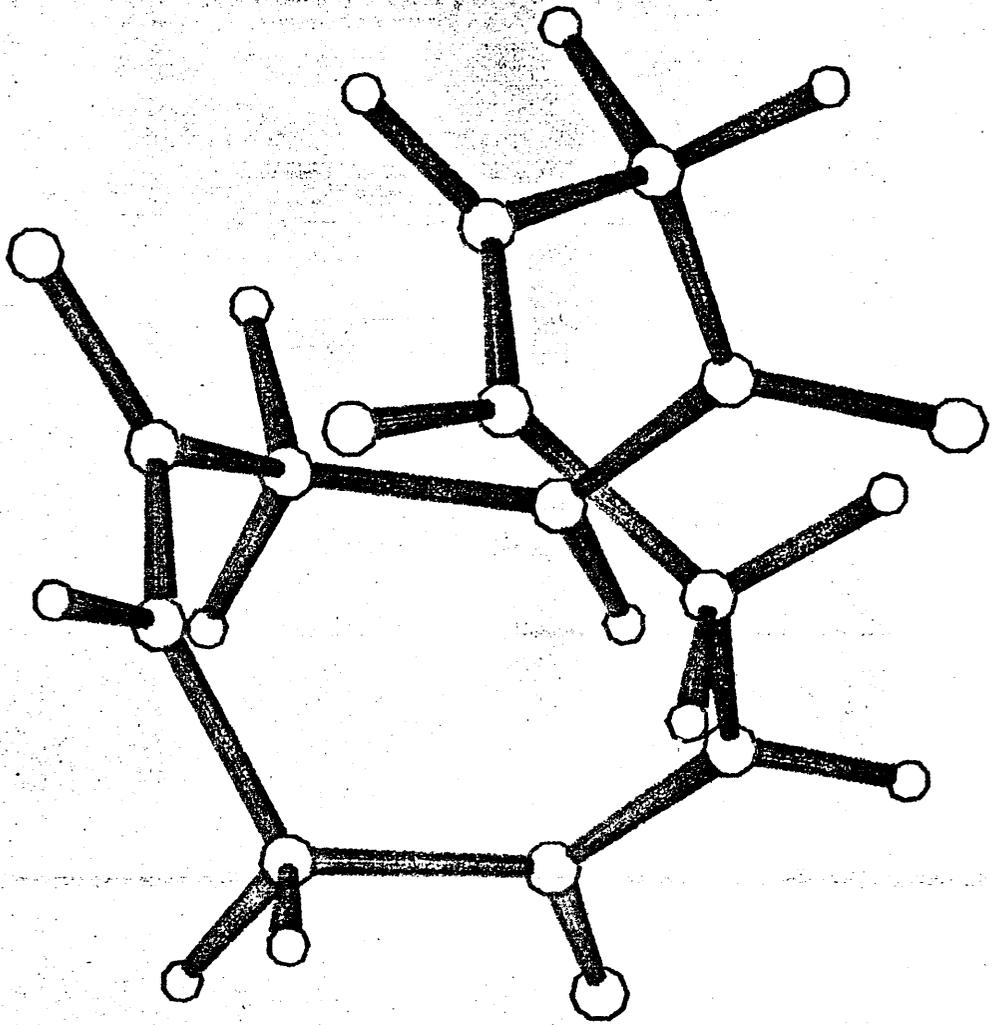
n	ϕ_n	ψ_n	ω_n
1	−47	−61	0
2	47	61	0
3	−47	−61	0
4	47	61	0

Figure 4.36 GL5−1652 CCCC S4

ATOM	X	Y	Z
C'1	1.69716	-0.86506	1.68436
O1	1.04917	-0.65783	2.72039
N1	1.22522	-0.59946	0.47725
H1	1.80126	-0.80638	-0.31583
CA1	-0.06214	0.06417	0.19914
H11	-0.50100	-0.40053	-0.68991
H12	-0.77620	-0.09044	1.01168
C'2	0.15218	1.53870	-0.08224
O2	0.16856	1.89890	-1.26952
N2	0.43838	2.39057	0.89810
H2	0.64211	3.33551	0.63628
CA2	0.39811	2.12768	2.35107
H21	-0.25565	1.29080	2.56206
H22	-0.10367	2.99026	2.80204
C'3	1.67490	1.93073	3.15122
O3	1.55604	2.01256	4.38392
N3	2.86569	1.66034	2.62628
H3	3.62589	1.57763	3.27267
CA3	3.25045	1.57552	1.20664
H31	2.41505	1.35466	0.55344
H32	3.61512	2.56601	0.91645
C'4	4.34711	0.57341	0.91167
O4	5.38304	1.00909	0.38594
N4	4.21645	-0.72332	1.16971
H4	5.02133	-1.28927	0.98095
CA4	3.08893	-1.44454	1.79819
H41	3.32517	-1.56853	2.85990
H42	3.04072	-2.44765	1.36164

Orthogonal Coordinates of: GL5-5251

AMIDE	CHI(C)	CHI(N)	TAU
1	0.09	-4.14	-176.97
2	-6.48	4.73	-4.02
3	1.33	-5.73	0.76
4	-1.45	5.43	1.17

Planarity of the amide groups of: GL5-5251
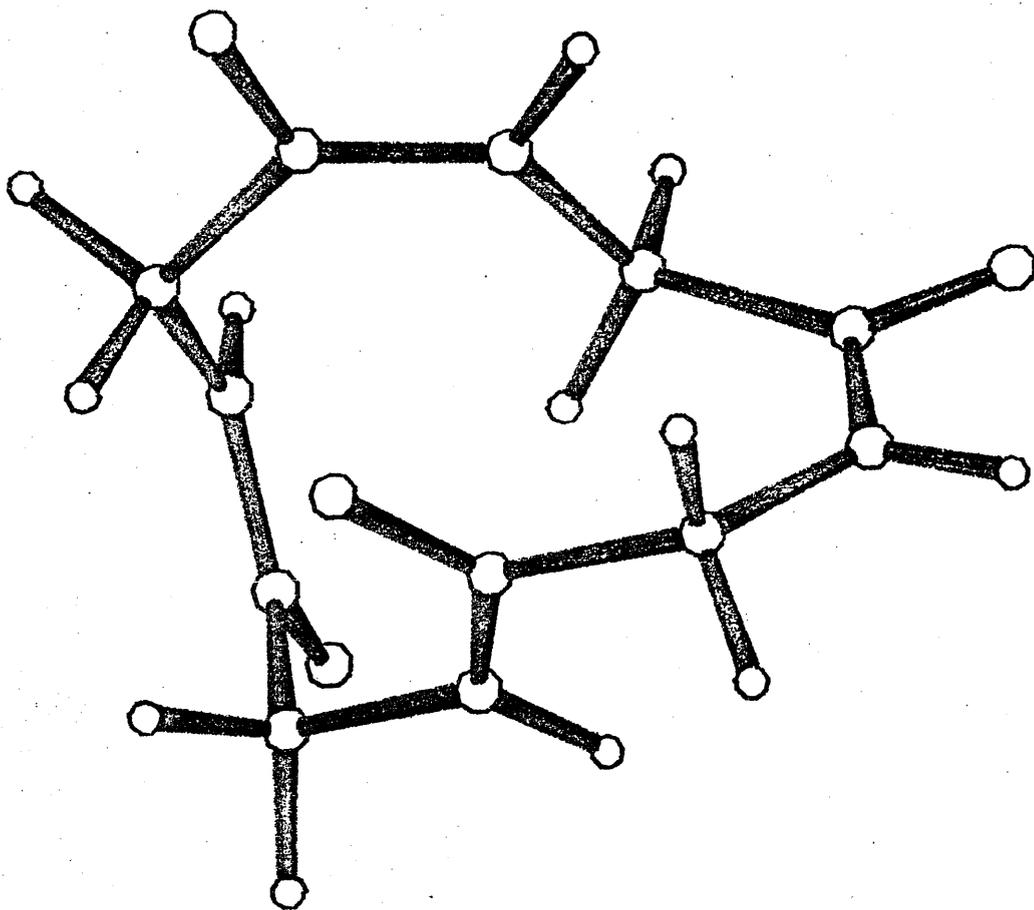
n	Φ_n	Ψ_n	ω_n
1	99	−74	−175
2	105	−17	−10
3	−148	60	4
4	27	55	−2

Figure 4.37 GL5-5251 CCCT

ATOM	X	Y	Z
C'1	2.09243	0.42346	1.36517
O1	1.71081	1.56574	1.66033
N1	1.40538	-0.39054	0.57493
H1	1.76446	-1.30514	0.38974
CA1	0.03604	-0.11893	0.11045
H11	0.01638	0.84952	-0.39930
H12	-0.30323	-0.87780	-0.60138
C'2	-0.86977	-0.07035	1.32196
O2	-1.08955	1.03830	1.83367
N2	-1.35321	-1.18109	1.86597
H2	-1.92051	-1.05463	2.68167
CA2	-1.32441	-2.55519	1.32135
H21	-1.97216	-2.55428	0.43886
H22	-1.80117	-3.22470	2.04487
C'3	-0.01837	-3.21600	0.92842
O3	0.10601	-3.59368	-0.24685
N3	0.93587	-3.45574	1.81826
H3	1.74419	-3.94775	1.49088
CA3	0.97955	-2.99313	3.21792
H31	-0.02680	-2.94871	3.64430
H32	1.57593	-3.67615	3.83144
C'4	1.59879	-1.61584	3.21756
O4	0.89694	-0.64453	3.53430
N4	2.77947	-1.48255	2.63001
H4	3.29800	-2.30013	2.35684
CA4	3.26137	-0.19657	2.09784
H41	4.09216	-0.34754	1.40070
H42	3.60119	0.44838	2.91421

Orthogonal Coordinates of: GL3-1662

AMIDE	CHI(C)	CHI(N)	TAU
1	-10.50	9.92	170.77
2	-3.83	9.55	-4.69
3	-3.84	4.67	-4.20
4	-13.44	13.37	169.38

Planarity of the amide groups of: GL3-1662
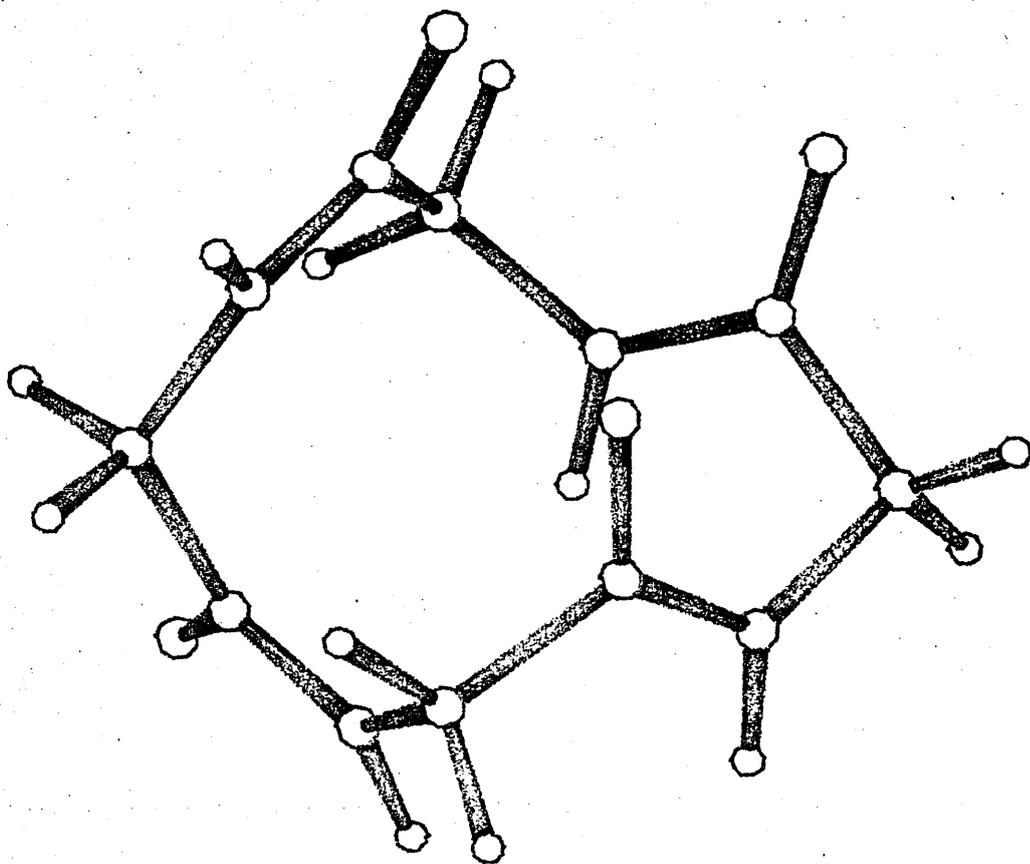
n	ϕ_n	ψ_n	ω_n
1	-63	-83	161
2	56	62	-11
3	-86	-54	-8
4	-44	-45	156

Figure 4.38 GL3-1662 CCTT

ATOM	X	Y	Z
C'1	1.04224	-0.61585	2.25882
O1	1.82897	-1.33237	1.62114
N1	0.02993	0.01153	1.67603
H1	-0.59285	0.56790	2.22807
CA1	-0.19504	0.02622	0.21850
H11	0.51121	-0.65139	-0.27053
H12	-1.20048	-0.35835	0.01935
C'2	-0.04797	1.38129	-0.45165
O2	-0.73246	1.60207	-1.46271
N2	0.80236	2.28926	0.01569
H2	0.84559	3.17270	-0.45363
CA2	1.61561	2.15955	1.23769
H21	2.09483	1.18001	1.24885
H22	0.92604	2.25399	2.07742
C'3	2.73990	3.14630	1.44832
O3	3.08539	3.89049	0.50999
N3	3.33606	3.17036	2.63433
H3	4.13263	3.76928	2.73459
CA3	2.96445	2.37685	3.82725
H31	3.36186	2.89909	4.70395
H32	1.88460	2.37475	3.96714
C'4	3.56664	0.98477	3.85832
O4	4.80243	0.93058	3.95492
N4	2.84564	-0.13418	3.82894
H4	3.36613	-0.99006	3.80999
CA4	1.38178	-0.30257	3.69999
H41	0.83655	0.57400	4.05368
H42	1.06001	-1.14489	4.32060

Orthogonal Coordinates of: GL5-3532

AMIDE	CHI(C)	CHI(N)	TAU
1	-9.07	4.63	172.07
2	-1.22	5.43	-0.90
3	-1.31	-1.47	-4.01
4	2.18	-6.02	-0.93

Planarity of the amide groups of: GL5-3532
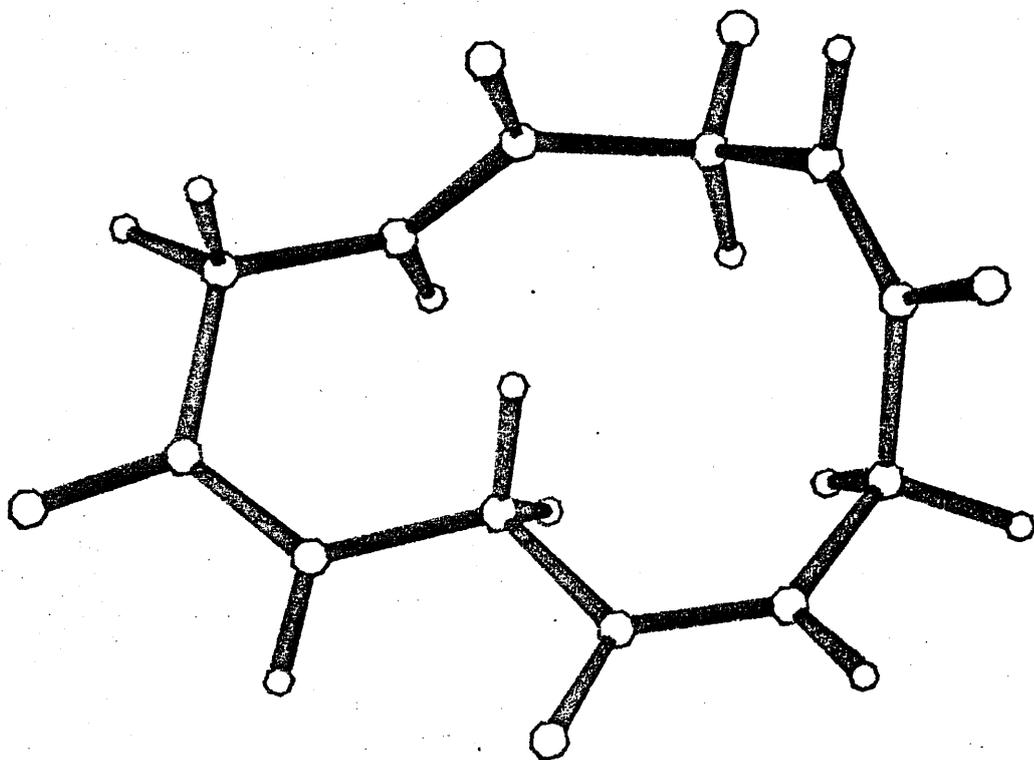
n	ϕ_n	ψ_n	ω_n
1	−114	31	165
2	167	170	−4
3	85	−117	−4
4	99	−135	3

Figure 4.39 GL5-3532 CCCT

ATOM	X	Y	Z
C'1	0.95431	0.68917	2.13923
O1	1.01648	0.71377	3.37849
N1	-0.04819	0.07927	1.52021
H1	-0.69323	-0.42043	2.10199
CA1	-0.25361	-0.09167	0.06874
H11	0.57274	-0.68873	-0.31454
H12	-1.17064	-0.65817	-0.09157
C'2	-0.37325	1.15618	-0.79418
O2	0.30271	1.18937	-1.83447
N2	-1.16854	2.16643	-0.46696
H2	-1.25042	2.90503	-1.13928
CA2	-2.06551	2.27311	0.69878
H21	-2.82328	1.48898	0.60290
H22	-2.59309	3.23116	0.64594
C'3	-1.47093	2.19122	2.08838
O3	-2.02871	1.43589	2.89990
N3	-0.40344	2.89698	2.43841
H3	-0.13177	2.84614	3.40167
CA3	0.35018	3.86625	1.62144
H31	-0.32666	4.69376	1.38608
H32	1.15846	4.28237	2.23177
C'4	0.98971	3.39501	0.33297
O4	0.80934	4.09257	-0.67752
N4	1.72001	2.26891	0.27474
H4	2.17525	2.10085	-0.59798
CA4	2.06879	1.37390	1.37745
H41	2.67705	1.93760	2.09197
H42	2.70513	0.57622	0.98026

Orthogonal Coordinates of: GL6-1616

AMIDE	CHI(C)	CHI(N)	TAU
1	-0.35	-7.37	-1.74
2	0.35	7.37	1.74
3	-0.36	-7.37	-1.74
4	0.35	7.37	1.74

Planarity of the amide groups of: GL6-1616
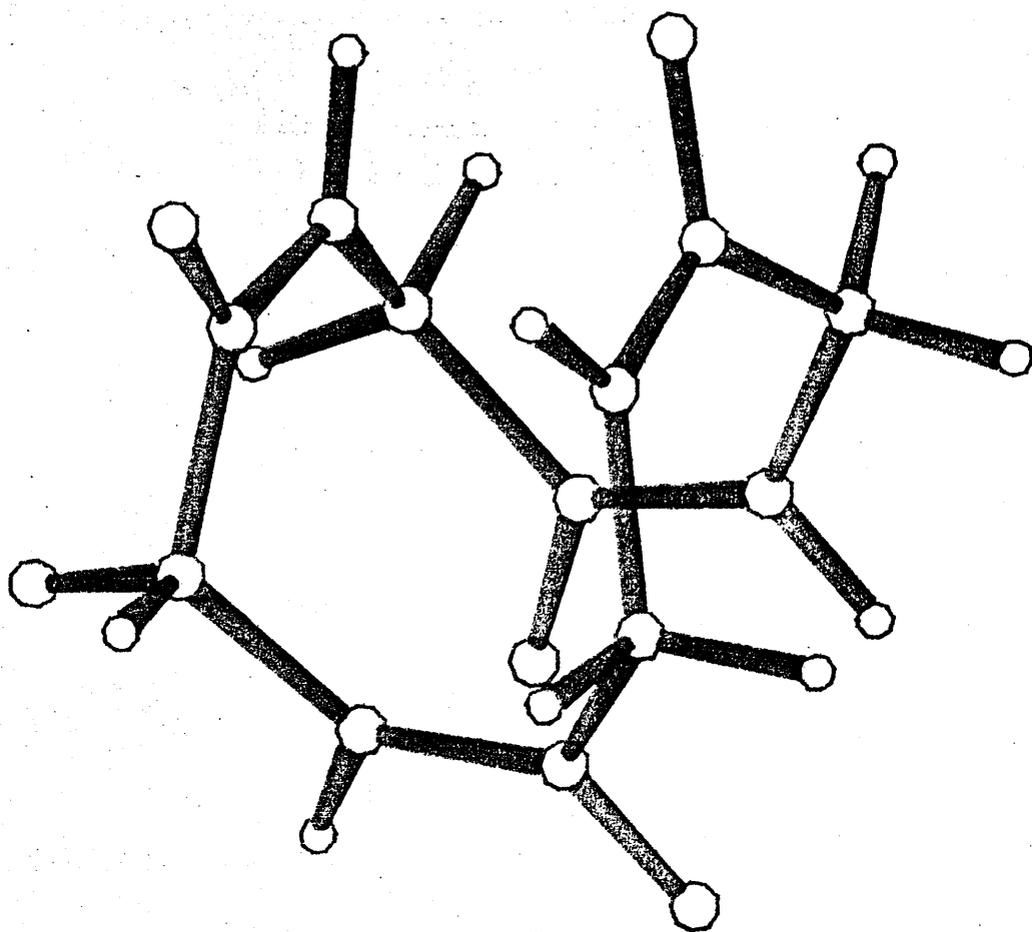
n	ϕ_n	ψ_n	ω_n
1	-60	-58	2
2	60	48	-2
3	-60	-48	2
4	60	48	-2

Figure 4.40 GL6-1616 CCCC   S4

ATOM	X	Y	Z
C'1	-0.61625	0.56099	2.73673
O1	-0.13405	0.85541	3.84123
N1	0.14568	0.54862	1.64768
H1	1.10549	0.80459	1.77813
CA1	-0.22570	0.24639	0.24899
H11	0.52688	-0.45066	-0.13391
H12	-1.17322	-0.28105	0.15344
C'2	-0.14017	1.48892	-0.61757
O2	0.99995	1.92916	-0.83189
N2	-1.20484	2.12194	-1.10132
H2	-1.02461	2.92743	-1.66847
CA2	-2.63264	1.76577	-0.97485
H21	-2.81577	0.93481	-0.29676
H22	-2.97417	1.43934	-1.96241
C'3	-3.50700	2.92200	-0.53588
O3	-4.35380	3.35307	-1.33365
N3	-3.36709	3.44220	0.67622
H3	-3.97889	4.19341	0.93068
CA3	-2.39236	3.03722	1.70625
H31	-1.70150	2.29493	1.31437
H32	-1.78933	3.91308	1.96541
C'4	-3.05805	2.54666	2.97562
O4	-3.76126	3.35194	3.60510
N4	-2.89285	1.29626	3.38766
H4	-3.32379	1.03530	4.25324
CA4	-2.09923	0.25164	2.71290
H41	-2.50440	0.11553	1.71263
H42	-2.23960	-0.70079	3.23442

Orthogonal Coordinates of: GL6-5235

AMIDE	CHI(C)	CHI(N)	TAU
1	-1.21	-1.94	-0.76
2	-2.21	3.32	-0.03
3	0.57	-0.32	1.08
4	0.29	-1.88	-1.72

Planarity of the amide groups of: GL6-5235
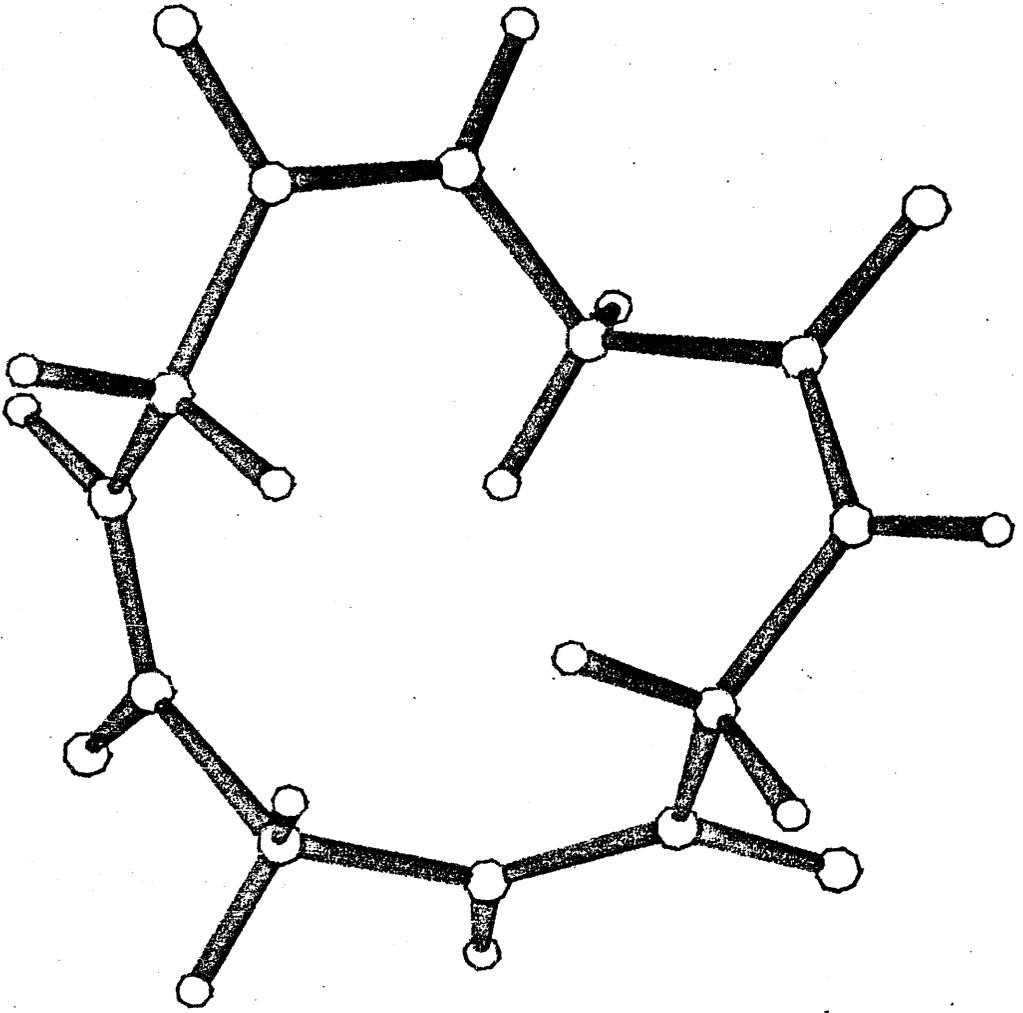
n	ϕ_n	ψ_n	ω_n
1	112	−109	0
2	132	−66	−3
3	−119	117	2
4	67	−128	−1

Figure 4.41 GL6-5235 CCCC

ATOM	X	Y	Z
C'1	-0.40385	1.05089	2.60227
O1	-1.19209	1.14383	3.55563
N1	-0.67929	0.35364	1.50628
H1	-1.59801	-0.03644	1.42458
CA1	0.19154	0.23770	0.31653
H11	1.19957	-0.07894	0.60337
H12	-0.20181	-0.53793	-0.34843
C'2	0.17975	1.54956	-0.44124
O2	-0.92730	2.08481	-0.60605
N2	1.29223	2.16273	-0.82861
H2	1.16202	2.99338	-1.37375
CA2	2.68822	1.68598	-0.72675
H21	3.30637	2.32139	-1.36960
H22	2.72206	0.68368	-1.16573
C'3	3.40717	1.63625	0.61160
O3	4.09448	0.62718	0.83299
N3	3.35720	2.61726	1.50954
H3	3.90743	2.49862	2.33793
CA3	2.55091	3.85401	1.47033
H31	3.23952	4.69306	1.36267
H32	1.88724	3.87862	0.60666
C'4	1.70740	4.08055	2.71357
O4	1.70703	5.21151	3.22357
N4	0.99086	3.08632	3.22653
H4	0.46293	3.26063	4.05943
CA4	0.95367	1.70759	2.71011
H41	1.60678	1.07204	3.31587
H42	1.35575	1.73990	1.70904

Orthogonal Coordinates of: GL6-6541

AMIDE	CHI(C)	CHI(N)	TAU
1	-2.68	8.01	0.47
2	5.19	-9.46	-1.66
3	1.85	-2.41	3.33
4	0.25	-3.25	0.01

Planarity of the amide groups of: GL6-6541
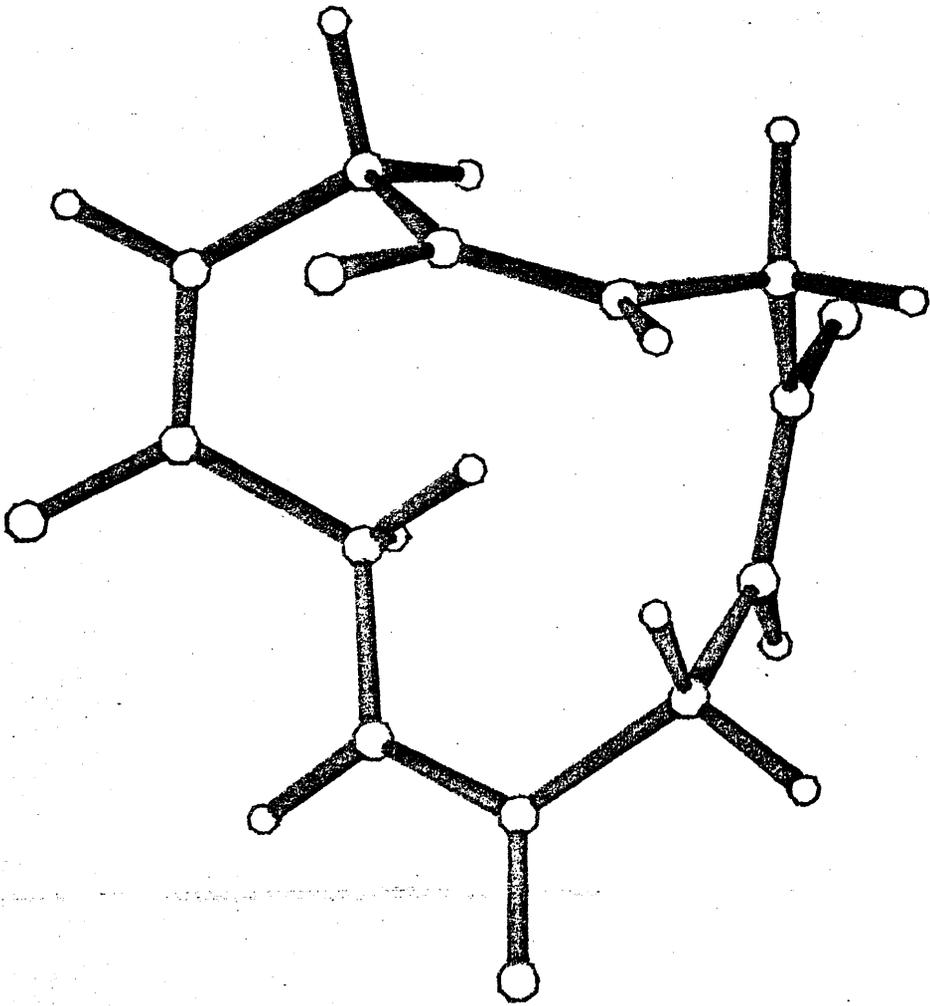
n	ϕ_n	ψ_n	ω_n
1	−72	130	−5
2	−72	−44	6
3	127	−46	5
4	−136	137	2

Figure 4.42 GL6−6541 CCCC

ATOM	X	Y	Z
C'1	1.56311	-0.54571	1.87049
O1	2.10160	-1.39084	1.13933
N1	0.53359	0.18564	1.48036
H1	0.16467	0.86608	2.11593
CA1	0.00544	0.25631	0.10743
H11	0.64941	-0.26106	-0.60799
H12	-0.98143	-0.21595	0.07824
C'2	-0.13439	1.72748	-0.22367
O2	-1.05322	2.33248	0.35100
N2	0.73469	2.39090	-0.97798
H2	0.52956	3.36205	-1.11190
CA2	1.82466	1.86960	-1.82975
H21	2.40987	2.72077	-2.19514
H22	1.34227	1.43338	-2.71041
C'3	2.83705	0.86217	-1.32410
O3	2.93330	-0.20603	-1.94825
N3	3.68198	1.14627	-0.34147
H3	4.35037	0.43365	-0.12349
CA3	3.68360	2.33553	0.53479
H31	4.70857	2.62176	0.79157
H32	3.20931	3.18674	0.03781
C'4	2.90989	1.97655	1.78244
O4	1.80171	2.50348	1.96073
N4	3.28860	0.87204	2.41002
H4	4.15395	0.44531	2.14430
CA4	2.32906	-0.02807	3.06878
H41	1.69592	0.50504	3.78518
H42	2.84326	-0.84336	3.58824

Orthogonal Coordinates of: GL3-6152

AMIDE	CHI(C)	CHI(N)	TAU
1	-18.33	10.82	167.01
2	6.63	-11.50	3.03
3	6.73	-5.37	3.23
4	18.55	-24.23	-165.28

Planarity of the amide groups of: GL3-6152
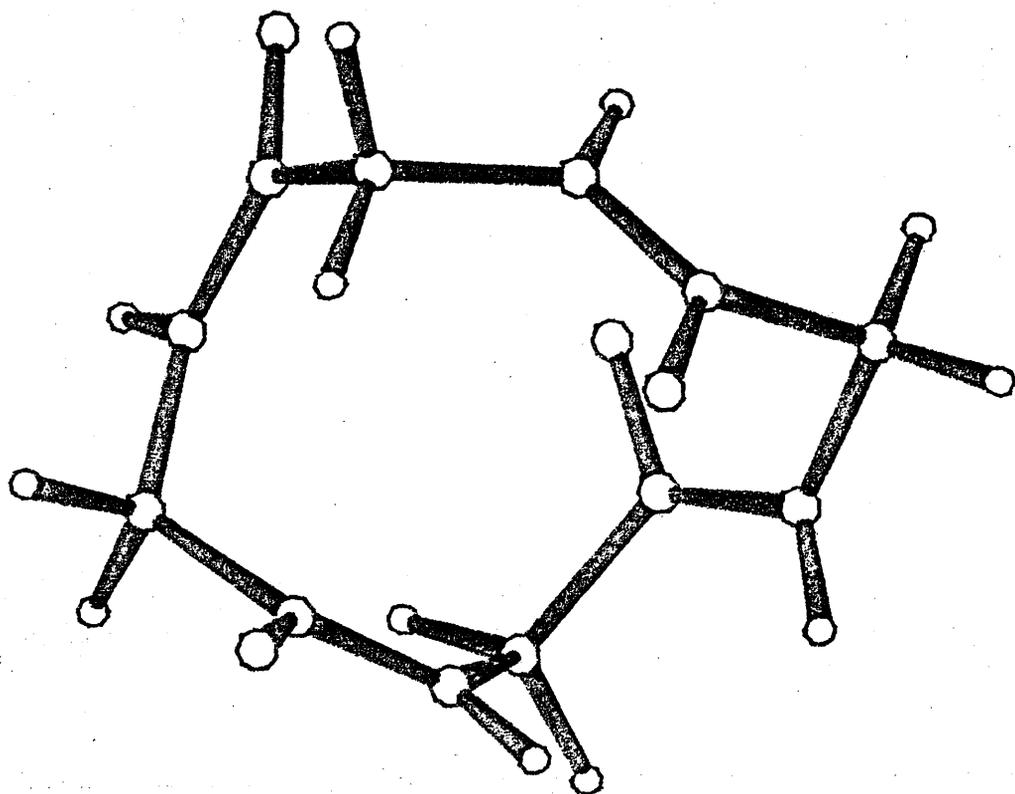
n	ϕ_n	ψ_n	ω_n
1	-131	102	152
2	-50	-64	12
3	94	52	9
4	66	-87	-144

Figure 4.43 GL3-6152 CCTT

ATOM	X	Y	Z
C'1	0.59766	1.73706	1.42658
O1	1.80294	1.77317	1.13811
N1	-0.23661	0.82965	0.93653
H1	-1.20431	0.87387	1.18375
CA1	0.11950	-0.11445	-0.13383
H11	0.97865	-0.70706	0.19631
H12	-0.69253	-0.81290	-0.35115
C'2	0.50540	0.67269	-1.36860
O2	1.71569	0.89292	-1.52983
N2	-0.37778	1.19092	-2.21799
H2	0.02509	1.67781	-2.99545
CA2	-1.84520	1.01741	-2.31796
H21	-1.99022	0.24449	-3.07981
H22	-2.27603	1.92807	-2.74406
C'3	-2.71040	0.60149	-1.14548
O3	-3.00708	-0.60034	-1.06102
N3	-3.24902	1.48243	-0.30947
H3	-3.84289	1.10620	0.40487
CA3	-3.02188	2.93977	-0.21996
H31	-2.44827	3.32746	-1.06519
H32	-3.98398	3.46127	-0.20098
C'4	-2.29976	3.16379	1.08967
O4	-2.84923	2.76719	2.12861
N4	-0.99769	3.38112	1.02498
H4	-0.58386	3.57102	0.13248
CA4	-0.05253	2.94304	2.06501
H41	-0.55507	2.66466	2.99663
H42	0.68091	3.72736	2.27725

Orthogonal Coordinates of: GL3-1255

AMIDE	CHI(C)	CHI(N)	TAU
1	-15.59	10.77	168.93
2	-4.23	9.63	-2.07
3	-6.80	4.87	-2.06
4	20.95	-20.73	-167.97

Planarity of the amide groups of: GL3-1255
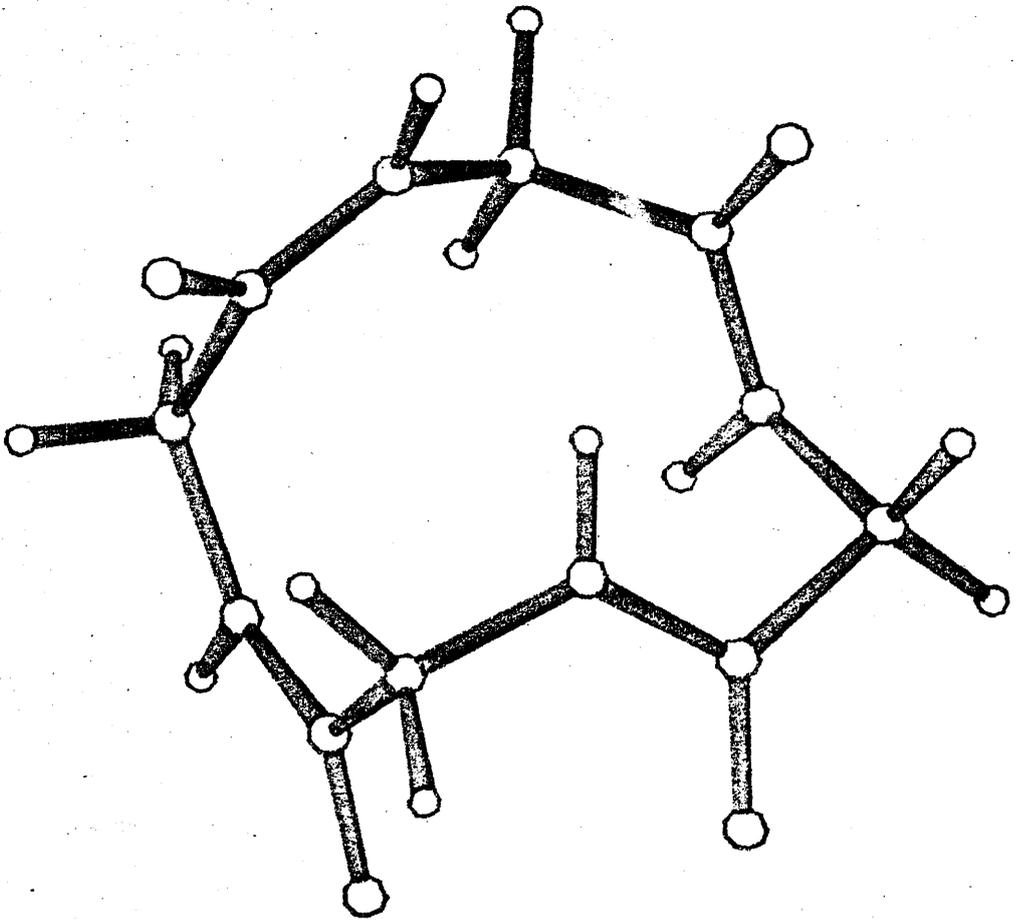
n	ϕ_n	ψ_n	ω_n
1	-60	-81	156
2	24	89	-9
3	-112	103	-8
4	104	-51	-147

Figure 4.44 GL3-1255 CCTT

ATOM	X	Y	Z
C'1	1.23036	-0.35522	2.08503
O1	1.40426	-0.21639	3.30581
N1	0.18330	0.21804	1.49776
H1	-0.43128	0.74416	2.08839
CA1	-0.25460	0.09247	0.09526
H11	0.19306	-0.79111	-0.35033
H12	-1.32516	-0.13447	0.13769
C'2	-0.11494	1.26056	-0.86862
O2	-1.07025	1.45509	-1.63631
N2	0.99402	1.98398	-0.99793
H2	0.99983	2.67188	-1.72601
CA2	2.18899	1.97941	-0.13389
H21	1.94235	1.54315	0.82971
H22	2.39946	3.03117	0.08727
C'3	3.50039	1.38751	-0.62670
O3	4.52877	2.02160	-0.34357
N3	3.59843	0.20379	-1.22590
H3	4.52359	-0.12254	-1.42825
CA3	2.50980	-0.65232	-1.73236
H31	1.60687	-0.06150	-1.85615
H32	2.79886	-0.92801	-2.75209
C'4	2.15884	-1.95623	-1.03262
O4	1.91150	-2.92411	-1.76872
N4	1.99875	-2.06943	0.28320
H4	1.69236	-2.95700	0.62300
CA4	2.33043	-1.08315	1.32809
H41	3.03245	-0.35418	0.93382
H42	2.90140	-1.63236	2.08424

Orthogonal Coordinates of: GL6-5254

AMIDE	CHI(C)	CHI(N)	TAU
1	5.99	-5.32	5.38
2	-6.04	5.34	-5.37
3	6.04	-5.34	5.37
4	-6.04	5.34	-5.37

Planarity of the amide groups of: GL6-5254
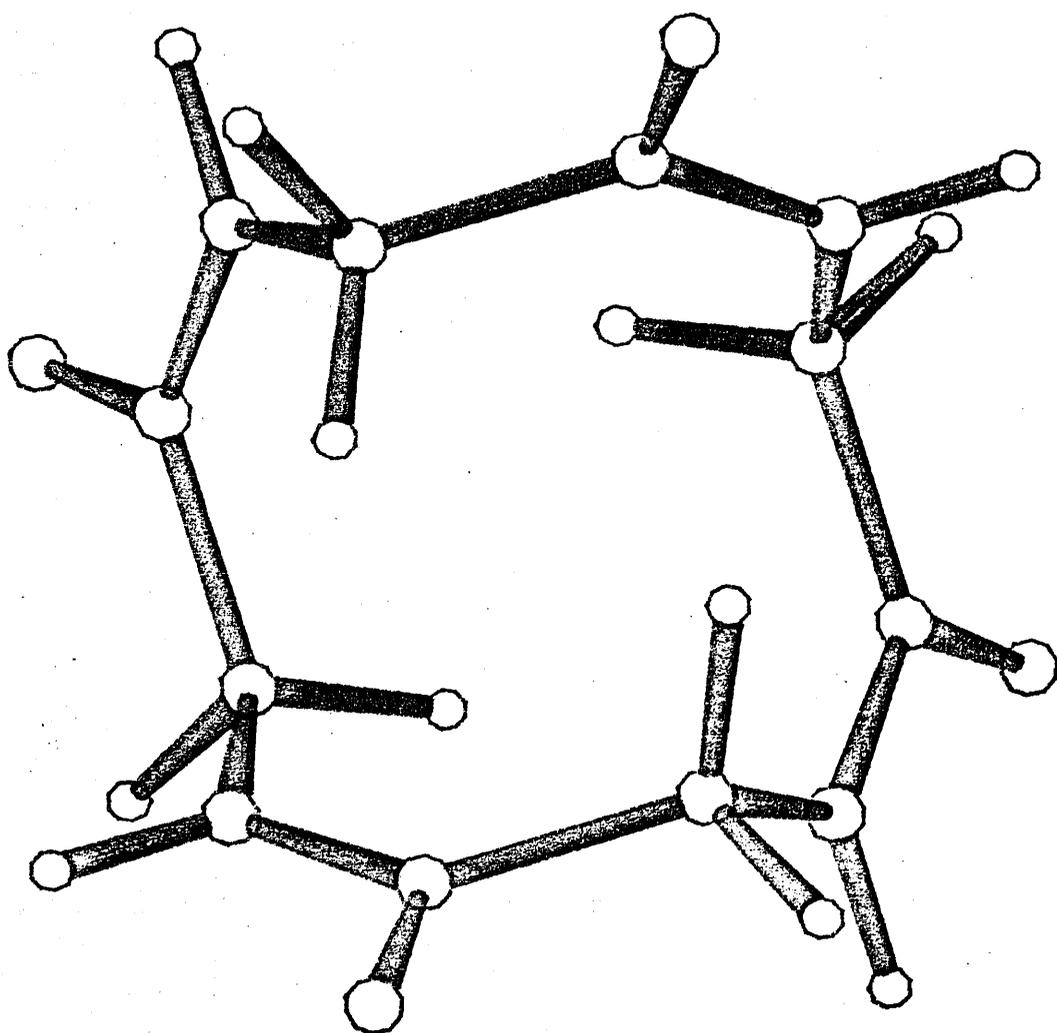
n	ϕ_n	ψ_n	ω_n
1	-109	47	11
2	109	-47	-11
3	-109	47	11
4	109	-47	-11

Figure 4.45 GL6-5254 CCCC S4

4.10 Discussion

The IR spectrum of sublimed cyclotetraglycyl suggests that all four amide bonds are identical but neither cis nor trans [11], and Scheraga's proof [12] that these amide groups cannot then be planar tends to support the calculated Global Minimum Energy Conformation. The calculated transoid amide bonds of the S4 conformation fit in very well with the experimental results since it is very unlikely that the all cis conformation could be a serious contender. In addition, a ring conformation very similar to the calculated S4 conformation has been observed in an X-Ray structure of dihydrochlamydocin [13]. This is in direct contradiction of a statement by Titlestad [11] that "..no crystal structures have yet been determined of cyclic tetrapeptides which do not adopt the cyclotetrasarcosyl [i.e. CTCT Ci] conformation".

The situation in solution is less straightforward and two conflicting studies have been undertaken, mainly using n.m.r. spectroscopy [11,14]. The Scandinavian group [11] studied cyclotetraglycyl in trifluoroacetic acid (TFA) and in water. They deduce that the solution conformation is not the same as the sublimed. Furthermore solid material crystallised from water adopted the same conformation as in solution. From the n.m.r. spectrum they deduce that this second conformation contains alternating cis and trans amides. This solution structure has been assigned to a cyclotetrasarcosyl-type (i.e. CTCT Ci) conformation. This conclusion was apparently reached as much on the basis of that conformation's occurrence in crystal structures as from the direct experimental evidence.

The Swiss group, under Grathwol, recorded the n.m.r. spectrum of cyclotetraglycyl in both TFA and dimethylsulphoxide (DMSO) [14] and interpret the results in terms of TTTT S4 conformation. There seems to be some doubt as to whether the samples used were authentic cyclotetraglycl – it is possible that the synthesis used could have yielded cyclo-octaglycyl [12,15,16]. These experiments are being repeated [15] and in all probability will result in conclusions similar to those found by the

Scandinavians.

The difference in conformation between the sublimed solid and the solution and crystalline states must be attributable to the presence of solvent. There are two possible mechanisms for this effect. Firstly protonation of the peptide groups (at the oxygen atoms) would increase the double bond character of the C'-N bond. This in effect means an increase in the barrier to free rotation around the bond from the 20 kcal/mole or so which is typical of an amide bond to about 60 kcal/mole corresponding to a double bond. This involves a considerable increase in Pitzer strain for the TTTT S4 conformation whereas the CTCT Ci is almost unaffected.

To test this hypothesis the models of the eight minimum energy conformations were "protonated" using the Chemical Graphics System and then re-optimised using a force-field reflecting the increase in the barrier to free rotation around the C'-N bond. The results are summarised in Table IV which shows a dramatic alteration in the relative energies of the conformers. The CTCT Ci conformation is now some 6 kcal/mole lower in energy than the TTTT S4. This indicates that protonation of cyclotetraglycyl should involve a conformational change as the Scandinavian experiments suggest. Further direct evidence comes from Winkler and Dunitz [17] who have shown from X-Ray and spectroscopic studies that protonation in a strained ring (such as cyclotetraglycyl) can induce trans -> cis isomerisations.

An alternative explanation for the behaviour of cyclotetraglycyl can be given in terms of hydrogen bonding rather than protonation. The TTTT S4 conformation is stabilised by weak hydrogen bonds across the "corners" of the ring (where the α-carbon atoms are the corners). This feature was observed by Karle in the crystal structure of dihydrochlamydocin [13]. Formation of strong solvent-solute hydrogen bonds (as could occur in both water and TFA) would remove the stabilisation of the all - trans conformation.

The CTCT Ci conformation has no intra-molecular H-bonds and therefore

CONFORMATION	CONFIGURATION & SYMMETRY		ENERGY
GL4–4532	CTCT	Ci	0.0000
GL4–5126	CTCT	Ci	0.7754
GL4–1531	CTCT		0.9583
GL4–2141	CTCT	C2	1.9409
GL1–6252	TTTT		5.3758
GL1–5252	TTTT	S4	6.2530
GL1–5212	CTTT		6.2644
GL1–5623	TTTT	Ci	7.7418

TABLE IV. Relative Energies (kcal/mole) after Protonation at the Oxygen Atoms.

formation of solvent-solute H-bonds would have very little effect on its stability.

4.11 Conclusion

The global minimisation algorithm has successfully located the GMEC of cyclotetraglycyl as well as a range of conformers close to the minimum in terms of energy.

4.12 References

1. P. de Santis and A. M. Liquori, Biopolymers, 10, 699 (1971).

2. G. N. Ramachandran and V. Sasiskharan, Adv. Prot. Chem., 23, 284 (1968).

3. C. M. Deber, V. Madison and E. R. Blout, Acc. Chem. Res., 9, 106 (1976).

4. Y. A. Ovchinnikov and V. T. Ivanov, Tetrahedron, 31, 2177 (1975).

5. N. Go and H. A. Scheraga, Macromolecules, 6, 525 (1973).

6. P. N. Lewis, F. A. Momany and H. A. Scheraga, Isr. J. Chem., 11, 121 (1973).

7. P. Groth, Acta Chem. Scand., 27, 3217 (1973).

8. see Appendix A in D. N. J. White and C. Morrow, Computers in Chemistry, (1978) in press.

9. D. N. J. White and M. H. P. Guy, J. Chem. Soc. Perk. II, 43 (1975).

10. F. K. Winkler and J. D. Dunitz, J. Mol. Biol., 59, 169 (1971).

11. K. Titlestad, Acta. Chem. Scand., B31, 641 (1977).

12. N. Go and H. A. Scheraga, Macromolecules, 3, 178 (1970).

13. J. L. Flippen and I. L. Karle, Biopolymers, 15, 1081 (1976).

14. R. Schwyzer, B. Iselin, W. Rittel and P. Sieber, Helv. Chim. Acta, 39, 872 (1956);

C. Grathwol, A. Tun-Kyi, A. Bundi, R. Schwyzer and K. Wuttrich, Helv. Chim. Acta, 58, 415 (1975).

15. K. Wuttrich, personal communication (1978).

16. J. D. Dale and K. Titlestad, Tetrahedron Letters, 379 (1978).

17. F. K. Winkler and J. D. Dunitz, Acta Cryst., B31, 278 (1975).

CHAPTER FIVE

Correlation of Observed and Calculated Conformations

5.1 Introduction

In attempting to correlate observed cyclotetrapeptide, and other, conformations with the calculated set of conformations various effects must be considered. The foremost of these must be the effect of substituents both at the nitrogen and α-carbon, this is discussed in detail in Sections 5.2 & 5.3.

The general effect of substitution will be to change the unsubstituted conformation in some way. An obvious example of this is a tetrapeptide composed of four different amino acid residues adopting a conformation which, for cyclotetraglycyl, would be symmetric. This would result in an observed conformation which was not symmetric with, however, a basic similarity between the observed and calculated ring torsion angles.

The degree of distortion will depend on the particular substituents and the pattern of substitution. This means that correlations must be looked for in terms of similarities between the ring conformations rather than exact correspondances of the torsion angle values.

5.2 N-Methylation

N-methylation removes the normal preference of peptides for trans amides i.e. cis and trans amides are nearly isoenergetic when N-methylated [28]. In an attempt to assess the importance of this effect the models of the eight lowest energy conformations were "N-methylated" in various ways. The Chemical Graphics System was used to produce models of the mono-, di- and tetra- N-methylated derivatives of these conformations. Di-N-methylation at adjacent groups and tri-N-methylation were not included since these patterns are not commonly observed — this is a consequence of the coupling reactions used to make cyclic tetrapeptides [1,7].

The results of these operations are summarised by Figures 5.1, 5.2 & 5.3 which show the ring conformation and energy (in kcal/mole) above the TTTT 94 conformation. The differences in the relative energies induced by

TTTT S4

CTCT
Ci

CTCT

Ci

0.00

0.04

−2.71

0.70

−1.77

1.12

−1.92
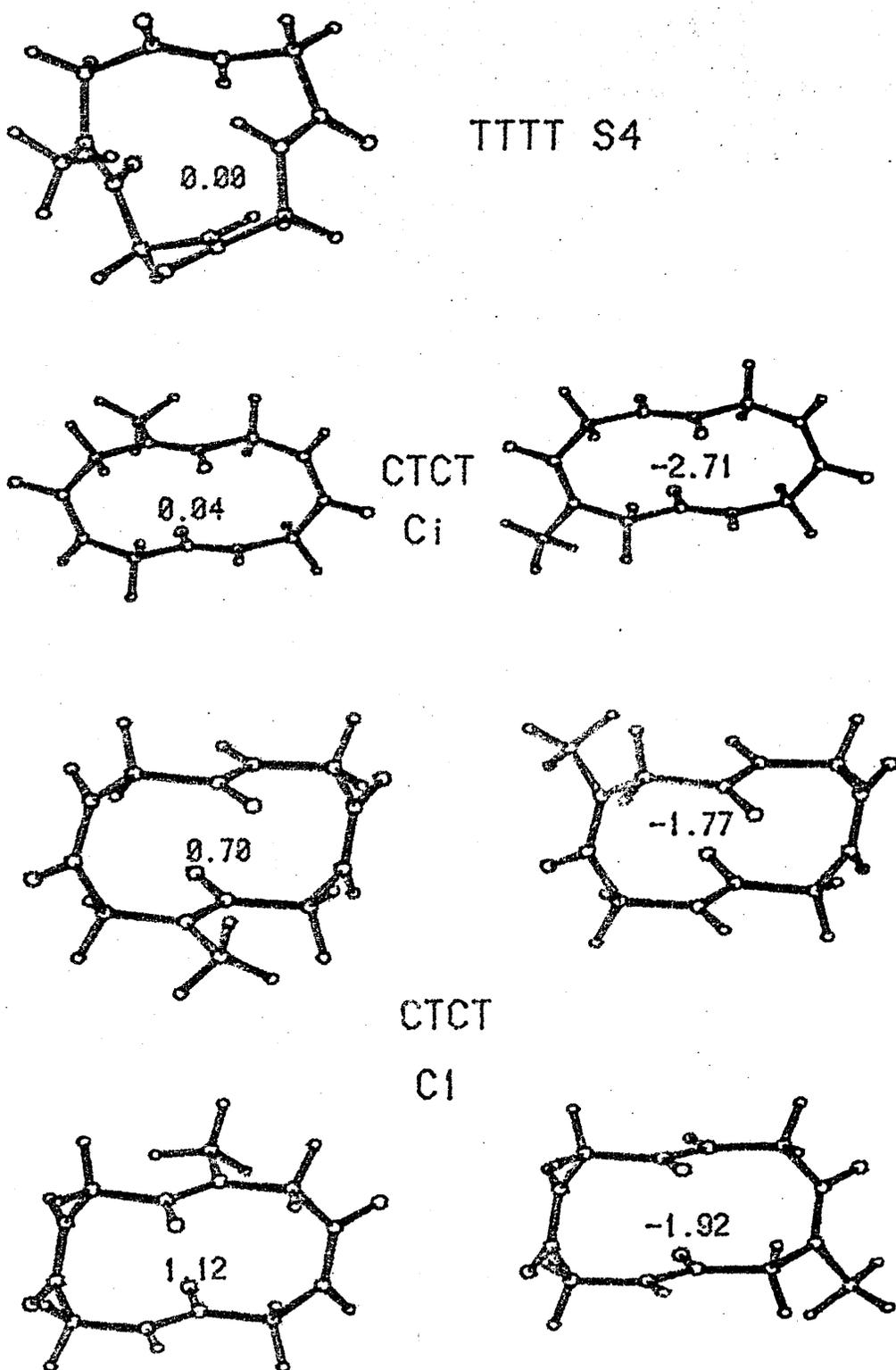
Figure 5.1 Mono-N-Methylated
derivatives with energies(in
kcal/mole) relative to the
TTTT S4 conformation.

TTTT
Ci

CTCT
Ci

TTTT
C1

Figure 5.1 Continued.

CTTT
C1

CTCT
C2

Figure 5.1 Continued

TTTT S4

0.00

−1.14 CTCT Ci

−6.70

−0.57 CTCT Ci

−2.74

3.54 TTTT Ci

1.09
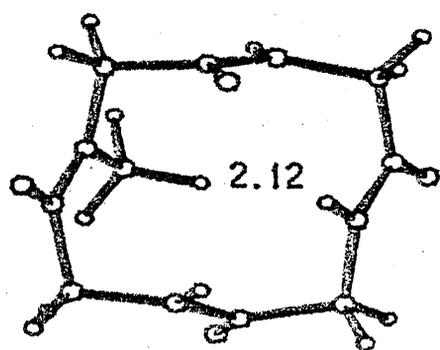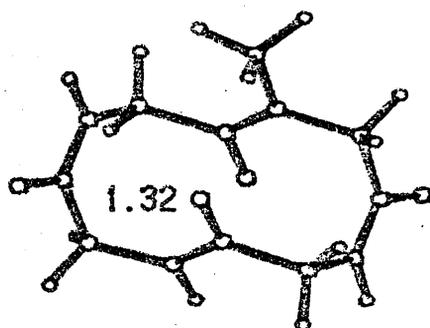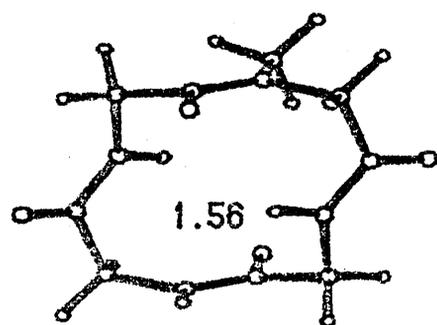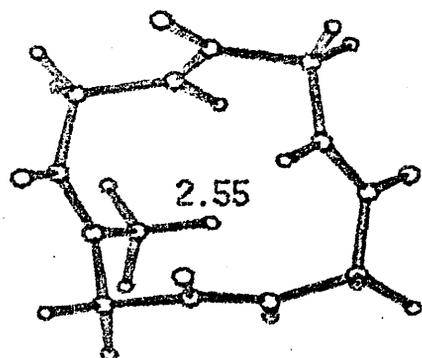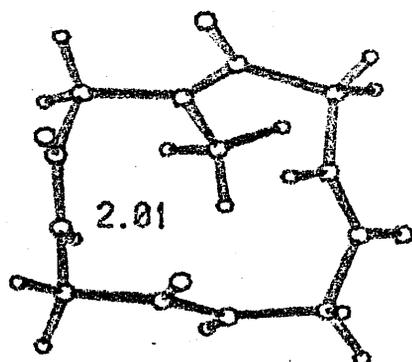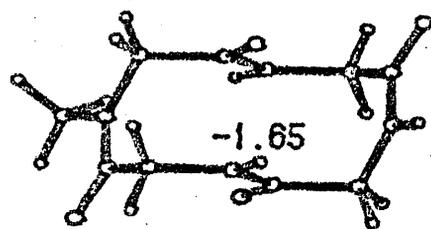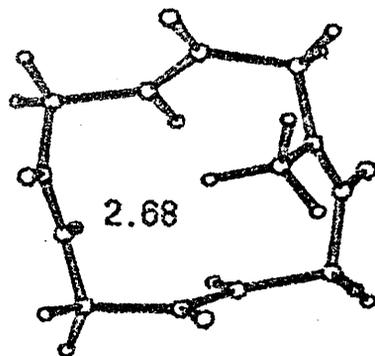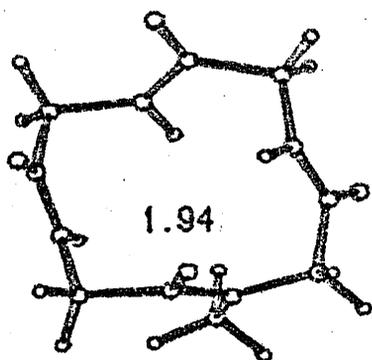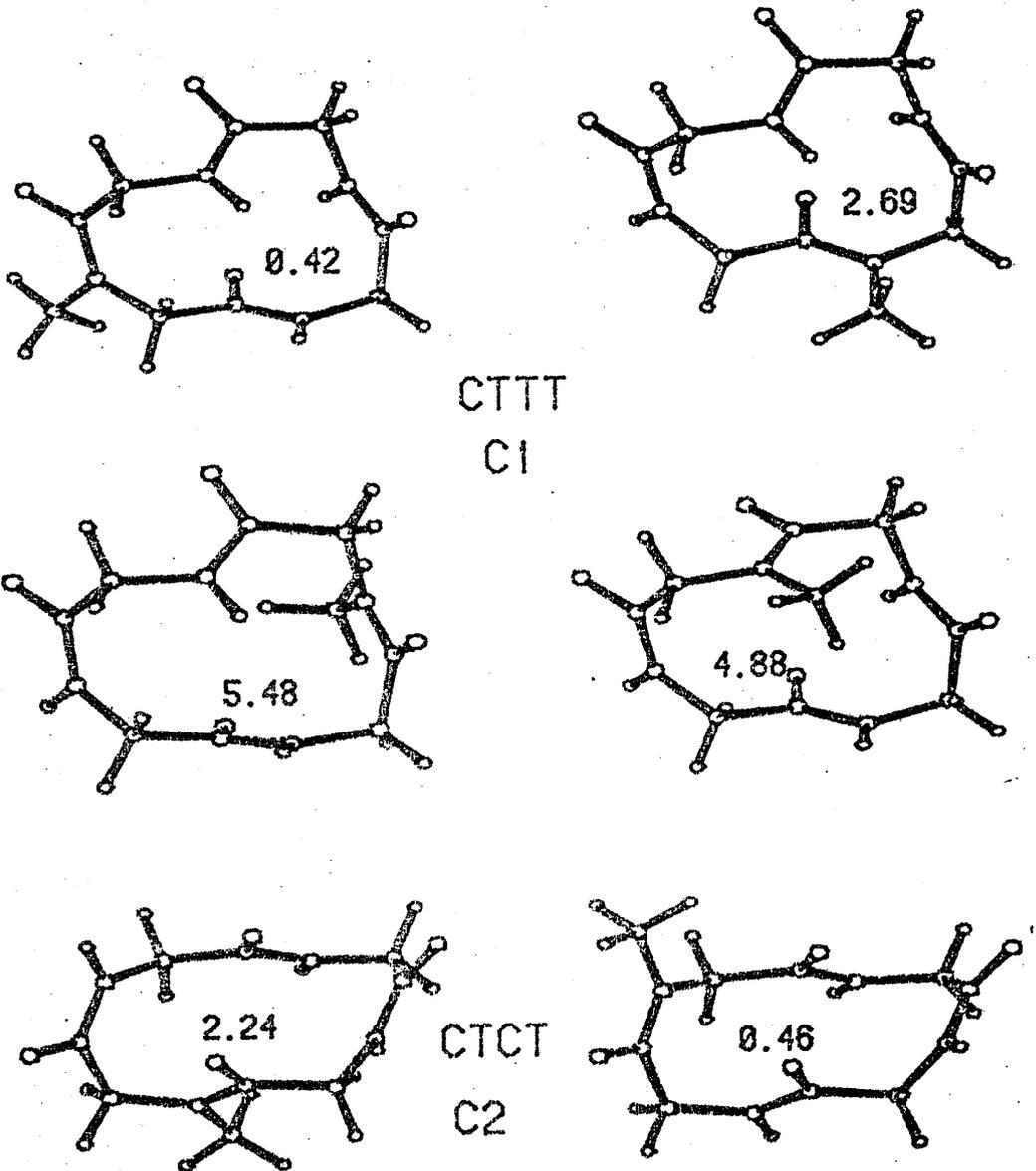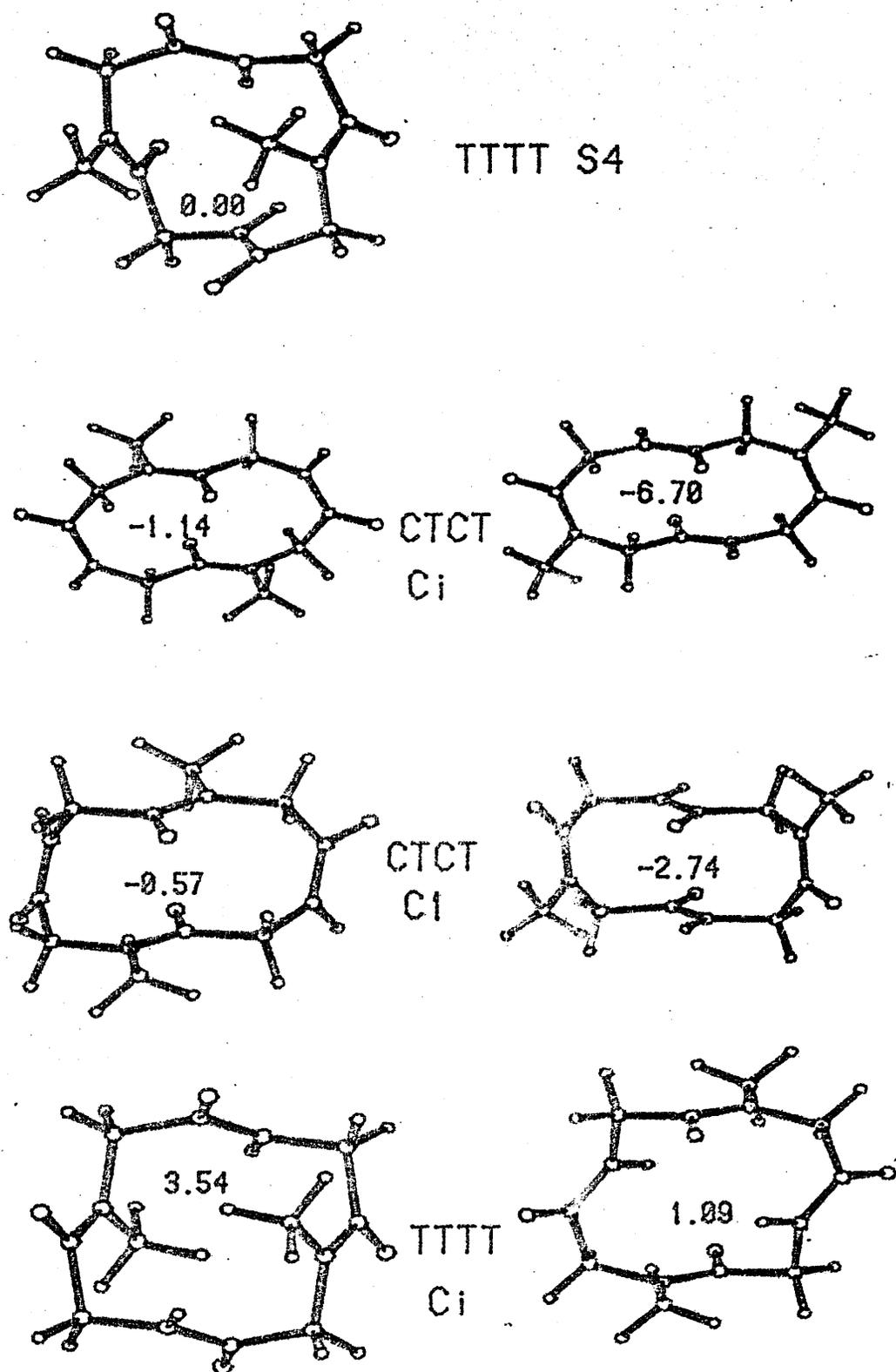
FIGURE 5.2 Di-N-Methylated
derivatives with energies (in
kcal/mole) relative to the
TTTT S4 conformation.

CTCT
Ci

TTTT
C1

CTTT
C1

CTCT
C2

0.09 -2.86

1.62 2.35

2.19 3.81

0.50 -3.13
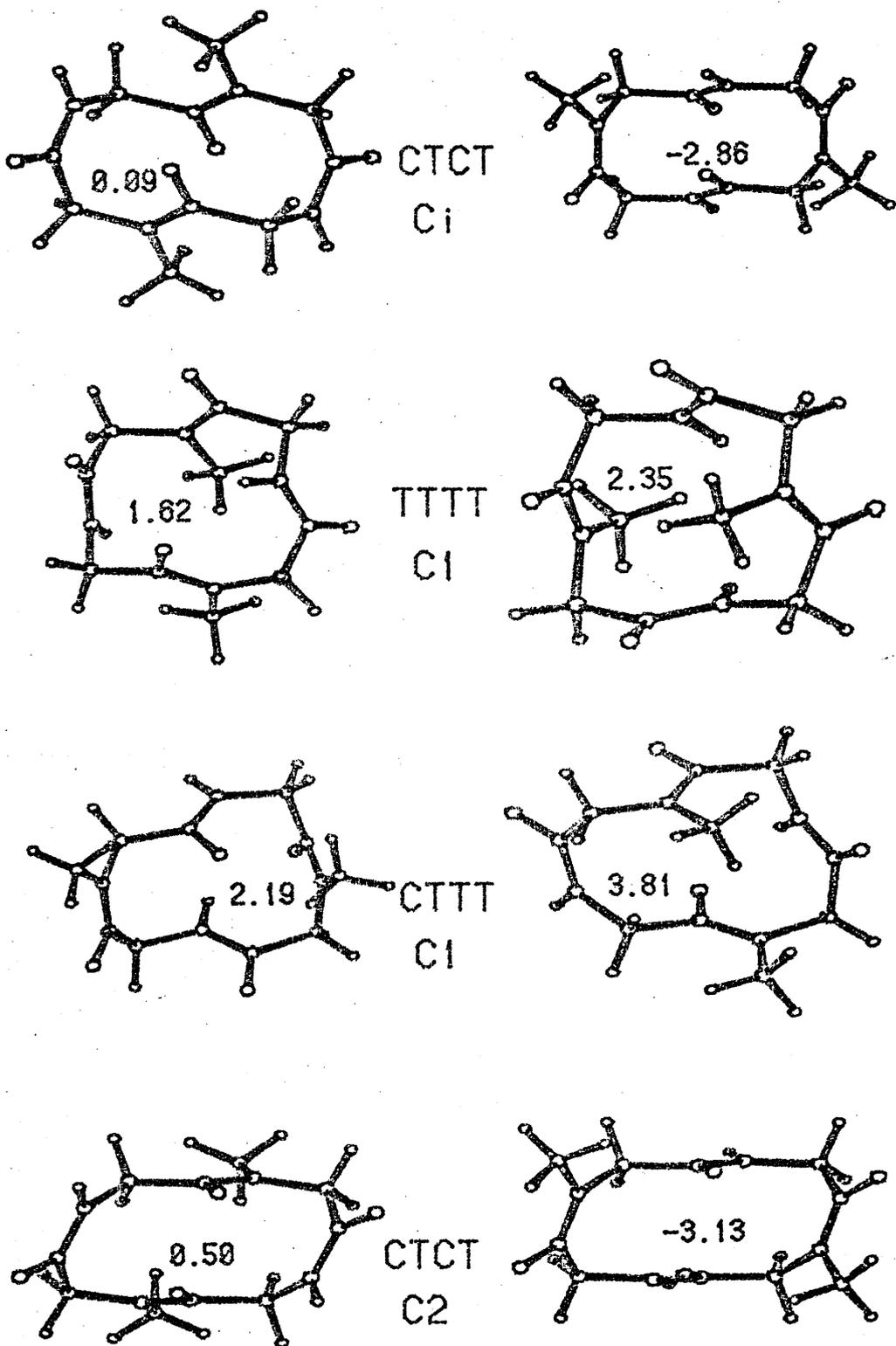
Figure 5.2 Continued
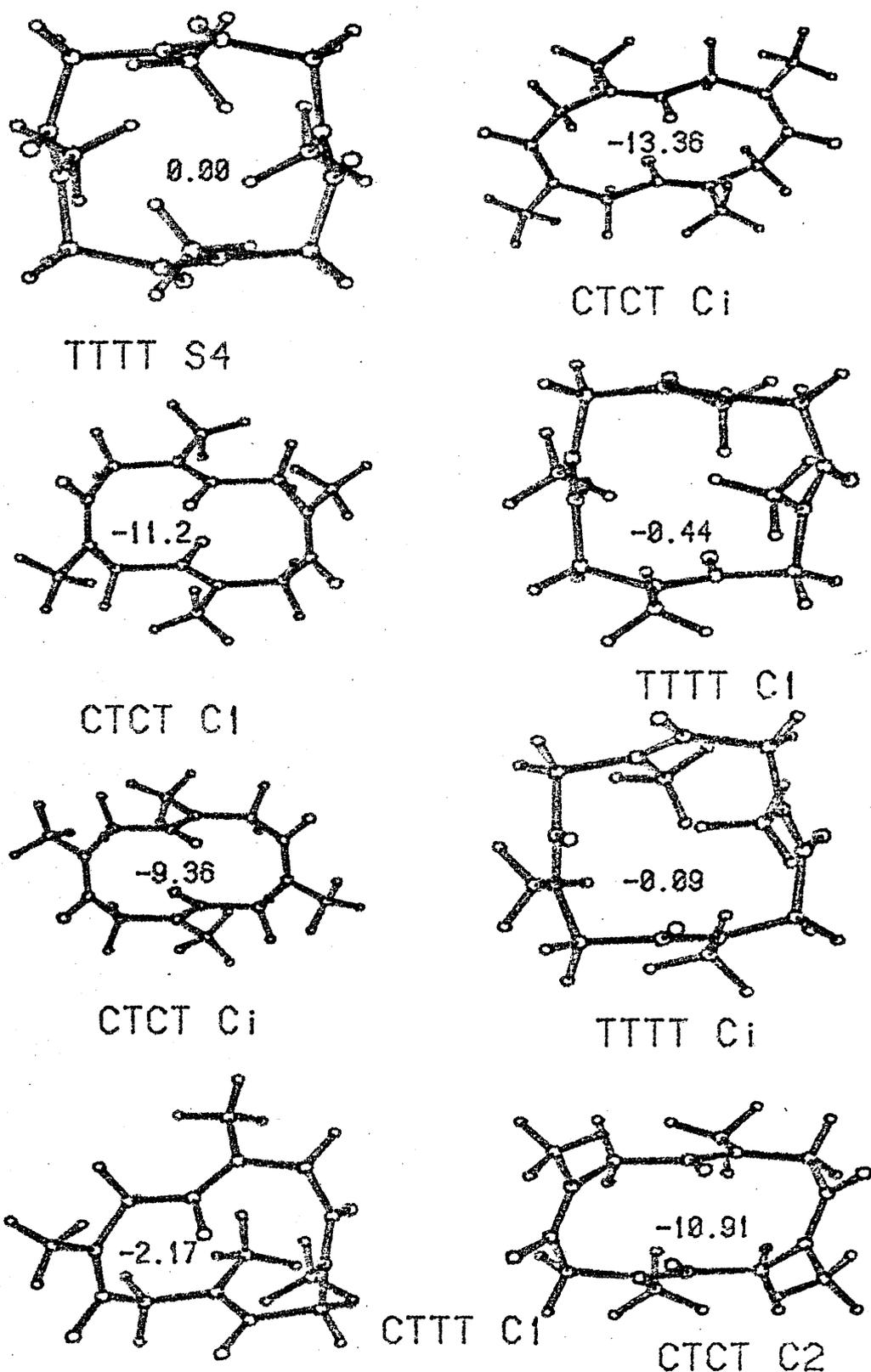
Figure 5.3 Tetra-N-Methylated derivatives with energies(in kcal/mole) relative to the TTTT S4 conformation.

the N-methylation are very marked especially in the case of the tetra-N-methylated derivatives where the TTTT S4 is now some 13 kcal/mole higher than the CTCT Ci. This pattern is repeated throughout the series of derivatives so that the CTCT Ci conformation (when N-methylated at one or both cis positions) is the minimum energy conformation for each derivative.

One particularly interesting result is that, for the di-N-methylated derivative the CTCT Ci is slightly lower in energy than the TTTT S4 even when the former is methylated at trans positions.

The minimisations of the various N-methyl derivatives were carried out on the pure diagonal (first stage) program only. The r.m.s. first derivatives of energy w.r.t. coordinates were all less than 0.1 kcal/mole/Å. The energies should therefore be regarded as reasonable approximations to the fully refined values.

5.3 Substitution at the α-Carbon

The steric effect of an α-carbon substituent is less easy to quantify than N-methylation. Both the actual size and configuration of the side chain are important. The reason for this configuration dependence lies in the fact that there are two distinct types of substituent positions.

The first of these has been termed "inner" since substitution at these positions causes the side chain atoms to partially overlap the ring. On the other hand "outer" positions do not involve this type of steric crowding. The most favourable substituent pattern for any ring conformation is that which minimises the number of inner substituents or ideally avoids them altogether. Thus for the CTCT Ci conformation this is the DDLL configuration which has no inner side chains (see Fig. 5.4a). The mirror image of the conformation must be considered as well as the original but in this instance this gives LLDD which is equivalent to DDLL.

It can be seen that the TTTT S4 conformation has only outer positions (see Fig. 5.4b) thus any configuration can be accommodated.
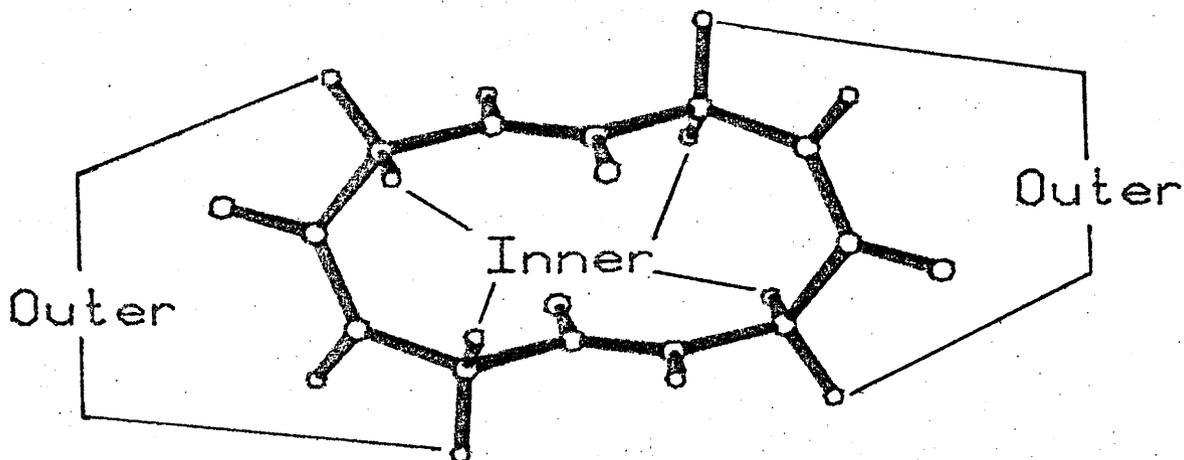
(a) CTCT Ci
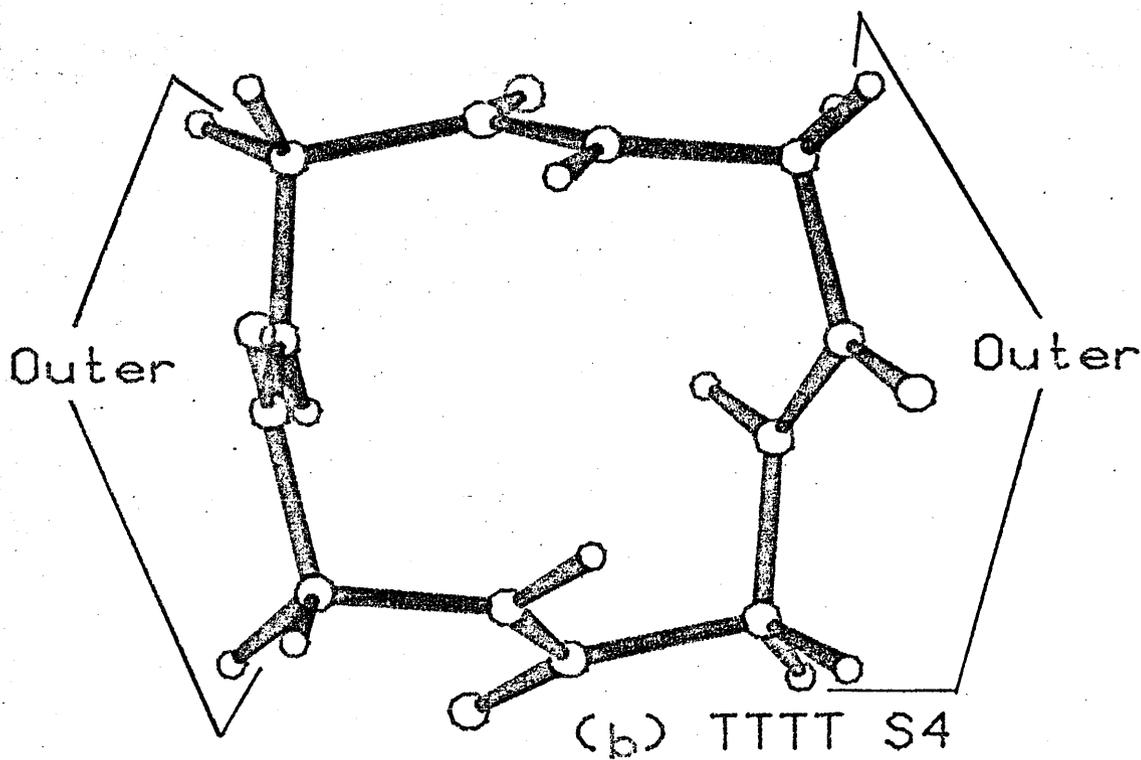
(b) TTTT S4

Figure 5.4 Inner and Outer
Substituent positions.
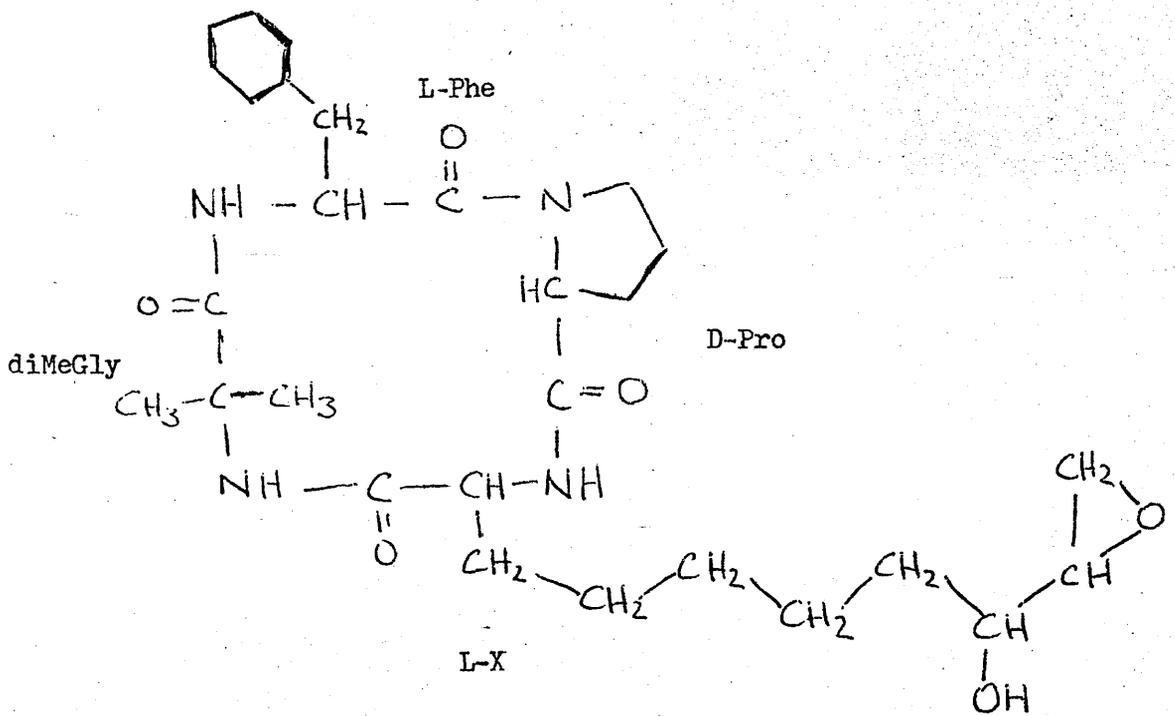
dihydrochlamydocin

This scheme of locating inner and outer substituent positions was extended to the lowest energy eight conformations and the resulting acceptable configurations are given in Table I.

A number of structures of cyclotetra -peptides and -depsipeptides have been experimentally determined. Comparisons between the observed and calculated ring conformations are given in the following sections together with an attempt to find the reasons for a compound's choice of conformation.

5.4 TTTT S4 Conformations

In addition to the sublimed form of cyclotetraglycyl [1] only one other example of the TTTT S4 has been observed. This is dihydrochlamydocin (c-Iabu-L-Phe-D-Pro-L-X) [2]. A comparison of the observed and calculated ring torsion angles is given in Table II.

The most important factor in this peptide is the D-prolyl residue. Extensive manipulation of Dreiding models has indicated that it is impossible to accommodate a D-Pro within a cis amide unit. This removes the possibility that the CTCT conformations could be favoured by having an N-substituted cis amide. The lowest energy all-trans conformation (the S4) is adopted since this is lower in energy than any of the other conformations which can accommodate the required configuration (ILDL) without inner sidechains.

5.5 CTCT Ci Conformations

This is by far the most widely adopted conformation with over half-a-dozen observed examples (both peptides and depsipeptides). The ring torsion angles of the calculated and observed conformations are given in Table III. In addition to the crystal structures there is also protonated cyclotetraglycyl [1] discussed earlier and Tentoxin derivatives [3,4,5,13,14] which will be discussed later.

CONFORMATION	ALLOWED CONFIGURATIONS
1. TTTT S4	ANY
2. CTCT Ci	DDLL
3. CTCT C1	LLLL & DDDD
4. TTTT Ci	ANY
5. CTCT Ci	LLLL & DDDD
6. TTTT C1	ANY
7. CTTT C1	LLLL, DLLL, DDLL, DDDL & DDDD
8. CTCT C2	LLLL, DLLL, DLDL, DDDL & DDDD

TABLE I. Sterically Uncongested
Side-chain Configurations.

	a	b
Φ_1	-104	-105
Φ_2	104	83
Φ_3	-104	-105
Φ_4	104	72
Ψ_1	65	94
Ψ_2	-65	-73
Ψ_3	65	105
Ψ_4	-65	-64
ω_1	163	162
ω_2	-163	-166
ω_3	163	156
ω_4	-163	-164

Ring Torsion Angles (degrees) of:

a Calculated TTTT S4
b c-Iabu-L-Phe-D-Pro-L-X^2

TABLE II. TTTT S4 Conformations.

	a	b	c	d	e	f
ϕ_1	-121	-121	-125	-121	-130	129
ϕ_2	-83	-94	-85	-91	-91	83
ϕ_3	121	121	125	124	128	-124
ϕ_4	83	94	85	94	91	-85
ψ_1	65	66	71	61	63	-70
ψ_2	166	170	166	171	166	-165
ψ_3	-65	-65	-71	-65	-67	68
ψ_4	-166	-170	-166	-165	-159	162
ω_1	2	6	-7	2	1	6
ω_2	-178	-171	-173	-169	-170	175
ω_3	-2	-5	7	-7	-5	-3
ω_4	178	171	173	175	-179	-175

Ring Torsion Angles (degrees) of:

a Calculated CTCT Ci
b c-Sar$_4$[6]
c c-(GlySar)$_2$[7]
d c-Gly-Sar$_3$[7]
e c-D-Ala-Sar$_3$[7]
f c-L-Ala-Sar$_3$[7]

Table III. CTCT Ci Conformations.

	a	b	c
ϕ_1	−121	−150	−120
ϕ_2	−83	−106	−90
ϕ_3	121	150	130
ϕ_4	83	106	60
ψ_1	65	51	60
ψ_2	166	142	160
ψ_3	−65	−51	−85
ψ_4	−166	−142	−150
ω_1	2	19	0
ω_2	−178	−178	180
ω_3	−2	−19	0
ω_4	178	178	180

Ring Torsion Angles (degrees) of:

a Calculated CTCT Ci
b c-D-MeVal-L-HyIv-L-MeVal-D-HyIv[8]
c D-MeAla(1)-Tentoxin (L1)[5]

TABLE III. Continued.

Cyclotetrasarcosyl adopts the calculated minimum energy conformation (tetra-N-methylated) and the correspondence between the observed [6] and calculated conformations is remarkable. This is also the case for c-(GlySar)$_2$ [7] which adopts the minimum energy di-N-methylated conformation.

The other two synthetic peptides (c-D,L-AlaSar$_3$ and c-GlySar$_3$ [7]) are both tri-N-Methylated and therefore one would expect them to adopt one of the CTCT conformations. They can both be accommodated by the CTCT Ci conformation with the N-methyls in cis positions and so this conformation is preferred as being lower in energy than any other.

The other examples are discussed elsewhere.

5.6 CTCT C2 Conformations
─── ──── ── ──────────────

Four observed structures correspond to the CTCT C2 conformation and these are shown in Table IV. The two Tentoxins [4,5] and the depsipeptide [10] are discussed later. The remaining depsipeptide c-D-HyIv-L-MeIleu-D-HyIv-L-MeLeu [9] is readily dealt with.

Di-N-Methylation has the effect of making all of the CTCT conformations lower in energy (when N-Methylated at cis positions) than the TTTT conformations. In general then, a di-N-Methylated peptide will adopt a CTCT conformation in preference to an all-trans one, if this can be done with the N-Methyls in the cis positions.

The configuration of the depsipeptide (DLDL) can be accommodated, without inner side chains, by the CTCT C2 (N-Methylated at cis positions) and all of the TTTT conformations. In this case the CTCT Ci conformation requires two bulky inner side chains or the N-Methyls at trans positions. The CTCT C2 conformation is preferred since it is some 3 kcal/mole lower in energy than any of the all-trans conformations. This is in contrast to dihydrochlamydocin which cannot be accommodated by any of the CTCT conformations.

	a	b	c
ϕ_1	-98	-108	107
ϕ_2	31	91	-74
ϕ_3	-98	-102	114
ϕ_4	31	66	-48
ψ_1	55	36	-57
ψ_2	152	146	-123
ψ_3	55	39	-34
ψ_4	152	121	-134
ω_1	-176	-180	172
ω_2	1	3	-1
ω_3	-176	177	-179
ω_4	1	6	-18

Ring Torsion Angles (degrees) of:

a Calculated CTCT C2
b c-D-HyIv-L-MeIleu-D-HyIv-L-MeLeu[9]
c D-MeVal-D-HyIv-D-MeVal-L-HyIv[10]

TABLE IV. CTCT C2 Conformations.

5.7 Tischenko Depsipeptides

A series of depsipeptide crystal structures have been determined by a Russian group under Tischenko [8,10,11,12]. The basic depsipeptide is c-(MeVal-HyIv) and four configurations have been synthesised and their structures determined: LDDD, LLDD, DDDL & DLLD.

These structures are remarkable in that all four configurations adopt different conformations — a result which can only be attributable to the changing pattern of side chain configuration. This is most striking proof that side chains play a deciding role in a cyclic peptide's choice of ring conformation.

The effect of replacing an amino acid by a hydroxy acid is to greatly increase the trans preference of the group. The hydroxy acid groups of depsipeptides will therefore invariably occupy trans positions. This in turn has a considerable effect on the conformation adopted by the depsipeptide.

As previously mentioned the DLLD configuration adopts the CTCT Ci conformation. This is because the configuration can be accommodated by that conformation without any inner side chains and with N-Me's at cis positions and hydroxy acids in trans positions. In contrast to this the LLDD configuration cannot be fitted into the CTCT Ci without having the hydroxy acids cis, nor can it adopt the CTCT C2 conformation without an inner side chain. The conformation adopted is the TTTT Ci, not the TTTT S4, as is shown by Table V.

The reasons for this choice are not immediately obvious although di-N-methylation reduces the difference in energy between the two lowest all-trans conformers to about 1 kcal/mole. Deformation of the hydroxy acids by the 17 degrees out of plane required by both of these conformations involves considerable strain and any small differences in the degree of deformation required would be enough to tip the balance in favour of one or other of the conformations.

	a	b
ϕ_1	104	76
ϕ_2	103	125
ϕ_3	-104	-76
ϕ_4	-103	-125
ψ_1	71	49
ψ_2	-63	-94
ψ_3	-71	-49
ψ_4	63	94
ω_1	-163	-163
ω_2	-163	-168
ω_3	163	163
ω_4	163	168

Ring Torsion Angles (degrees) of:

a Calculated TTTT Ci
b c-L-MeVal-L-HyIv-D-MeVal-D-HyIv[11]

TABLE V. TTTT Ci Conformations.

	a	b
ϕ_1	−121	−126
ϕ_2	−83	−117
ϕ_3	121	−116
ϕ_4	83	99
ψ_1	−166	−147
ψ_2	65	83
ψ_3	166	65
ψ_4	−65	−41
ω_1	178	166
ω_2	2	5
ω_3	−178	174
ω_4	−2	−14

Ring Torsion Angles (degrees) of:

a Calculated CTCT Ci
b c-L-MeVal-D-HyIv-D-MeVal-D-HyIv[12]

TABLE VI. Transition Conformer.

The DDDL configuration is more straightforward in that the conformation adopted is the lowest energy available (since di-N-methylation makes the CTTT C1 higher in energy than the CTCT C2). The cis positions can be occupied by N-methylated amides with the hydroxy acids in trans positions without causing any steric crowding.

The final configuration to be considered is LDDD and as can be seen from Table VI the conformation is based on the CTCT Ci. One amide group deviates from the calculated conformation in such a way as to form a transition state between the CTCT Ci and C2 conformations as is illustrated by Figure 5.5.

None of the CTCT conformations can accommodate the LDDD configuration with the hydroxy acids trans without at least one inner side chain. It appears that this is preferable to one of the all-trans conformations involving non-planar hydroxy acids which is hardly surprising since the di-N-methylated CTCT Ci is more than 5 kcal/mole lower in energy than the corresponding all-trans conformations. This conformation is chosen as the most favourable and the hindered side chain has been moved to a less crowded position by the deviation from the calculated conformation.

5.8 Tentoxin and Derivatives

Considerable interest has been shown in Tentoxin which is a phytotoxic metabolite of the fungus Alternaria tenuis [15]. When applied to germinating seedlings Tentoxin causes chlorosis in some plant species but has little or no effect on others [16]. This selective toxicity has been attributed [17] to the presence of a Tentoxin binding site in chloroplast coupling factor 1 (CF1) a key protein involved in ATP synthesis [18]. Resistant species are thought to contain a form of CF1 which does not bind Tentoxin.

The configuration and sequence of the amino acids in Tentoxin is well established [6,13,14] and has been confirmed by total synthesis [19]. The conformation adopted is not so well defined and two different conformations
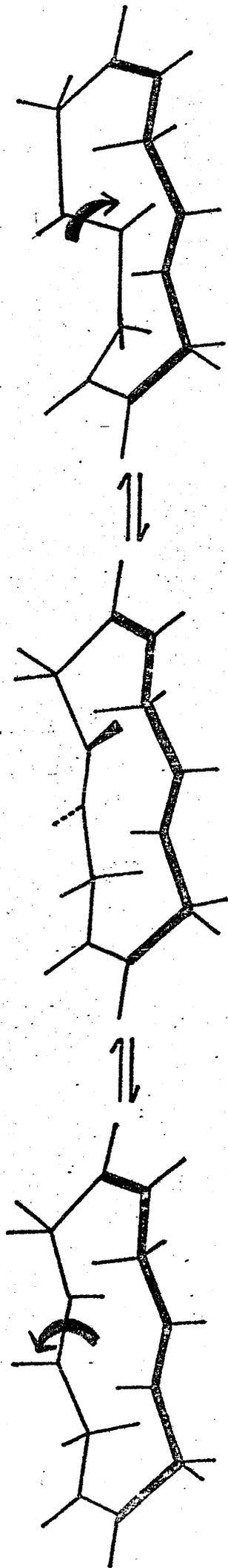
Figure 5.5. CTCT Ci -> CTCT C2 Interconversion with Observed Intermediate Conformation.
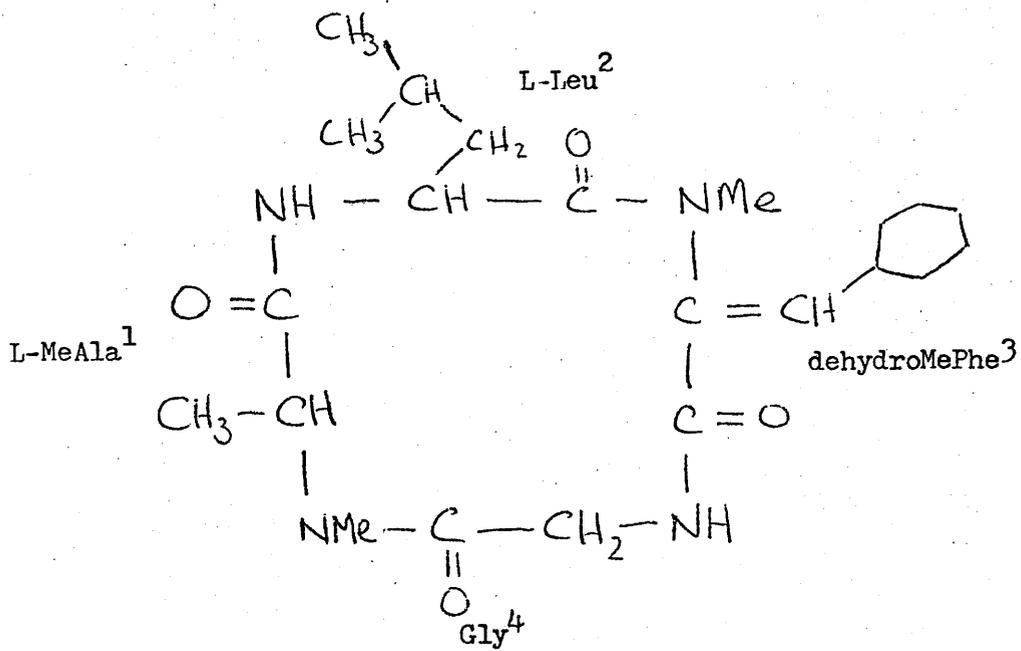
Tentoxin
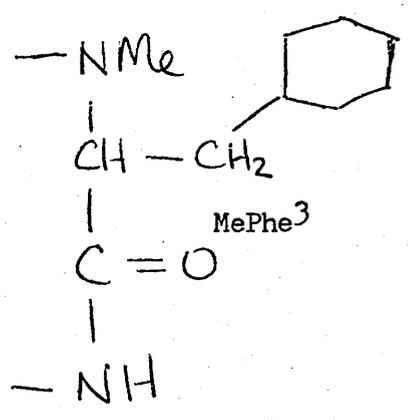
dihydroTentoxin

have been suggested.

Meyer et al [3,13,14] propose that Tentoxin adopts the CTCT Ci conformation on the basis of n.m.r. studies [14] and a crystal structure of dihydroTentoxin [13]. The conformation of the dihydro derivative is indeed the CTCT Ci with the N-methylated groups in the cis positions. This requires that the methyl side chain of the alanine residue be in an inner position. This tends to suggest that a single inner methyl group can be tolerated in some circumstances. Meyer interprets the n.m.r. spectrum of Tentoxin in terms of a conformation similar to that observed for the dihydro derivative.

Rich [4] has interpreted the n.m.r. spectrum of Tentoxin in favour of a CTCT C2 conformation which corresponds (Table VIII) to the calculated conformation number 10. Rich replaced the L-MeAla(1) residue of Tentoxin by L-Pro which has the effect of fixing part of the conformation [4]. L-Pro(1)-Tentoxin cannot adopt the CTCT Ci conformation without the bulky leucine side chain being in an unfavourable inner position. Despite this the n.m.r. spectrum and biological activity of L-Pro(1)-Tentoxin are very similar to Tentoxin itself. This strongly suggests that they both adopt the same conformation and that it is NOT the CTCT Ci.

The complete interpretation of the n.m.r. spectrum of L-Pro(1)-Tentoxin was facilitated by the restriction of the L-Pro residue and a complete set of torsion angles were obtained (see Table VIII). These angles were obtained using the method of Bystrov [20] and are estimated within +/- 20 degrees. If the proposed conformation is correct it is difficult to explain the difference in conformation between Tentoxin and its dihydro derivative. It is possible that dihydroTentoxin adopts different conformations in solution and in the solid state, although this is unlikely in view of the homogeneity of conformation previously observed [1,7].

The situation is further confused by D-MeAla(1)-Tentoxin for which Rich [5] has observed three conformers. One of these is the CTCT Ci and

	a	b	c	d
ϕ_1	80	60	77	60
ϕ_2	−143	−140	−19	−60
ϕ_3	−64	−80	75	90
ϕ_4	−2	−2	121	140
ψ_1	−109	−135	−166	−160
ψ_2	72	70	−68	−60
ψ_3	−39	−30	55	−20
ψ_4	−78	−90	−67	−80
ω_1	−1	0	−4	0
ω_2	180	180	178	180
ω_3	3	0	1	0
ω_4	−179	180	176	180

Ring Torsion Angles (degrees) of:

a Calculated CTCT C1 (Conformer 13)
b D-MeAla(1)-Tentoxin   (L2)[5]
c Calculated CTCT C1 (Conformer 3)
d D-MeAla(1)-Tentoxin   (U)[5]

TABLE VII. Asymmetric Tentoxin.

	a	b	c
ϕ_1	-79	-87	-80
ϕ_2	-121	-125	-120
ϕ_3	-79	-90	-90
ϕ_4	-121	-125	-125
ψ_1	-14	5	-10
ψ_2	69	60	70
ψ_3	-14	-20	-20
ψ_4	69	60	70
ω_1	3	0	0
ω_2	179	180	180
ω_3	3	0	0
ω_4	179	180	180

Ring Torsion Angles (degrees) of:

a Calculated CTCT C2 (Conformer 10)
b Tentoxin[4]
c L-Pro(1)-Tentoxin[4]

TABLE VIII.
Tentoxin C2 Conformers.

another the asymmetric conformation No. 13. These two conformations are in equilibrium and have a biological activity approaching that of Tentoxin. This immediately suggests that the conformation of Tentoxin could be the CTCT Ci as indicated by the crystal structure of dihydroTentoxin. The balance of probability seems to indicate that this is indeed the case and that Rich's proposed conformation applies only to L-Pro(1)-Tentoxin and not to Tentoxin itself. The effect of the dehydro residue on conformation is not clearly defined and further work is required on this point.

The other conformer of D-MeAla(1)-Tentoxin observed by Rich corresponds to an asymmetric conformation (calculated No. 3) which is similar in some respects to the CTCT Ci. Rich's estimated torsion angles are shown in Table VII.

5.9 Other 12-membered Rings

The similarities between the observed and calculated conformers even in cases with bulky substituents led to a consideration of the general applicability of the calculated conformer set to other 12-membered ring systems. Of these the most "peptide-like" systems are the 12-crown-4 ethers (1,4,7,10-tetraoxa-cyclododecanes) and cyclododeca-1,4,7,10-tetraenes.

The structures of four 12-crown-4 ether complexes have been determined by X-Ray experiments. In three of these [21,22,23] the conformation adopted is analagous to the TTTT C4 (see Table IX) while the fourth [24] is a TTTT Ci conformation (Table X). The reason for this seems to be that sodium and calcium ions directly complex to the ring oxygens and therefore all four oxygens need to be on the same side of the ring (as is the case with the TTTT C4). The magnesium ion is fully hydrated and therefore the ether conformation is not critical to complex formation. This is illustrated by Figures 5.6 and 5.7.

This result would tend to suggest that TTTT Ci might be the preferred conformation of 12-crown-4 ethers.

	a	b	c	d
ϕ_1	84	−84	−79	78
ϕ_2	84	−84	−81	81
ϕ_3	84	−84	−80	79
ϕ_4	84	−84	−80	83
ψ_1	59	−56	−60	59
ψ_2	59	−56	−59	60
ψ_3	59	−56	−56	56
ψ_4	59	−56	−59	59
ω_1	−165	165	162	−165
ω_2	−165	168	165	−165
ω_3	−165	165	163	−164
ω_4	−165	168	166	−163

Ring Torsion Angles (degrees) of:

a Calculated TTTT C4
b 12-Crown-4 Calcium Chloride[21]
c 12-Crown-4 Sodium Hydroxide[22]
d 12-Crown-4 Sodium Chloride[23]

TABLE IX. 12-Crown-4 C4 Conformers

	a	b
ϕ_1	104	98
ϕ_2	103	91
ϕ_3	-104	-98
ϕ_4	-103	-91
ψ_1	71	83
ψ_2	-64	-71
ψ_3	-71	-83
ψ_4	64	71
ω_1	-163	-147
ω_2	-163	-116
ω_3	163	147
ω_4	163	116

Ring Torsion Angles (degrees) of:

a Calculated TTTT Ci
b 12-Crown-4 Magnesium Chloride[24]

TABLE X. 12-Crown-4 Ci Conformer.
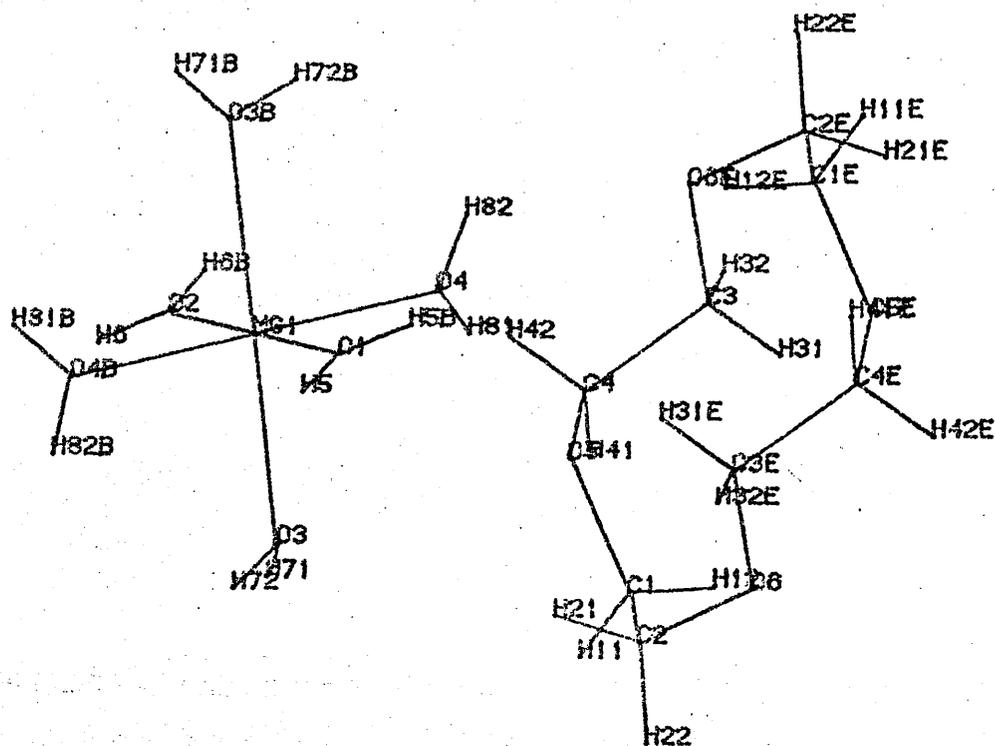
Figure 5.6 Crystal Structure of
12-Crown-4 ether complex with Mg
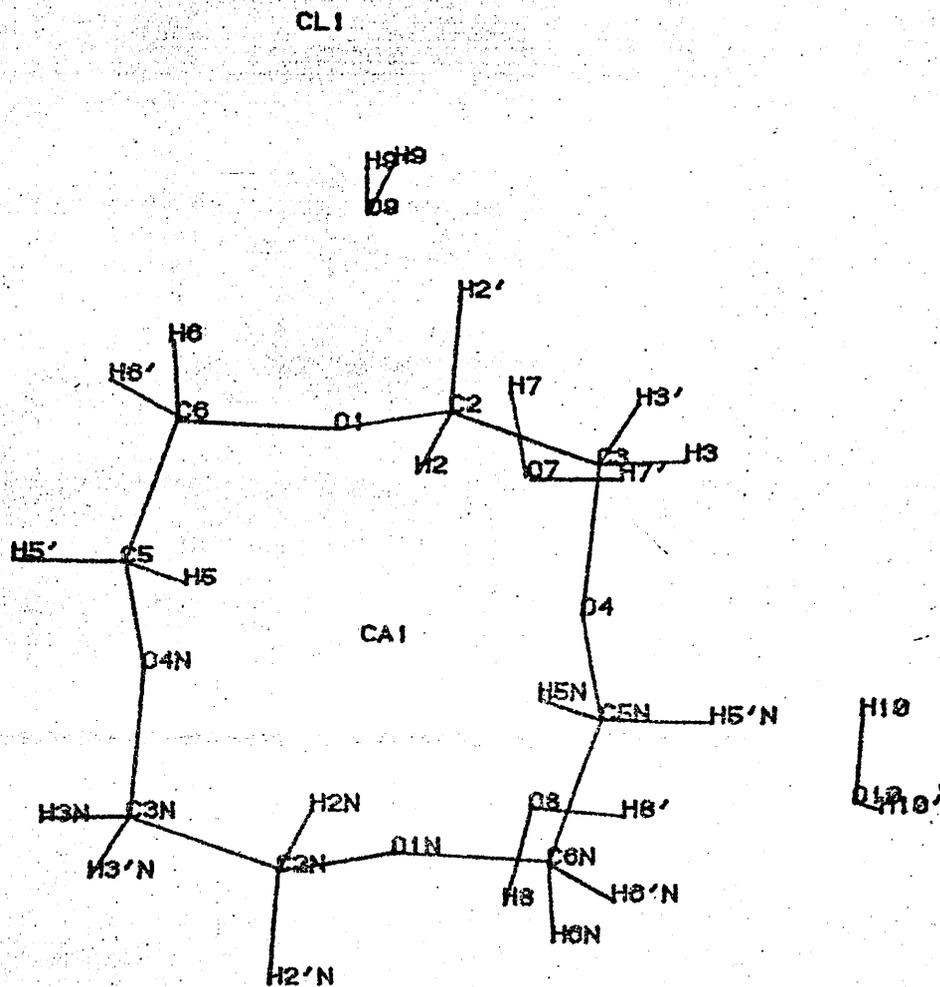
Figure 5.7 Crystal Structure of
12-crown-4 ether complex with Co

No crystal structures have been determined of cyclododecatetraenes and in fact the only conformational evidence comes from an n.m.r. study of a tetrabenzo compound (I).

(I)
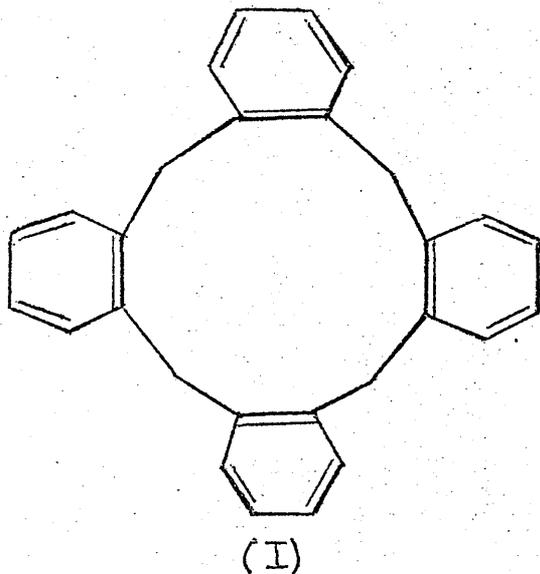
The conformation of this has been assigned [25] to a C2h "sofa-like" structure which could correspond to one of the calculated all-cis C2 conformers. However the bulk of the substituents makes a direct comparison impossible.

Two further 12-membered rings are of interest one of which is cyclododecane [26] while the other is azacyclododecane [27] (cyclododecane with one carbon atom replaced by nitrogen). Somewhat surprisingly both of these compounds adopt conformations closely resembling the TTTT C4 as is shown by Table XI.

	a	b	c
ϕ_1	84	−68	64
ϕ_2	84	−70	66
ϕ_3	84	−67	67
ϕ_4	84	−69	70
ψ_1	59	−69	71
ψ_2	59	−67	71
ψ_3	59	−70	69
ψ_4	59	−68	68
ω_1	−165	161	−154
ω_2	−165	155	−169
ω_3	−165	163	−155
ω_4	−165	155	−168

Ring Torsion Angles (degrees) of:

a Calculated TTTT C4
b Cyclododecane[26]
c Azacyclododecane[27]

TABLE XI. Cyclododecanes.

5.10 References

1. K. Titlestad, Acta. Chem. Scand., B31, 641 (1977).

2. J.L. Flippen and I.L. Karle, Biopolymers, 15, 1081 (1976).

3. W.L. Meyer, L.F. Kuyper, R.B. Lewis, G.E. Templeton and S.H. Woodhead, Biochem. Biophys. Res. Commun., 56, 234 (1974).

4. D.H. Rich and P.K. Bhatnagar, J. Amer. Chem. Soc., 100, 2212 (1978).

5. D.H. Rich and P.K. Bhatnagar, ibid, 100, 2218 (1978).

6. P. Groth, Acta. Chem. Scand., 24, 780 (1970).

7. J.P. Declercq, G. Germain, M. van Meersche, T. Debaerdemaeker, J. Dale and K. Titlestad, Bull. Soc. Chim. Belg., 84, 275 (1975).

8. G.N. Tischenko, Z. Karimov and V.V. Borisov, Biorg. Khim., 1, 378 (1975).

9. J. Konnert and I.L. Karle, J. Amer. Chem. Soc., 91, 4888 (1969).

10. G.N. Tischenko, N.V. Nazimova, V.I. Andrianov and Z. Krimov, Biorg. Khim., 1, 386 (1975).

11. Z. Karimov, A.M. Mikhailov and G.N. Tischenko, ibid, 2 (1976).

12. G.N. Tischenko, V.I. Smirnova, L.N. Zeibot, N.V. Nazimova and V.I. Adrianov, ibid, 2, 895 (1976).

13. W.L. Meyer, L.F. Kuyper, D.W. Phelps and A.W. Cordes, J. Chem. Soc. Chem. Comms., 339 (1974).

14. W.L. Meyer, L.F. Kuyper, D.W. Phelps and A.W. Cordes, J. Amer. Chem. Soc., 97, 3802 (1975).

15. G.E. Templeton, Microb. Toxins, 8, 160 (1972).

16. N.D. Fulton, K. Bollenbacher and G.E. Templeton, Phytopathology, 55, 49 (1974).

17. J.A. Steele, T.F. Uchytil, R.D. Durloin, P. Bhatnagar and D.H. Rich, Proc. Nat. Acad. Sci. U.S.A., 73, 2245 (1976).

18. H.S. Penefsky, Enzymes 3rd ed., 10, 375 (1974).

19. D.H. Rich and P. Mathiaparanam, Tetrahedron Letters, 4037 (1974); D.H. Rich, P. Bhatnagar, P. Mathiaparanam, J.A. Grant and J.P. Tam, J. Org. Chem., 43, 296 (1978).

20. V.F. Bystrov, V.T. Ivanov, S.L. Portnova, T.A. Balashova and Yu. A. Ovchinnikov, Tetrahedron, 29, 873 (1973).

21. P.P. North, C. Steiner, F.P. van Remoortere and F.P. Boer, Acta. Cryst., B32, 370 (1976).

22. F.P. Boer, M.A. Neuman, F.P. van Remoortere and E.C. Steiner, Inorg. Chem., 13, 2826 (1974).

23. F.P. van Remoortere and F.P. Boer, ibid, 13, 2071 (1974).

24. M.A. Neuman, E.C. Steiner, F.P. van Remoortere and F.P. Boer, ibid, 14, 734 (1975).

25. J.D. White and B.D. Gesner, Tetrahedron, 30, 2273 (1974).

26. J.D. Dunitz and H.M.M. Shearer, Helv. Chim. Acta., 43, 18 (1960).

27. J.D. Dunitz and H.P. Weber, Helv. Chim. Acta., 47, 1138 (1964).

28. P.N. Lewis, F.A. Momany and H.A. Scheraga, Isr. J. Chem., 11, 121 (1973).

APPENDIX

Program GLOMIN

A.1 Introduction
─── ────────────

The algorithm, described in Chapter 4, for locating the GMEC of cyclic molecules originally consisted of several different computer programs. Transfer of data between these programs was by means of disk files which required varying amounts of manual editing and correction. To avoid this and to streamline the whole process it was decided to unify the process into a single program. The resulting program requires only one input and one output file and all control variables are set up interactively when the program is started.

The program checks all typed data for validity and rerequests any incorrectly entered items. The program can also be aborted at any stage without loss of data and restart information is provided.

The following section describes the program dialogue and the following points should be noted:
(1) [Y/N] implies a yes/no answer;
(2) [I] implies one or more integer answers; &
(3) [F] implies one or more real number answers.
The symbol '>' is used as a prompt where a list of items is required. The immediately preceding question should be answered the required number of times.

A.2 Example Dialogue
─── ─────── ────────

The program first prints the contents of an information file giving an up‑to‑date summary of restrictions, hints, bugs etc. followed by the requests for the various items required.

DOES THE FOLLOWING DATA REFER TO A HOMOPOLYMER [Y/N] ?
If the answer to this is yes the program can then check for sequences of generators which are cyclically redundant. This check is not appropriate in cases where the ring is composed of dissimilar units.

NUMBER OF VARIABLE TORSION ANGLES [I]

There are n-3 potentially variable torsion angles in the n-atom linear chain to be folded into an n-membered ring. These can be regarded as independent variables leaving 3 dependent variables (see later) which are only meaningful once the folding process has been completed. However in some cases it is necessary to fix some of the normally variable torsion angles as for example in rings containing double bonds or amino acids.

TORSION ANGLES BY PAIRS [Y/N] ?

Normally torsion angles would be specified singly however peptides require a pairwise specification of two torsion angles (Φ & Ψ). Both torsion angles and generators should be entered in pairs if a yes answer is entered here. Otherwise the message:

***YOU HAVE SELECTED INDIVIDUAL TORSION ANGLES.

is printed and torsion angles and generators should be entered singly.

ARE THE BOND ANGLES TO REMAIN FIXED [Y/N] ?

In order to achieve ring closure in a small ring (< 8 atoms or < 4 amino acids) it is necessary to allow the valence angles to vary. A no answer here enables this facility and the following message is given:

***BOND ANGLES WILL VARY DURING RING CLOSURE.

and bond angles are varied by up to 20 degrees in an attempt to get the chain ends to the required separation.

NAME OF FILE:

The name of the file containing the starting coordinates etc. should be entered at this point. Only those atoms which will form the target ring are required - pendant atoms (e.g. Hydrogens) should be omitted from the file. The title contained in this file is printed together with any errors in it.

ENTER NO OF GENERATORS [I]

The number of (pairs of) torsion angle values to be used as generators is

entered here. Thus for cyclotetraglycyl as described earlier this would be 6 pairs while for a cycloalkane this would be 3 singles (60, -60 & 180 degrees).

ENTER A GENERATOR (IN DEGREES) [F]

>

The prompt symbol '>' will appear for the number of times specified by the last response. In reply the user should enter a single generator, or generator pair, as appropriate.

ENTER NUMBER OF EXCLUSION STRINGS (MAX 10) [I]

For some types of compounds it is possible to predict that certain combinations of generators can not result in cyclic conformations. An example of this is cycloundecane for which any generator sequence which includes three consecutive appearances of the same generator makes it impossible to get the chain ends within the required separation. Thus in this case one would enter "3" here and in response to the further prompt:

ENTER AN EXCLUSION STRING

>

one would enter 111, 222 & 333. This results in all generated conformers which contain any of the specified sequences being rejected (e.g. 11122332 and 11223331 would be rejected).

ENTER TARGET END-TO-END DISTANCE AND MAXIMUM DEVIATION THEREFROM [F]

The target end-to-end distance depends on the type of bond to be formed at the ring junction. If it was simply a Carbon-Carbon single bond then 1.54 (Ångstroms) would be input here. The program checks that the ends of the chain, after folding, are the required distance apart. A simple pattern search procedure is used to attempt to close those rings which do not initially have the correct end-to-end distance. This involves varying the torsion angles (and valence angles, if allowed) by up to 20 degrees. Those which do not close to within the specified deviation are rejected. (A

deviation of +/- .1 Å from the target distance is usually suitable.)

ENTER SHORTEST TOLERABLE NONBONDED DISTANCE [F]

ENTER MAXIMUM PERMISSIBLE NUMBER OF SUCH CONTACTS [I]

These two questions control the degree of steric checking performed on each generated conformer. All nonbonded contacts shorter than the specified distance are counted and if the total exceeds the number entered in response to the second prompt the conformer is rejected. This is a very gross check since only the ring atoms are involved. The values chosen are not critical - no nonbonded contacts shorter than 1.9 or 2 Ångstroms gives reasonable results.

DO YOU WANT TO RESTRICT THE DEPENDENT TORSION ANGLES [Y/N] ?

As previously mentioned, 3 of the ring torsion angles are undefined until ring closure has been completed. It may be necessary to restrict the acceptable values of one or more of these. This would arise where one of the torsion angles was a double bond or an amide bond. The required details can be set by responding to the following prompts in the appropriate way:

ENTER NUMBER OF DEPENDENT T.A.'S TO BE RESTRICTED [I]

ENTER T.A., DESIRED VALUE & TOLERANCE

>

Information about the starting conditions is required since it is possible to halt the program in mid-execution and restart it at a later stage. This is entered in response to the two prompts:

ENTER STARTING CONFORMER NUMBER [I]

ENTER STARTING GENERATOR SEQUENCE [I]

A starting conformer number of n indicates that the next generated conformer which passes all of the checks will be the nth to do so. This is 1 for the first run and is obtained from the restart information printed when the program is aborted in other cases. The terminal output on aborting the program is as follows:

***ABORT REQUEST RECEIVED.

RESTART INFORMATION:

NEXT CONFORMER GENERATED WOULD HAVE BEEN NO. n

CURRENT GENERATOR SEQUENCE IS mmmmmm

The printed values are used to restart the program so as to avoid duplication.

The final input required is the name of a disk file which will contain the coordinates of the generated conformers and is entered in response to:

OUTPUT FILE NAME:

This file is written in a format which can be read by the Chemical Graphics System described in Chapter 3. This system can then be used to add Hydrogens and other pendant atoms prior to energy minimisation.