Conformational Studies on some Inhibitors of Thermolysin and EC 3.4.24.11

A Thesis presented for the degree of Doctor of Philosophy

in

The Faculty of Science

of

The University of Glasgow

by

Fiona Ruth Ferguson Forrest

March,1987.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ACKNOWLEDGEMENTS

I am deeply indebted to Professor G.A.Sim and Dr. J.K.Tyler for their advice, help and support during the period of this research.

Thanks are also due to Dr.A.J.Baker and Dr.P.R.Mallinson who were ever willing to lend an ear and to many other members of the Chemistry Department.

Special thanks go to Dr.M.S.Tute and Dr.M.Snarey of Pfizer Central Research, Sandwich, Kent, whose guidance was invaluable, and who, along with Dr.D.N.J.White, initiated this project.

I would also like to thank the Science and Engineering Research Council and Pfizer Central Research for their financial support during the period of this research.

Finally, I would like to express my gratitude to my family and friends for their help, encouragement and good humour during the three years, especially Maureen T. Travers, Dr. Marie A. Russell, Dr. David H. Kitson and my husband Dr. Douglas Hayes for his patience and encouragement during the write up of this Thesis.

Fiona Ruth Ferguson Forrest.

To Robert and Jean Forrest.

Contents

Summary

Molecular modeling is an exciting new approach in the field of drug design. A molecule's activity as a drug is dependent upon its conformation i.e. the three dimensional arrangement of its functional groups.

Molecular mechanics coupled with interactive computer graphics is an excellent technique for studying the conformations of small molecules and this thesis presents a discussion of this technique and its application to some inhibitors of the enzymes Thermolysin and EC 3.4.24.11.

An introduction to the role of molecular mechanics in computer assisted drug design is given in Chapter One. Chapter Two describes the calculation of steric energies and the potential functions used to do this, while the procedures used for conformation generation and energy minimisation are discussed in Chapter Three (Appendix A lists one of the programs used for the conformation generation).

The MOL software package is an interactive modeling system for small molecules which has been developed at Glasgow and Chapter Four contains a description of this system and its options.

The structure and function of many enzymes is of prime importance in drug design since drugs often act by inhibition of enzymic pathways. Since very few enzyme structures are known from X-ray crystallography the primary source of information is enzyme kinetics and Chapter Five discusses some important properties of enzymes, such as their functions as catalysts, substrate and inhibitor kinetics and evolutionary pathways.

EC 3.4.24.11. (a mammalian enzyme) and Thermolysin (EC 3.4.24.4, a bacterial enzyme) have very similar substrate specificities and are inhibited at comparable rates by several inhibitors. Very little is known about the active site of EC 3.4.24.11., however, the structure of Thermolysin is well known from X-ray crystallography and in Chapter Six the known features of active sites of both enzymes are discussed and compared.

It is known that two isosteric inhibitors of Thermolysin, β-Phenylpropionyl-L-Phe and Carbobenzyoxy-L-Phe, bind in completely opposite orientations. This unexpected phenomenancan be explained by conformational studies on the two inhibitors. In both cases the enzyme does not bind the lowest energy conformer, however in each case it binds the conformer which has both low energy and gives a good fit to the active site.

The results of conformational studies of inhibitors of EC 3.4.24.11. are also discussed in Chapter Six and comparison of the low energy conformers with inhibitors of Thermolysin (from X-ray data) point to differences in the positions and sizes of the S_1 and S_2' subsites of the two enzymes. Both enzymes are highly specific for hydrophobic groups in the S_1' site, with Phe being the most favoured residue. From the X-ray studies of Thermolysin it is known that the Phe side chain is bound with torsion angles of roughly 170° and 80°. It is interesting to note that these torsion angle values are very common in many of the low energy conformers and it seems reasonable to assume that the high level of specificity arises from the fact that the side chain of the Phe, Leu etc. are normally in the most favourable position for binding to the S_1' hydrophobic pocket without rearrangement. Several other differences

between the two enzymes lead to the conclusion that their functional similarity is more likely to be a product of evolutionary convergence on function than of divergence from a common ancestor.

Chapter Seven contains a brief summary of the programming which was done during the course of this project paying particular attention to both the motivation and methods involved in programming an array processor and as previously mentioned Appendix A contains a listing of the GLOMIN program - a conformation generation program which was modified to run on an array processor during the course of this project.

Finally, Chapter Eight contains a discussion of work carried out in conjunction with Professor A.Y.Meyer while he was on sabbatical leave at Glasgow University. This work shows that force fields can be modified quite simply to give more transferable non-bonding parameters by the inclusion of an electrostatic term. A paper on this work has been published in the Journal of Computational Chemistry and a copy of this paper is given in Appendix B.

Chapter One

Introduction

In recent years there have been major advances in the field of medical research. To keep pace with this progress the pharmaceutical industry has invested vast resources in the field of drug research and development. The industry has also responded to the growth in public awareness about disease and medical treatment by attempting to improve the range, potency and safety of many drugs.

The method most commonly used in drug design centres on an 'active analog' approach in which medicinal chemists use reasoning and intuition to design new drugs by modifying existing potent compounds - a time consuming and expensive process which is not always successful. One of the failings of this approach is that, although it is appreciated that the activity of a compound and its three dimensional structure are directly related, the synthetic chemist generally deals with two dimensional images of molecules which may not bear much resemblance to their three dimensional structures.

An exciting new approach to drug design is currently being developed which eliminates the 2-d, 3-d problem. This approach involves the use of theoretical chemistry and interactive computer graphics in the study of molecular properties and structures.[1,2] High resolution graphics terminals provide an excellent medium with which to display large, or complex structures in three dimensions[3-6] and various theoretical techniques exist with which structural properties can be calculated.

At the moment this approach is still relatively novel, but it is hoped that eventually the prediction of potent compounds using

computational techniques will be at least as reliable as the current trial and error methods. There is so much interest in the concept of computer-assisted drug design that many universities and most large pharmaceutical companies now have some form of molecular modeling facility.

The area to which computational chemistry has been applied in this project is the study of structure-activity relationships, which play a key role in drug design. The majority of drugs act by blocking metabolic pathways - usually by enzyme inhibition. The structures of the enzymes, inhibitors and their complexes are therefore very important. X-ray crystallography has provided us with highly accurate information on the binding of a few enzyme-inhibitor complexes and molecular modeling has enabled us to elucidate many of the pertinent features of the binding interactions. However, the structures of several enzyme-inhibitor systems remain undetermined and it is in this area that the predictive powers of computational chemistry may best be used.

There are several experimental and theoretical methods currently available with which information about molecular structure and energy can be determined. Experimental techniques include spectroscopic methods (such as nuclear magnetic resonance, infrared, ultraviolet, optical rotatory dispersion and circular dichroism) and diffraction methods (such as X-ray, electron and neutron diffraction). All of which will give highly accurate information about molecular geometry, but little information about thermodynamic properties. On the other hand calorimetry will give good thermodynamic data, but little geometric information.

None of the individual experimental techniques can give reliable geometries and thermodynamic information , but such results can be obtained using the theoretical methods currently available. These include ab initio calculations and semi-empirical molecular orbital methods such as Extended Huckel Theory (EHT), neglect of non-bonded differential overlap (NNDO), intermediate neglect of differential overlap (INDO), modified intermediate neglect of differential overlap (MINDO), and complete neglect of differential overlap (CNDO). These methods are capable of giving reasonable geometries and thermodynamic data depending on their application. However, they are all highly c.p.u. intensive which restricts their use to small molecules.

Molecular mechanics (or the force field method) is much less demanding on computer time. It gives reasonable geometries and thermodynamic information[7] on the system by making good use of experimental data in conjunction with a mathematical representation of the system.

For the purpose of molecular mechanics a molecule is considered as a collection of atoms held together by forces which can be described using potential functions derived from classical mechanics. The set of empirically derived potential functions chosen to reproduce particular molecular properties is known as a 'force field'[8]. Several force fields exist which employ different potential functions . The nature of the individual functions is not important so long as the set of functions chosen gives good agreement with experimental data. The overall energy of the molecule can be calculated using these potential functions with parameters which have been obtained from existing experimental data.

The structures of large molecules can therefore be modelled

using information derived from the observed properties of smaller molecules. The energy of the molecule is minimised by altering the positions of the atoms. The only disadvantage of this method is the large body of reliable experimental data needed to parameterise the force field as results of this method are only as reliable as the force field parameters being used. Many force fields exist which can be used for alkanes[9], alkenes[10], hydrocarbons[11], peptides[12] and even proteins[13,14].

There are several reviews on molecular mechanics[15-21] and computer assisted drug design[22-27] which are relevant to this discussion.

Molecular Mechanics Studies

The compounds studied during the course of this project are either inhibitors of the mammalian enzyme EC 3.4.24.11.[28] or its bacterial equivalent, Thermolysin[29]. The enzyme EC3.4.24.11. has been isolated from several tissues, most notably the kidney[28] and the brain[30], where its role is that of an endopeptidase cleaving small peptides at the N terminus of hydrophobic residues. In the kidney it has been shown to cleave peptides such as angiotensin and bradykinin[31] whereas in the brain it is thought to play a role in the degradation of naturally occurring enkephalins[32]. The inhibition of EC 3.4.24.11. is of interest to the pharmaceutical industry for several reasons, including the likelihood of inhibitors finding possible application as analgesics[33].

Both Thermolysin and EC 3.4.24.11. are zinc metalloendopeptidases with similar substrate specificities, but with different sensitivities to certain inhibitors[34]. The two enzymes are

therefore not identical and we would like to compare features of their active sites to see how they differ. As yet the structure of EC 3.4.24.11. has not been solved; however, the X-ray structures of Thermolysin plus several inhibitors[35-40] have been published.

With the known structure of Thermolysin as a starting point, it has been possible to use low energy conformers of the inhibitors with binding information to devise possible inhibitor-enzyme interactions for the unknown active site of EC 3.4.24.11. This binding information may prove useful in the search for new, active compounds.

Chapter Two

Section 2.1

Molecular Mechanics

One of the most important principles of molecular mechanics calculations is that for any molecule there is a minimum energy conformation for which there are optimal values for bond lengths, bond angles and torsion angles. Deviations from this equilibrium state are unfavourable and give rise to an increase in molecular potential energy.

For any conformation of a molecule the potential, or steric energy can be expressed in terms of the sum of the energies arising from several different types of intramolecular interactions i.e.:-

$$V_s = V_l + V_\theta + V_w + V_{nb} + V_x + V_q + V_{cross}$$

The intramolecular interactions being considered are in order; bond stretching or compression, valence angle deviations (Baeyer strain), torsion angle variations (Pitzer strain), non-bonded (van der Waals) interactions, out-of-plane bending at trigonal, planar regions of the molecule, coulombic interactions and cross terms.

The first six terms are assumed to be independent of one another and cross terms, which express coupling between different components, tend to be ignored since they are small and only of real relevance in the calculation of vibrational frequencies.

It is important to note that by itself the absolute magnitude of V_s has no physical

meaning. However, differences in V_s for different geometries of the same molecule are appropriate for comparison to experimentally observable physical properties such as rotational barriers, or conformer populations.

The empirically derived set of equations which are used to calculate each interaction's contribution to the steric energy and the derivation of the parameters used in these equations will be discussed in the rest of this chapter.

Section 2.2

Bond Stretching or Compression

The expression for evaluating the energy involved in the stretching, or compression of a bond can be derived from Hooke's Law, in which bonded atoms are considered to be two masses joined by a spring undergoing simple harmonic motion. This gives us the following equation:-

$$V_1 = (1/2)\, k_1\, (l - l_o)^2$$

where V_1 is the bond stretching energy, k_1 is the bond stretching force constant, l is the observed bond length and l_o is the calculated equilibrium bond length.

The force constant k_1 has a range of values, typically between 300 and 1200 kcal mol^{-1}Å^{-2} and has a different value for every bond type in the molecule - the larger the bond order, the more strongly the atoms are held together and the larger is the value for k_1. Stretching bonds is therefore a high energy process - indeed the

deformation of a bond from equilibrium by 0.1Å can cause an energy increase of between 1.5 kcal mol^{-1} and 6 kcal mol^{-1}. Since bond stretching / compression is so energetically unfavourable it is reasonable to assume that bonds within a molecule will adopt values close to their equilibrium lengths.

Section 2.3

Angle Bending

Again Hooke's law is used to derive the expression for calculating the contribution bond angle bending makes to the total energy. For angles which deviate from the equilibrium value by up to 15° the following equation is used:-

$$V_\theta = (1/2)\, k_\theta\, (\theta - \theta_o)^2$$

where V_θ is the energy associated with angle bending, k_θ is the angle bending force constant, θ is the observed angle and θ_o is the calculated angle which varies with the degree of substitution. The above equation gives good results for deviations of up to 15° from the equilibrium value, but as distortions become greater angle bending cannot be described in terms of simple harmonics[1]. For such cases a more elaborate equation must be used:-

$$V_\theta = (1/2)\, k_\theta\, [\, (\theta - \theta_o)^2 - k_\theta' \, (\theta - \theta_o)^3 \,]$$

where k_θ' is the anharmonicity force constant.

Section 2.4

Torsion Angle Variations

The spatial relationship of any four consecutively bonded atoms in a molecule can be described by a torsion angle.

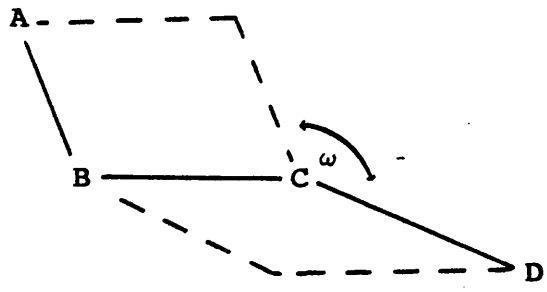


Fig 2.1

The value of the torsion angle for A B C and D can be defined as the angle w measured about B C from plane A B C to plane B C D i.e. looking along B C the torsion angle is the angle through which A must be rotated to eclipse D (see above diagram). According to the Prelog-Ingold convention[2] the torsion angle is designated by a +ve or -ve sign depending on whether the direction of rotation is clockwise or anti-clockwise, up to a value of 180^{o}.

Variations in torsional values in a molecule (i.e. rotations about single bonds) give rise to changes in that component of the molecule's steric energy known as the torsional strain.[3,4] The energy arising from such torsion angle variations can be calculated using:-

$$V_w = (1/2)\, k_w\, (1 + S \cos n w)$$

where V_w is the torsional energy, k_w is a force constant known as the barrier to free rotation, S has a value of +1 for eclipsed minima and -1 for staggered minima and accounts for vector direction, n is the periodicity of rotation (i.e. the number of times a minimum is found during a 360° rotation of the torsion angle) and w is the observed torsion angle.

Section 2.5

Non-bonded Interactions

The energy arising from non-bonded interactions is generally the most important component of the total energy since such interactions are a major factor in determining the stability of molecular conformations.

The concept of non-bonded, or van der Waals interactions is used to explain the interactions between atoms which are not joined to each other, or to a common atom.

The potential function which describes non-bonded interactions is derived from two independent forces which are related to interatomic separation (see Fig 2.2). There is a strong, repulsive force over short distances and a weaker, attractive force over larger distances.

The stronger force arises from the repulsion which occurs when the filled electronic orbitals of neighbouring atoms start to overlap. The expressions commonly used to describe this repulsion are:-

$$V = Ae^{(-Br)}$$

and

$$V = A / r^n$$

a)

R = Infinity

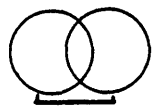
At large interatomic distances the weak attractive force predominates.

b)

R = 5Å

At a few Angstroms separation the attractive and repulsive forces begin to compete.

c)



R=1Å

At short interatomic separation the electron shells begin to overlap and the repulsive force predominates.

Fig. 2.2.

where A and B are constants for two interacting atoms, r is the interatomic separation and n has a value of 12 although occasionaly 9 is used.

Notice that the repulsion term is very large at small interatomic separations and very small at large interatomic separations. Therefore this force is only of real consequence for small interatomic separations.

The second component of the non-bonded force is the weak attractive force, or London dispersion force which results from induced dipole / dipole coupling. Electrons in molecular orbitals can synchronise their motion giving rise to a weak attractive force known as the London energy. This attraction varies with the sixth power of the interatomic separation and can be expressed by the function:-

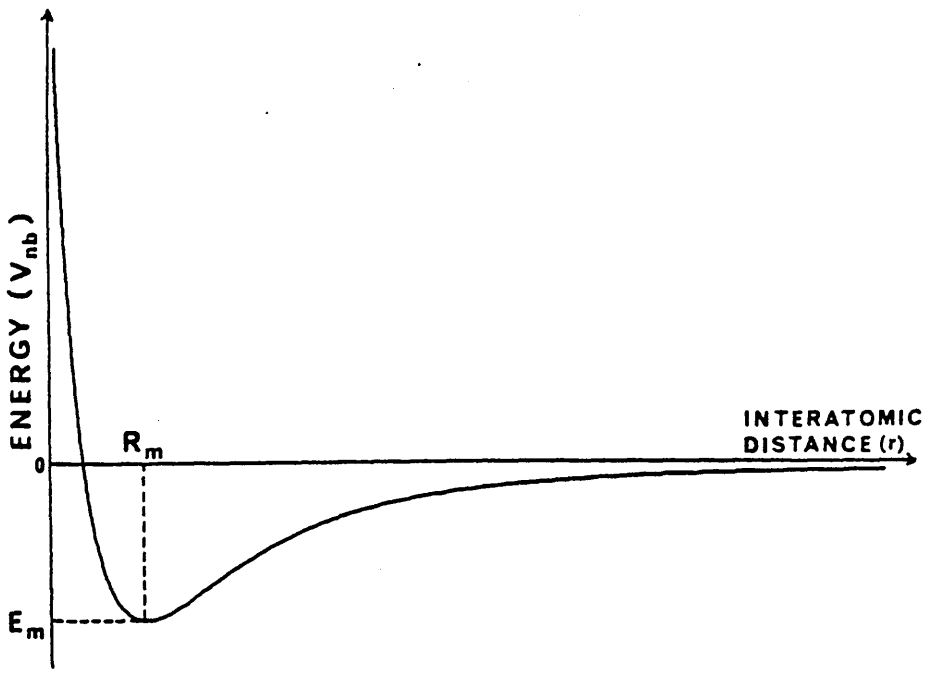
$$V = -P / r^6$$

where P is a constant related to the polarisability of the atoms and r is the interatomic separation.

This 'inverse sixth' attractive energy decreases rapidly with increasing r, but not nearly as rapidly as the van der Waals repulsion energy.

The functions describing the attractive and repulsive forces can be combined to give a single potential function representing non-bonded energy. The general shape of this function is shown in Fig. 2.3.

The non-bonded potential function can be described by either of two equations:-



Van der Waals Interaction Energy

R_m is the minimum interatomic distance at which the attractive and repulsive forces are balanced and E_m is the energy associated with this distance.

Fig. 2.3.

a) The Buckingham Potential[5]

$$V_{nb} = Ae^{(-Br)} - (P / r^6)$$

or

b) The Lennard-Jones 6 / 12 Potential[6]

$$V_{nb} = (A / r^{12}) - (P / r^6)$$

In the Lennard-Jones potential[6] A and P are related to the minimum energy point of the function and its associated distance and are defined by:

$$A = -E_{min} (R_{min})^{12}$$

$$P = -2E_{min} (R_{min})^{6}$$

For molecular mechanics calculations there are several problems arising from the use of these functions for the calculation of non-bonded interactions. The first problem is that the above functions were derived to explain intermolecular interactions and not intramolecular interactions. The above functions adequately represent the interactions between two atoms in different molecules in the gaseous phase which are independent of any non-bonded interactions with other atoms. It is obvious however that this is a very simplistic approximation to apply to the the non-bonded interactions occuring within a molecule, especially when we consider that the induced dipole moment from two interacting electrons may be considerably disturbed by

the dipole moments of other atoms within the molecule.[7] It is also true that the functions do not take into account the fact that atoms on opposite sides of the molecule will be shielded from one another by the electron density of the molecule. This means that the calculated interaction between two atoms will remain constant regardless of whether or not there is another atom directly between them.[8]

Another problem that arises with this approach is that the atoms are regarded as being spherical and therefore they have a consistent van der Waals radius. Although this is generally true for large atoms it is not always the case for smaller atoms.

However despite the many failings of this approach it still appears to give relatively accurate results. This is probably due to the fact that any errors arising either cancel each other out or are averaged out over the whole molecule.

There is very little to chose between the two potential functions as they both appear to work equally well, however the Lennard-Jones potential has two distinct advantages over the Buckingham potential since a) it is quicker to compute a power than an exponential function and b) the Buckingham curve inverts at small distances and atoms which are too close together collapse into a common nucleus.

The function used for this work is a modified form of the Buckingham potential:-

$$V_{nb} = E_{min} \left((-2 \, (R_{min})^6 / r^6) + e^{(12(1-(r / R_{min})))} \right)$$

However it is hoped that the Lennard-Jones function will be used for future work.

Section 2.6

Out-of-Plane Bending

The minimum energy position for a trigonal planar atom (olefinic carbon, carbonyl carbon) is where the central atom and its three substituents are all coplanar. Deformations within the plane are taken care of by the angle bending term (section 2.3) as they work against different restoring forces to out-of-plane deformations which distort the pi electron system's energy.[9]

Deviations from planarity can be measured by calculating the improper torsion angle for the molecule (see Fig. 2.4).

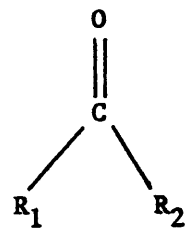


Fig. 2.4

If the atoms are coplanar when we look along C=O the angle between R_1 and R_2 should be 180°. If there is any out-of-plane bending the improper torsion angle will be less than 180° and the following expression is used to calculate the out-of-plane bending energy:-

$$V_x = (1/2)\, k_x\, (180 - X)^2$$

where k_x is the out of plane bending force constant and X is the calculated value of the improper torsion angle.

Section 2.7

Coulombic Interactions

Since Rutherford first proposed his model of the atom in 1909 there have been many theories as to the nature of the chemical bond from which have arisen the principles of Quantum Mechanics and Molecular Orbital Theory.

It is now generally accepted that when a bond is formed between two atoms the electronegativity of each atom is the factor which determines the nature of the bond which has been formed. The electronegativity of an atom is a function of the charge on the nucleus and the size of the atomic radius and can be taken as a measure of the atom's ability to attract free electrons. The larger the charge on the nucleus and the smaller the atomic radius the larger will be the atom's electronegativity and the more readily will the atom be able to attract electrons.

Differences in the electronegativities of the atoms forming a bond give rise to various degrees of bonding of which there are two extremes, pure covalent bonds and ionic bonds.

A pure covalent bond is formed between two identical atoms. In this type of bond both atoms will attract the bonding electrons equally well, so the electrons will be shared equally between the atoms. Good examples of this type of bonding are homonuclear diatomic gases such as N_2, O_2 and H_2.

In ionic bonding one atom has a much greater electron affinity than the other and there is a complete transfer of one electron from one atom to the other. Ionic bonding occurs in salts such as $Na^+ Cl^-$.

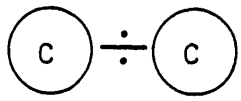
The bonding which occurs in organic molecules is generally somewhere between the two extremes described above. The atoms forming bonds in a molecule tend to be similar, but not identical. The electron density of the bond tends to be concentrated towards the atom which is able to exert the stronger attraction resulting in the bond having polarity (see Fig. 2.5). The atoms in a molecule tend to have partial charges. Interactions between these partial charges give rise to the coulombic component of the steric energy and the equation used to calculate this interaction is derived from Coulomb's Law:-

$$V_q = (332\ q_i\ q_j\ /\ D\ r_{ij})$$

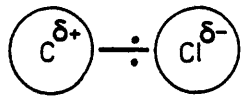
where q_i and q_j are the partial charges for atoms i and j and r_{ij} is the interatomic separation between i and j. The conversion factor 332 is included to give the coulombic energy, V_q, units of kcal mol^{-1} so that it is compatible with the other energy terms. D is a scaling factor used to account for the dielectric permittivity of the medium separating i and j, in vacuo the dielectric constant is used and this has a value of one, in bulk solution in water a value of D = 80 is used and to represent a microenvironment dielectric permittivity D = 4 is used as this has been shown to reproduce certain properties in enzyme active sites and gives good structures for amino-acid crystals. A value of four was used throughout this work.

When calculating the coulombic energy the molecule is split into charge groups (eg. amide groups, carboxylate groups, protonated nitrogens etc.) and intra group interactions are ignored. The partial charges for such calculations can be obtained from CNDO / CNDO2 type

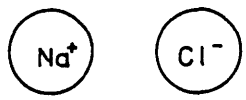
Bond Catagories



a) A pure covalent bond - electron pair are shared equally by the two atoms.



b) A polar covalent bond - electron pair are shared unequally by the two atoms giving rise to a dipole.



c) An ionic bond - electron transfer has occurred giving rise to charged atoms (ions).

Fig. 2.5.

calculations.

Section 2.8

Cross Terms

Many different force fields have been developed by individual groups and there are no strict rules governing how many or what types of potential energy functions should be used. Several existing force fields employ cross terms eg.:-

Bond-Bond terms

$$V_{ll} = k_{ll} (l - l_o) (l' - l_o')$$

Bond-Angle terms

$$V_{l\theta} = k_{l\theta} (l - l_o) (\theta - \theta_o)$$

Torsion-Angle terms

$$V_{w\theta} = k_{w\theta} (w - w_o) (\theta - \theta_o)$$

Angle-Angle-Torsion terms

$$V_{\theta\theta w} = k_{\theta\theta w} (\theta - \theta_o) (\theta' - \theta_o') (w - w_o)$$

In general these cross terms have been introduced to overcome specific weaknesses in existing force fields especially when dealing with strained structures. However, it has been found that cross terms

play a larger role in the calculation of vibrational frequencies than for general molecular structure and they have not been used for this work.

Section 2.9

The Force Field Method

Force Fields and Vibrational Analysis

The force field method originated in vibrational spectroscopy where it was developed in an attempt to predict vibrational frequencies from a simple mathematical model.

The principles of the force field method were derived from the assumption that a molecule with N atoms (i.e. 3N coordinates) has a strain free geometry, X_o, at which its potential energy is zero. Deviations from this strain free state result in an increase in potential energy, V_s, which can be expressed in terms of a Taylor expansion series:-

$$V_s = V_o + \sum_{i=1}^{3n} \left(\frac{\partial V}{\partial x_i}\right)_0 \Delta x_i + \frac{1}{2} \sum_{i,j=1}^{3n} \left(\frac{\partial^2 V}{\partial x_i \partial x_j}\right)_0 \Delta x_i \Delta x_j + \frac{1}{6} \sum_{i,j,k=1}^{3n} \left(\frac{\partial^3 V}{\partial x_i \partial x_j \partial x_k}\right)_0 \Delta x_i \Delta x_j \Delta x_k + \text{higher terms}$$

where x_i, x_j and x_k represent small displacements in atomic coordinates.

V_o is by definition, zero at the potential energy minimum where the first derivatives disappear from the above equation. A harmonic approximation can be obtained by truncating to the quadratic terms.

The second derivatives are expressed as force constants giving:-

$$V_s = (1/2) \sum_{i,j=1}^{3n} f_{ij} \Delta x_i \Delta x_j$$

Vibrational frequencies can then be calculated from the force constants using the secular determinant derived from Newton's Laws of motion.

The first step in developing a force field is the selection of a set of potential functions with which certain molecular properties can be described. The parameters for such functions must then be chosen from reliable existing experimental data such that they reproduce observable properties for a group of test compounds, with the ultimate goal that the force field can be applied to additional compounds for which there are no experimental data.

The first force field formulated along these lines was the Central Force Field, in which the potential functions were defined in terms of interatomic distances. This method gave good results for ionic compounds, but was of very little use for organic compounds and is used very rarely today.

The two types of force fields commonly used in vibrational spectroscopy today are the Valence Force Field[10] and the Urey-Bradley Force Field.[11] With both force fields the potential functions are derived from classical mechanics eg. Hooke's Law and are formulated for the molecule's internal coordinates i.e. bonds, valence angles and torsion angles. The main difference between the two force fields is that the Urey-Bradley force field also includes 1,3 non-bonded terms. Although, in theory, each molecule should have its own force field,

when the Valence force field was extended to include cross terms it was found to be transferable within a group of molecules with reasonably accurate results.

Force Fields and Molecular Mechanics

Once the force field method had become established in vibrational spectroscopy its potential application to other areas became noticed. Various force field methods have now been developed for molecular mechanics calculations by individual groups using a range of different potential functions and parameters. Several of the force fields currently in use appear to be converging and it is possible that a single force field will eventually be attained capable of accurately reproducing molecular properties.

Although the potential functions may differ, most molecular mechanics calculations employ modified versions of the Valence or Urey-Bradley type force fields. A Valence force field was used for this work which included non-bonded interactions for atoms which were not bonded to each other, or to a common atom, unlike the Urey-Bradley force field which does include geminal interactions. The Valence force field used also included an out-of-plane bending term and a coulombic term, but omitted any cross terms which were felt to be more relevant for the calculation of vibrational frequencies.

The parameterisation of a force field relies heavily on existing experimental data from several sources. Vibrational analysis can provide the force constants for bond length deformations, valence angle distortions and out-of-plane bending. Diffraction experiments can provide equilibrium bond lengths and angles as can microwave spectroscopy which also provides torsional barriers. It is important

that the initial parameterisation should involve a range of structures from aliphatic compounds to highly strained systems if the force field is to be able to cope with unusual steric constraints. The parameters must be carefully chosen using reliable experimental data since the force field can only be as accurate as the parameters it contains.

Once a force field has been designed the parameters and potential functions will need to be refined to give better agreement with observed structural properties. There are two methods of refinement - a trial and error method and a least squares procedure.[12]

With the trial and error method the force field parameters are varied logically by the user to give more accurate results. Whereas the least squares procedure, which was designed by Lifson and Warshel,[12] involves comparing the results given by a particular parameter set with experimentally obtained data. Any differences between the observed and calculated properties are noted and the sums of the squares of these differences are minimised by iteratively varying the potential constants. Neither approach seems to offer a significant advantage over the other since they both appear to work equally well.[10]

Force Fields and Thermodynamics

So far we have only discussed force fields in their roles in vibrational spectroscopy and molecular mechanics giving vibrational frequencies, molecular geometries and their relative conformational energies. However force fields have also been parameterised to give heats of formation, crystal packing arrangements and transition state structures and reactivities.

From molecular mechanics we get the hypothetical potential energy

associated with a motionless molecule. However, the molecule also has rotational, vibrational and translational energy. To calculate the heat of formation, or enthalpy of the molecule we must add the energies arising from such motions to the potential energy.

The enthalpy H_{RT} associated with full excitation of rotational and translational degrees of freedom is given by:-

$$H_{RT} = 3RT$$

where R is the gas constant and T is the absolute temperature in K.

Vibrational modes are only partially excited and the enthalpy associated with vibration H_v is given by:-

$$H_v = RT \sum_{i=1}^{3n-6} u_i / ((e^{u}i) - 1)$$

where $u_i = h\nu_i/kT$, k = Boltzmann's constant, h = Planck's constant, N = the number of atoms in the molecule and the ν_i are the fundamental vibrational frequencies.

A correction must also be made to include the molecular zero point vibrational energy E_v^o which can be calculated using the function:-

$$E_v^o = N^o (1/2) h \sum_{i=1}^{3n-6} \nu_i$$

where N^o is Avogadro's number.

Once the molecule's heat of formation has been calculated the calculation of its other thermodynamic properties is fairly straightforward.

The current approach to force field development is to design a force field which will give the best overall agreement between calculated and observed geometric, vibrational and thermodynamic properties even if some accuracy is lost in any one area. This is known as the Consistent Force Field method.[13]

Chapter Three

Conformation Generation and Energy Minimisation

Section 3.1.

Introduction

A molecule's steric energy is dependent upon the positions of its atoms, i.e. its conformation therefore before we begin any molecular mechanics calculations we must first ascertain its three dimensional structure.

For any molecule containing N atoms the relationship between the steric energy and the atomic positions can be represented by a 3N+1 - dimensional energy hypersurface, from which can be found all the minimum positions for the molecule. The lowest energy minimum is known as the Global Minimum Energy Conformer (GMEC).

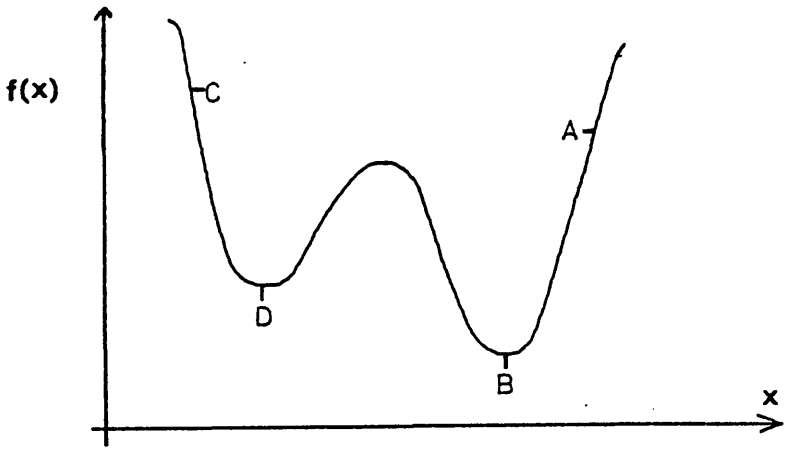
For a simple molecule there will probably be a few well defined minima, however for a large molecule the energy hypersurface may contain several hundred (or even thousand) minima.

The energy hypersurface can be represented by a function f(**x**) where **x** is the vector representing the position of the molecule on the hypersurface and the variables of **x** are the atomic coordinates. The ideal method of locating all the minima on the hypersurface would be to use a Grid Search procedure,[1] which involves evaluating the function f(**x**) at a series of points covering the molecule's conformational space and examining the function values to find the low energy positions.

This method is not practicable to use for a large molecule due to the complexity of the functions and the large number of variables which would have to be computed. It can however be applied where only

one torsion angle in a molecule is being varied and is commonly used to study the backbone torsion angles in amino acids and peptides.[2,3]

The most common method used to locate the minima involves a two stage process of conformation selection and energy minimisation. Two stages are required because trial structures from, for example, molecular modeling very rarely correspond to conformers which exist at the minima of a hypersurface.



From the above diagram positions A and C on the hypersurface are likely to correspond to trial structures whose energies must be minimised to give structures which correspond to the minima at B and D respectively.

Locating the GMEC can prove very difficult for large molecules which have several minima. If B corresponds to the GMEC its location depends on starting from a conformation which is close to it eg. the conformation which exists at A. Whereas minimisation of the conformation which exists at C will lead to the conformer existing at the local minimum position D. No minimisation procedure exists which can take us from position C to the GMEC at B.

The minima reached using the two stage process will be those

minima closest to the starting conformations and it is highly likely that low energy conformers found using this process will correspond to local minima like D. The selection of starting conformations is therefore a very important process.

Section 3.2.

Trial Conformation Selection

Although the energy hypersurface for the molecule is defined in terms of atomic coordinates for conformation selection, it is much easier to use the variables involved in the molecular mechanics potential functions i.e. the molecule's internal coordinates. However, bond lengths and angles are represented by relatively 'hard' functions, so, on the assumption that they will adopt positions close to their optimal values, the main variable determining molecular conformation will be the value of the torsion angles.

The torsion angle values within any molecule will be determined by the types of bonds involved. It has long been understood that rotation could occur about single bonds. Originally it was thought that such rotation was free from any constraints, but by the 1940's it had been demonstrated that there were energy barriers involved giving rise to restricted rotation.

Restricted rotation can be easily demonstrated with a simple molecule such as ethane.

The barrier to rotation about the carbon - carbon bond is best described in terms of the changes in energy which arise as the torsion angle H1 - C1 - C2 - H4 is rotated through 360° (see Fig 3.1).

As the torsion angle is rotated from 0° - 360° three minima and three maxima are encountered. The energy maxima occur where the hydrogens of C1 are eclipsing the hydrogens of C2 i.e. where there is the shortest possible distance between the two sets of hydrogens. The energy minima occur where the hydrogens of C1 are in a staggered relationship with the hydrogens of C2 i.e. where there is the greatest seperation between the two sets of hydrogens.

The torsional barrier to rotation about the carbon - carbon bond is approximately 3 kcal mol^{-1}. This energy is readily available at room temperature and the molecule will undergo rapid rotation passing between the two extremes of maxima and minima and giving an infinite number of conformations.

How then are the conformations which are to be minimised selected? The answer is very simple. Molecular mechanics deals with molecules at absolute zero i.e. when they are motionless, so the conformations which are selected should be those which accurately represent the molecule's structure at low temperatures.

In ethane the barrier to rotation is small, therefore at room temperature interconversion between the different minima is rapid and there is no way of measuring how long the molecule spends in each of its conformations. However, lowering the temperature of the surroundings slows down rotation and several low temperature NMR studies have shown that when a molecule has a rotatable bond it will tend to spend most of its time in the low energy conformers corresponding to the minimum energy positions for the torsion angles.

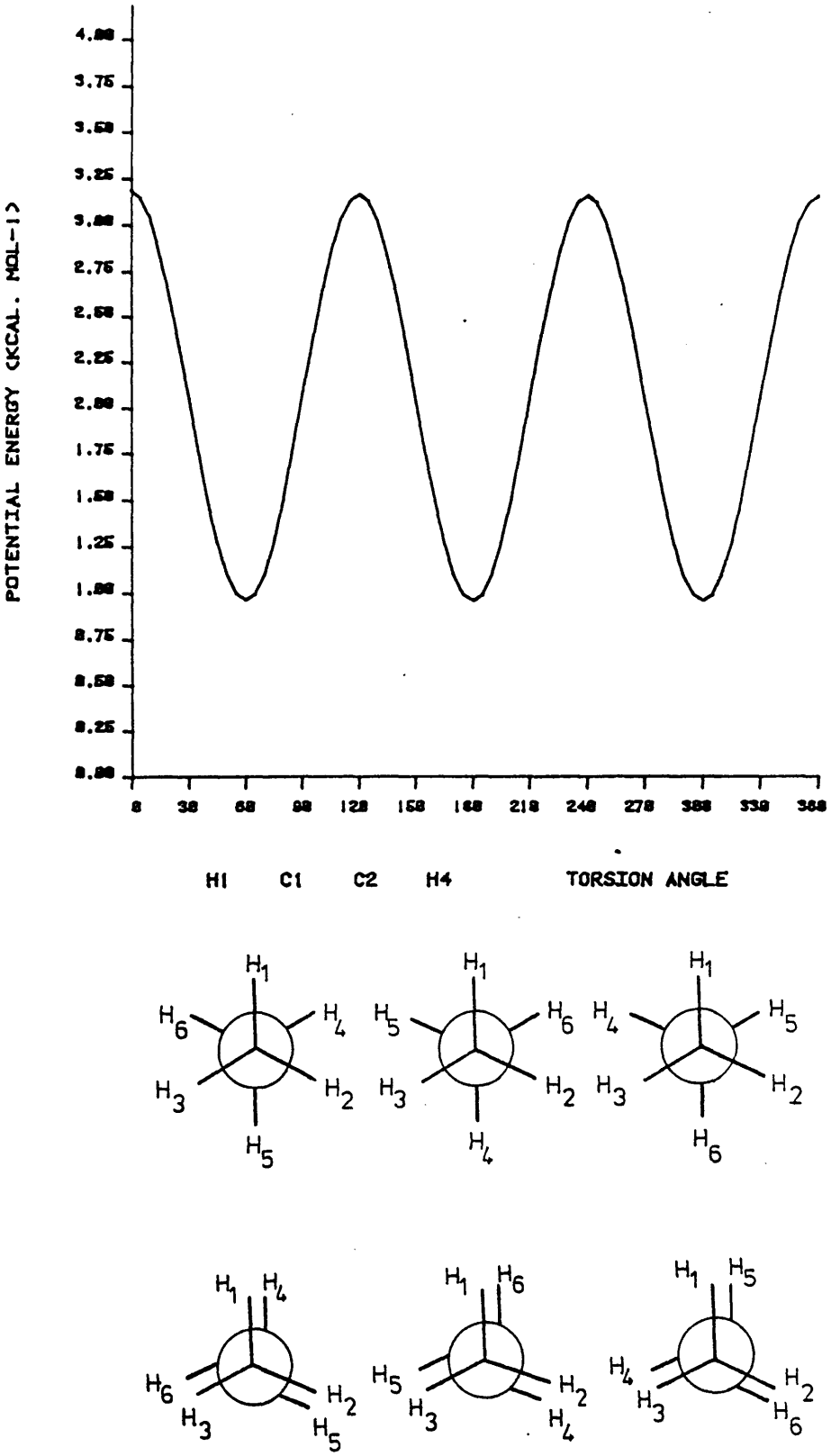
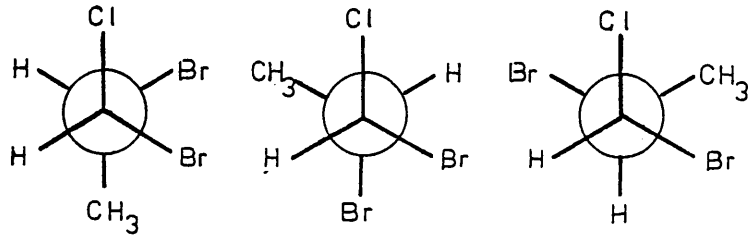

POTENTIAL ENERGY (KCAL. MOL-1)
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
H1
C1
C2
H4
TORSION ANGLE
H1
H6
H4
H3
H2
H5
H1
H5
H6
H3
H2
H4
H1
H4
H5
H3
H2
H6
H1
H4
H6
H3
H2
H5
H1
H6
H5
H3
H2
H4
H1
H5
H4
H3
H2
H6

Fig. 3.1.

It is these conformations which must be identified for molecular mechanics studies. For ethane this is a relatively simple task since it has only one rotatable bond giving rise to three minima, which due to the symmetry of the molecule are identical stereoisomers. This means there is only one low energy conformation to study and since it is the lowest possible energy conformation it will be the GMEC.

Finding the GMEC for other molecules is not nearly so easy. If, for example, there is asymmetry across the rotatable bond eg. 1-chloro-1-bromo-2-bromo-ethane there will be three distinct low energy conformers which will be populated according to their relative energies, the GMEC being the most highly populated state.



In general most molecules have more than one rotatable bond and such bonds are generally asymmetric. This makes it very difficult to locate low energy conformations. To gain some idea of the complexity of the task imagine that if a molecule has 5 rotatable bonds each having a possible 3 minima this would give 3^5 = 243 low energy conformers. If the molecule has 10 rotatable bonds there would be 3^{10} possible low energy conformations on which to perform molecular mechanics calculations - obviously a difficult if not impossible task.

Since most of the molecules studied in this project had upwards of 10 rotatable bonds it was necessary to define certain criteria

with which sensible starting conformations could be generated. Two distinct approaches were used to generate the trial structures studied and these methods will now be described in more detail.

The GLOMIN program

This program was originally designed to fold peptide chains to give cyclic compounds. The program was rewritten during the course of this project to run in the array processor and a program listing is given in Appendix A.

The function of the program is to set designated torsion angles to values which have been input by the user. If there are four torsion angles and the user inputs a set of torsion angle values eg. 60°, 180° and -60°, the program will generate a total of 3^4 conformations having every possible permutation of torsion angle values from 60° 60° 60° 60° to -60° -60° -60° -60°.

Before the program can be used the torsion angles and a set of torsion angle values must be chosen. Since most of the molecules studied in this project had several rotatable bonds, certain criteria were employed to restrict the number of torsion angles of interest. The following is a worked example for Phe-[N]-L-Phe-pAB (see Fig.3.2) and the method was employed for several other compounds.

It can be seen from Fig. 3.2 that the conformation of Phe-[N]-L-Phe-pAB can be described by 11 torsion angles namely:-

1) C37 C39 C40 C2 2) C39 C40 C2 N6 3) C40 C2 N6 C13

4) C2 N6 C13 C9 5) N6 C13 C9 N10 6) C13 C9 N10 C51

7) C9 N10 C51 C52 6) C54 C46 C43 O44 9) C40 C2 C1 O3

10) N6 C13 C26 C25 11) C13 C26 C25 C23

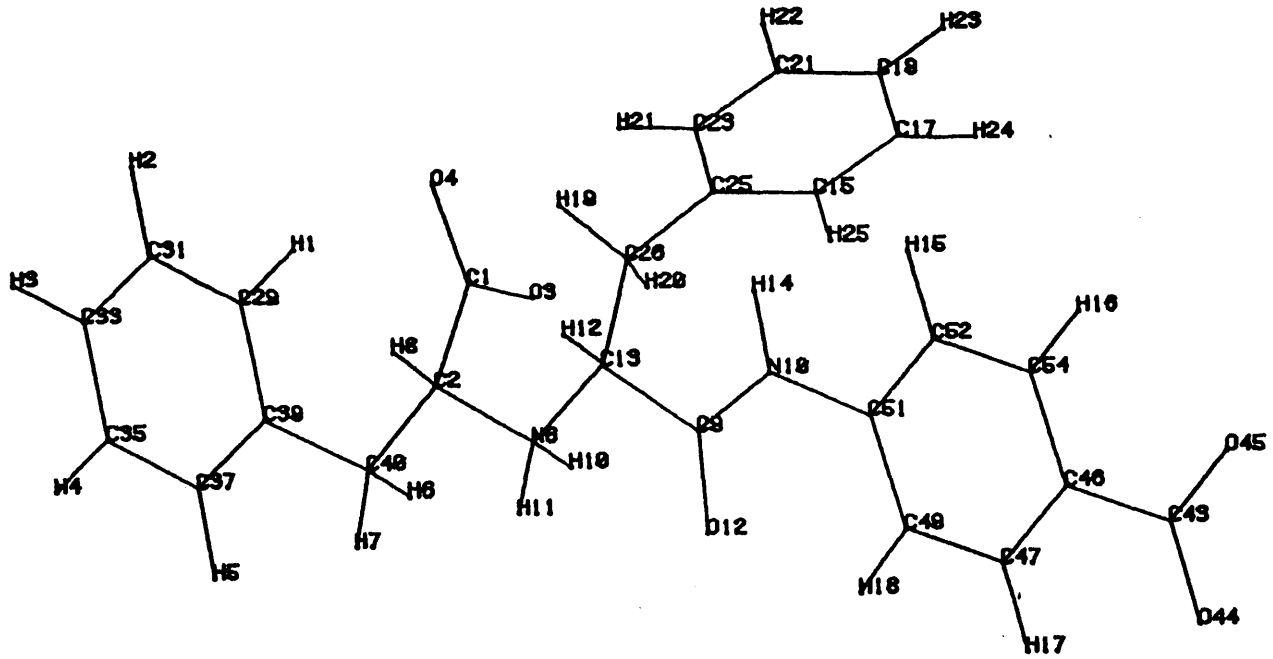

H22
H23
C21
O18
H21
O23
C17
H24
H2
O4
C25
O15
H19
H25
C31
H1
O26
H15
H3
C1
O3
H20
H14
O29
H16
C33
H12
C52
H8
C13
C54
C2
N19
C39
C51
C35
N6
O9
O45
H10
C40
C46
H4
O37
H6
H11
O43
O12
C49
H7
O47
H5
H18
O44
H17

Fig. 3.2.

To restrict the number of torsion angles we must locate those torsion angles which by their nature can be discounted, or set to favourable positions. Initially all the torsion angles are set to 180°. From Fig 3.3 it can be seen that torsion angle 6, an amide bond has a well defined minimumat 180° so it can be left at 180°. From Fig 3.4 torsion angle 7 has a minimumat 160° and the carboxylate group defined by torsion angle 8 will tend to lie approximately planar with the phenyl ring so it can be set at 180°. Fig 3.5 shows that torsion angle 1 has a minimum value at approximately 90° (phenyl rings tend to have minima between 70° and 110°) and finally from Fig. 3.6 we see that torsion angle 9 has a minimum position at 60° (which again tends to be favourable for a carboxylate group in the middle of a chain whereas torsion angle values of 120° tend to be more favourable for terminal carboxylate groups).

Setting all the above torsion angles to their favourable positions leaves 6 Torsion angles which is still too many for this method. The next criterion used is to separate the back bone torsion angles from the side chain torsion angles. From Figs. 3.7 and 3.8 it can be seen that the minimum positions for torsion angles 10 and 11 are at 170° and 70° respectively so they are set to these values.

All but four torsion angles have now been eliminated and Figs. 3.9 - 3.12 give the minima for these four:-

C39 C40 C2 N6	290° (-170°)
C40 C2 N6 C13	60° to 140°
C2 N6 C13 C9	240° to 290°
N6 C13 C9 N10	300° 120°

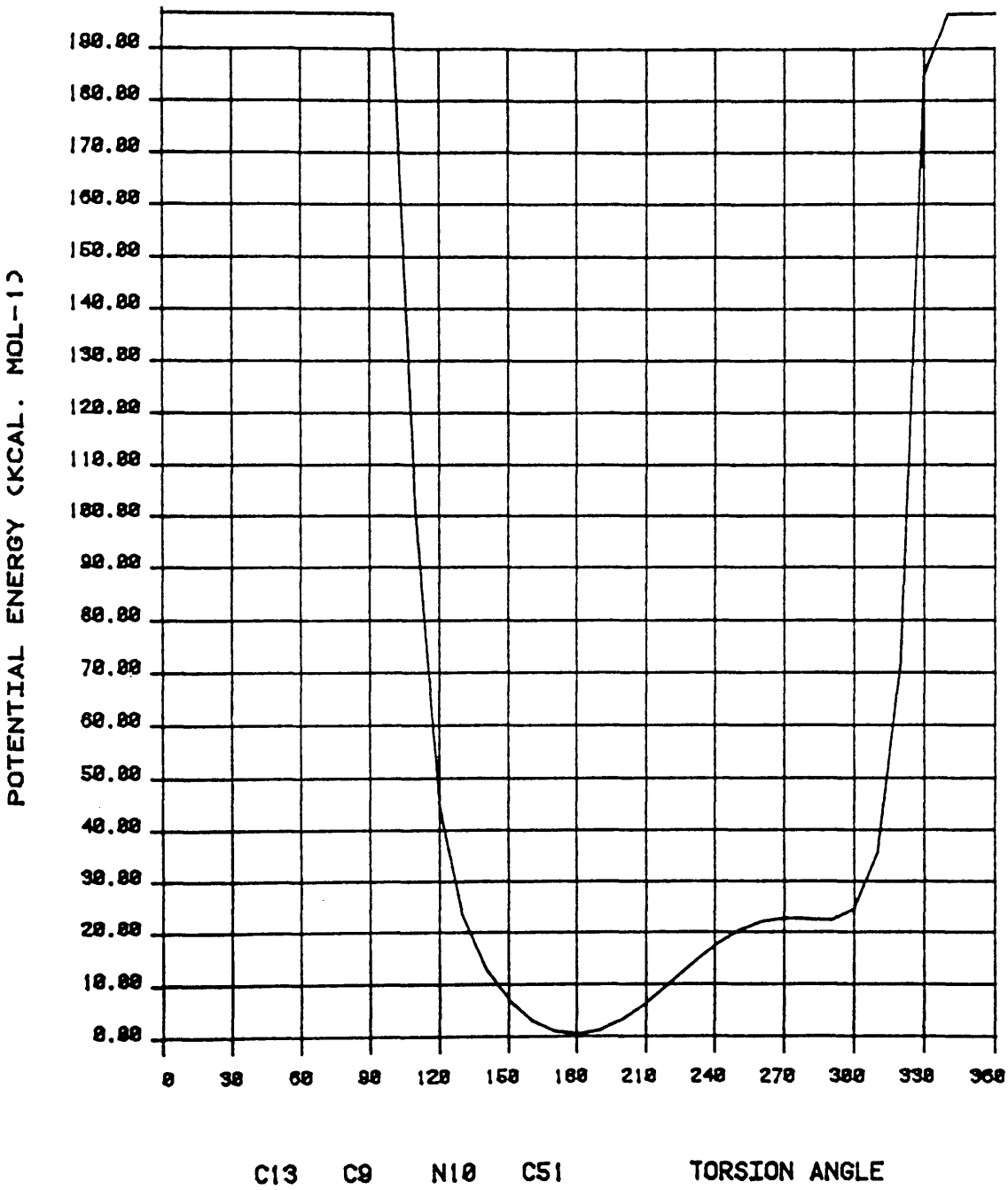

POTENTIAL ENERGY (KCAL. MOL-1)
190.00
180.00
170.00
160.00
150.00
140.00
130.00
120.00
110.00
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
C13 C9 N10 C51
TORSION ANGLE

Fig. 3.3.

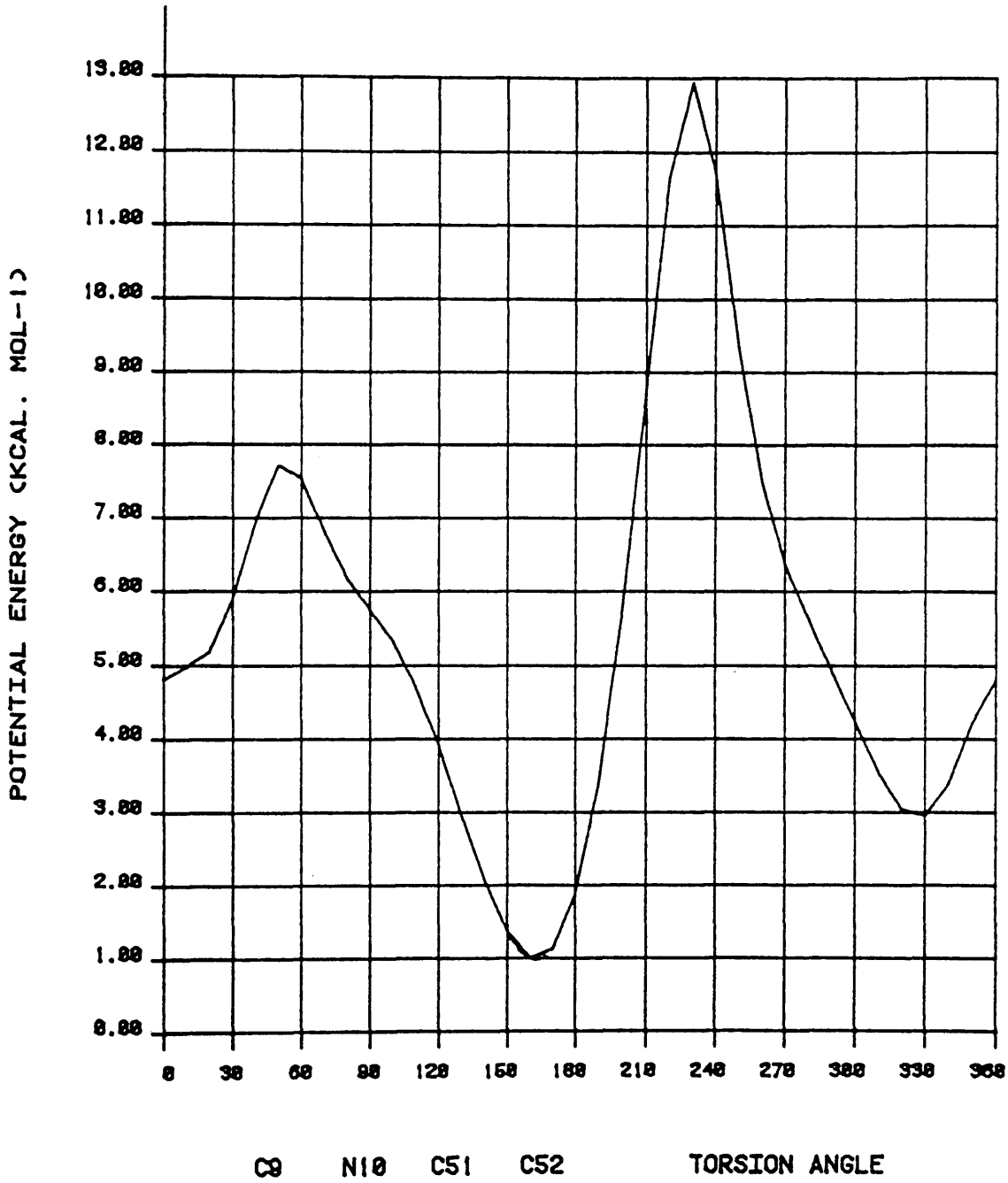

POTENTIAL ENERGY (KCAL. MOL-1)
13.00
12.00
11.00
10.00
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
C9 N10 C51 C52 TORSION ANGLE

Fig. 3.4.

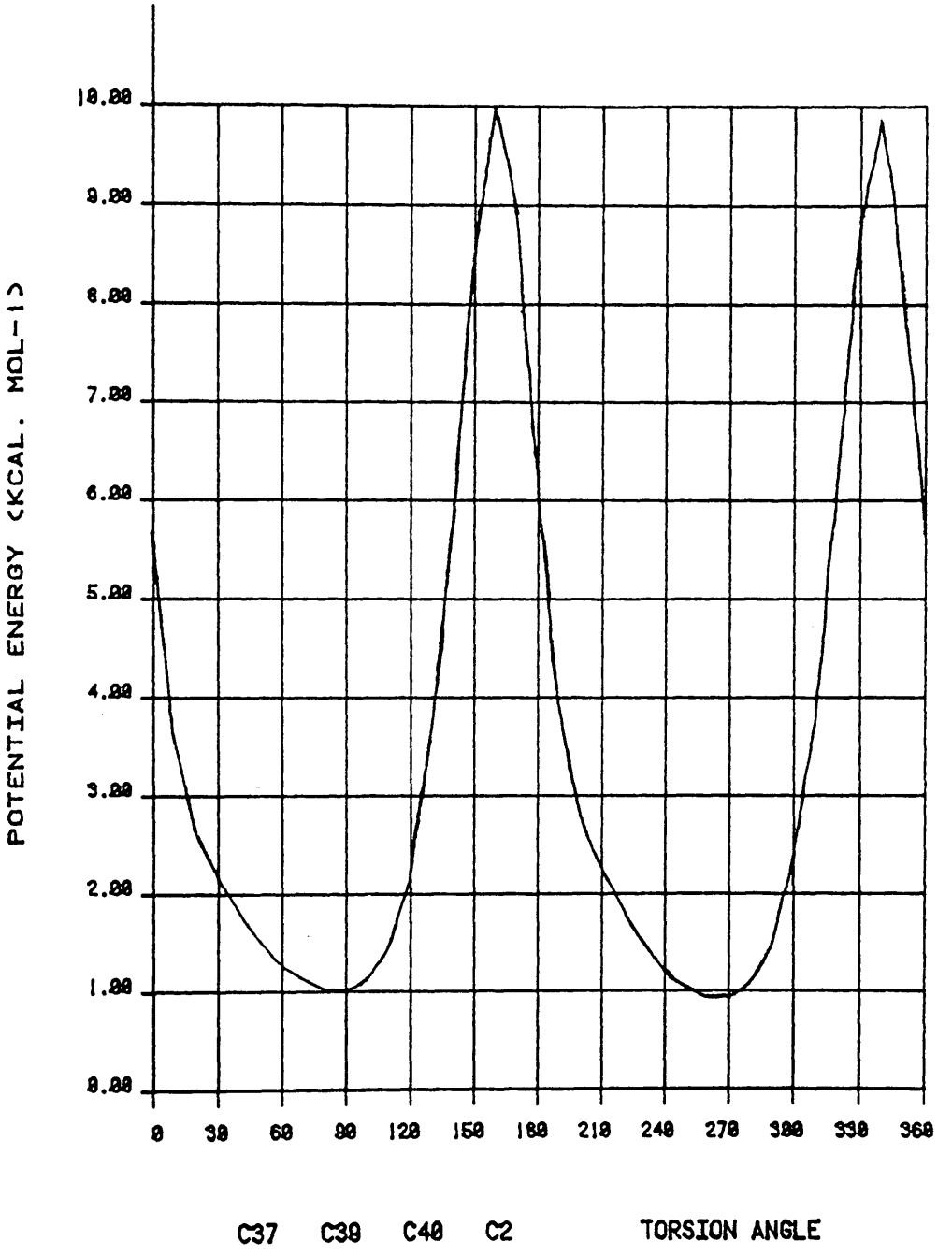

POTENTIAL ENERGY (KCAL. MOL-1)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
C37 C39 C40 C2
TORSION ANGLE

Fig. 3.5.

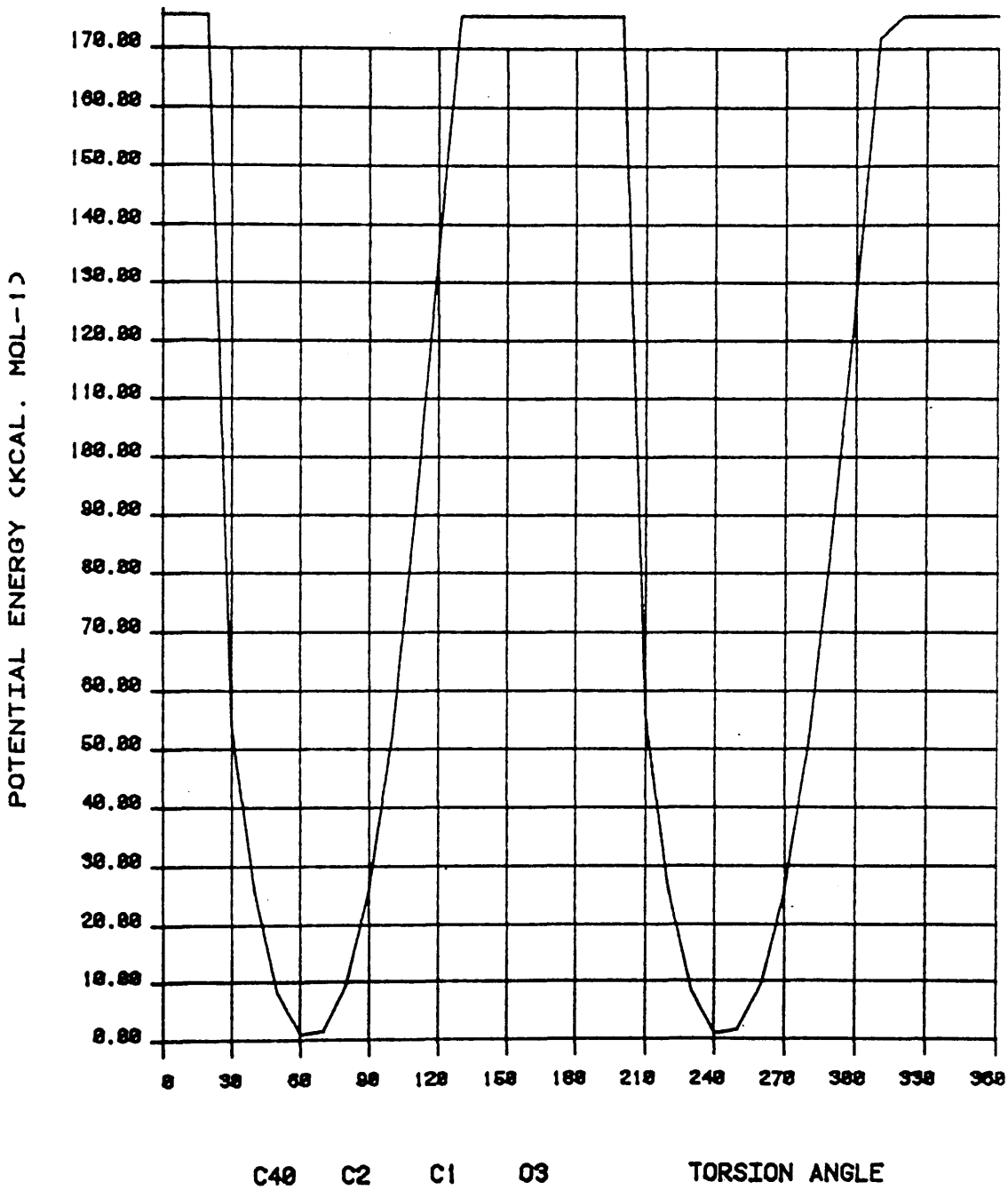

POTENTIAL ENERGY (KCAL. MOL-1)
170.00
160.00
150.00
140.00
130.00
120.00
110.00
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
C40 C2 C1 O3
TORSION ANGLE

Fig. 3.6.

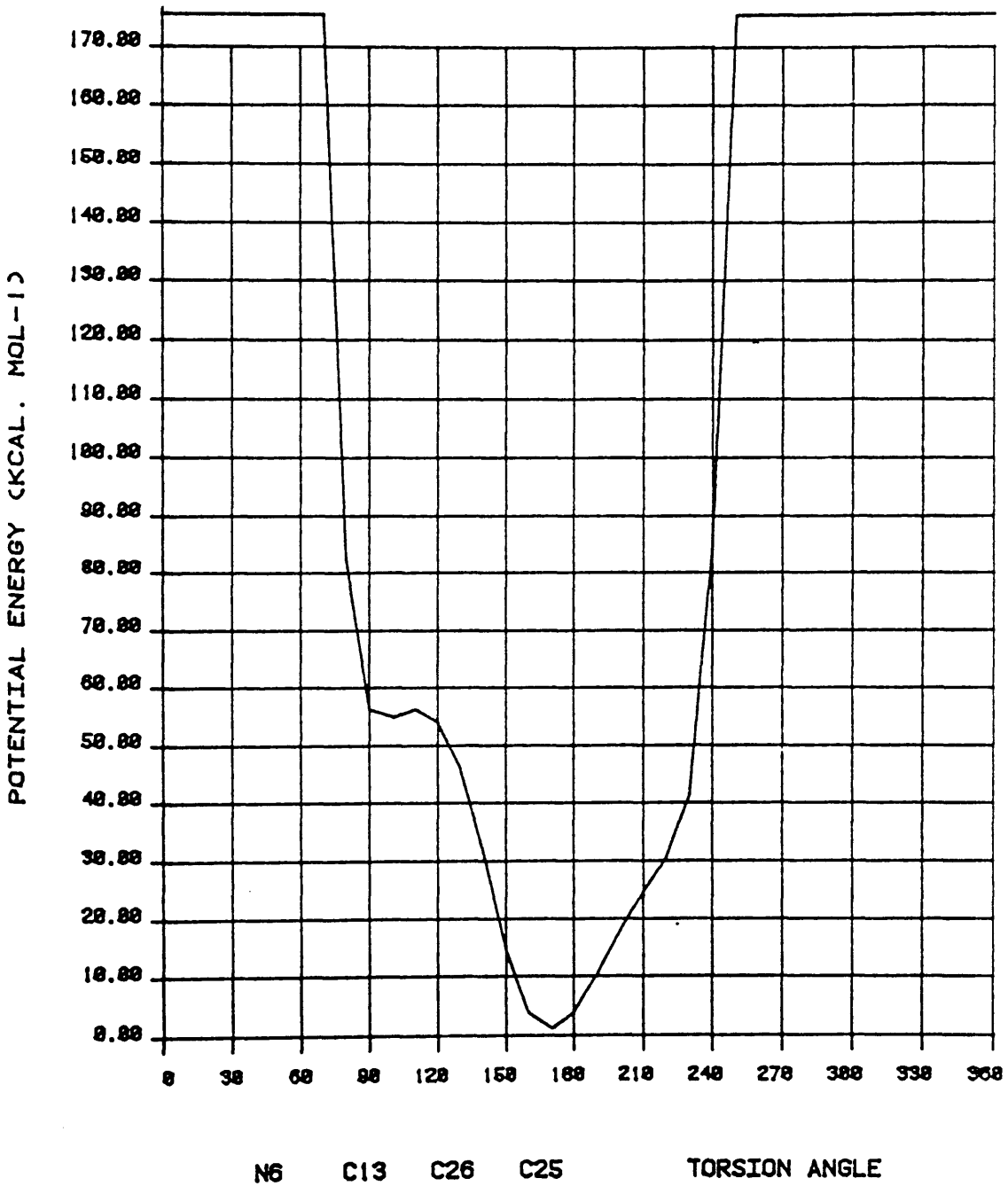

POTENTIAL ENERGY (KCAL. MOL-1)
170.00
160.00
150.00
140.00
130.00
120.00
110.00
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
N6 C13 C26 C25 TORSION ANGLE

Fig. 3.7.

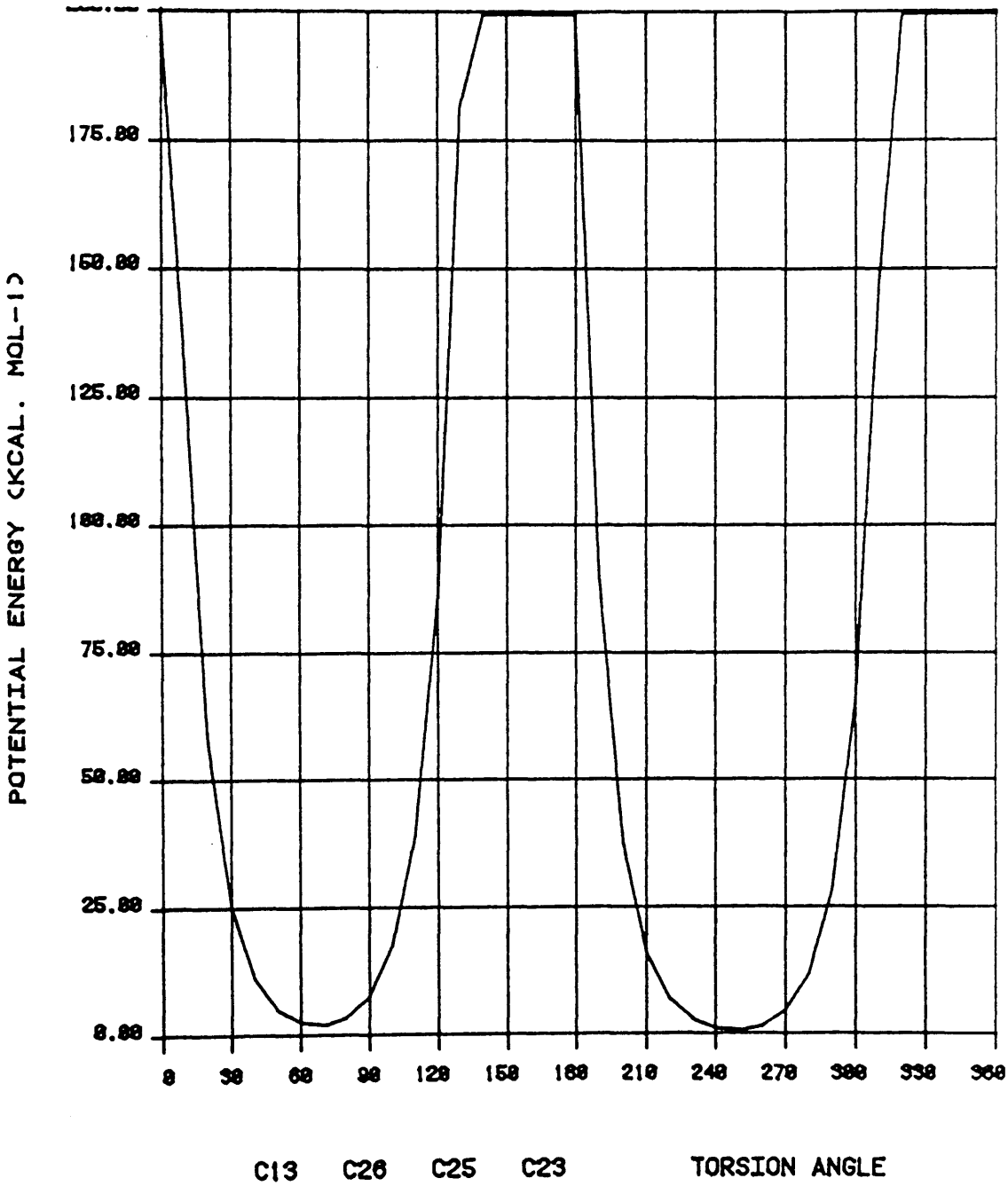

POTENTIAL ENERGY (KCAL. MOL-1)
175.00
150.00
125.00
100.00
75.00
50.00
25.00
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
C13 C26 C25 C23
TORSION ANGLE

Fig. 3.8.

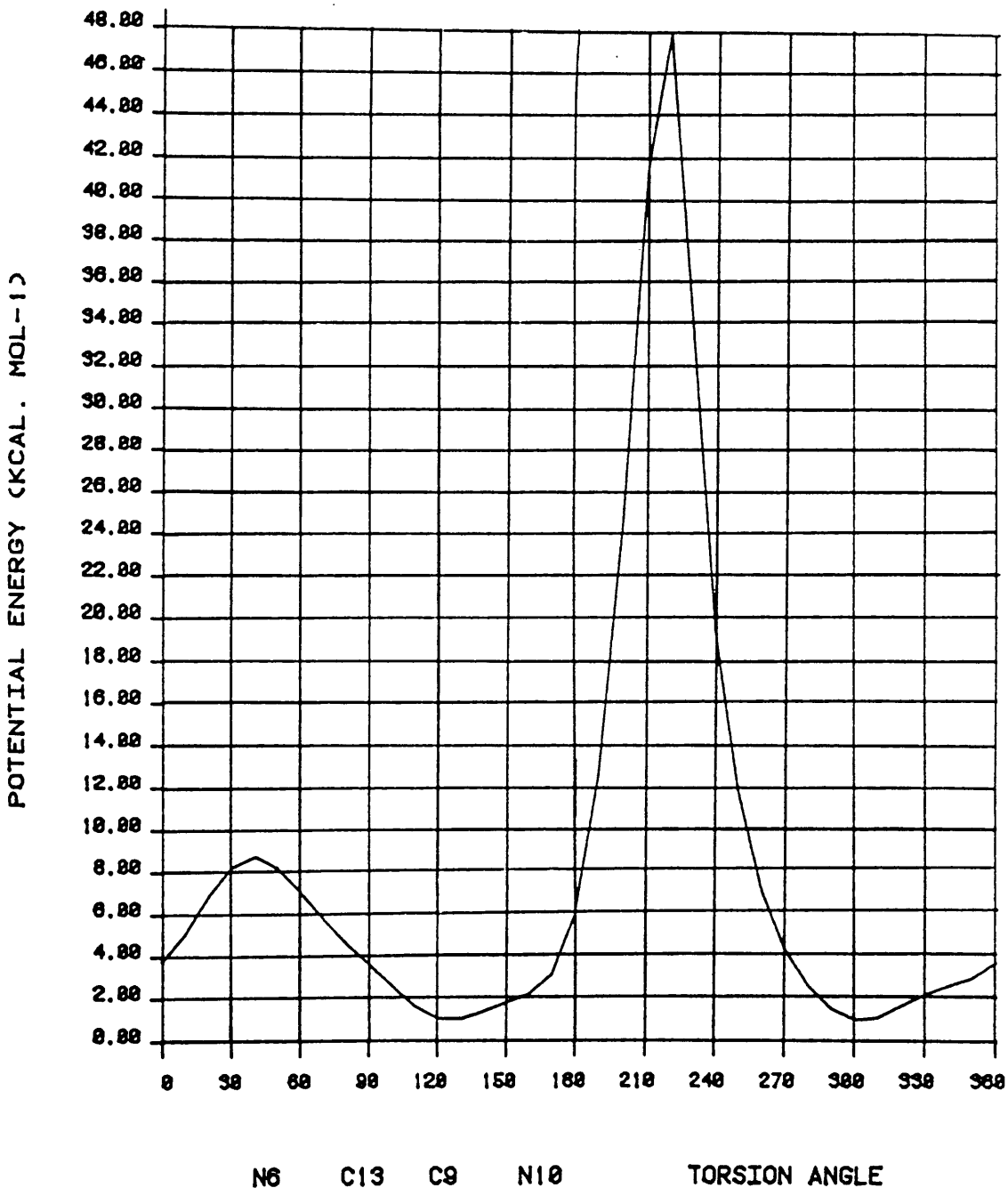

POTENTIAL ENERGY (KCAL. MOL-1)
48.00
46.00
44.00
42.00
40.00
38.00
36.00
34.00
32.00
30.00
28.00
26.00
24.00
22.00
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
N6 C13 C9 N10 TORSION ANGLE

Fig. 3.9.

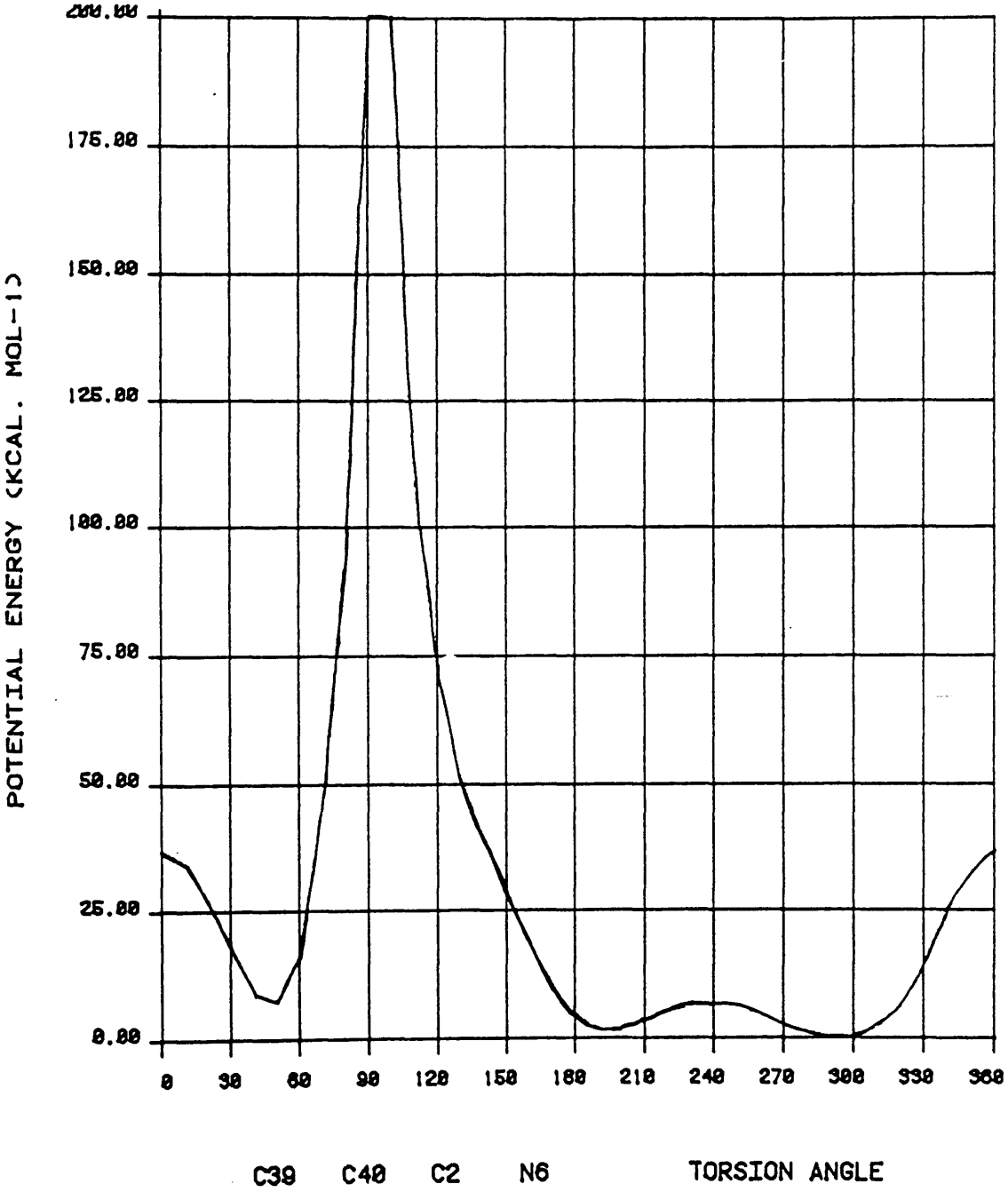

POTENTIAL ENERGY (KCAL. MOL-1)
200.00
175.00
150.00
125.00
100.00
75.00
50.00
25.00
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
C39 C40 C2 N6
TORSION ANGLE

Fig. 3.10.

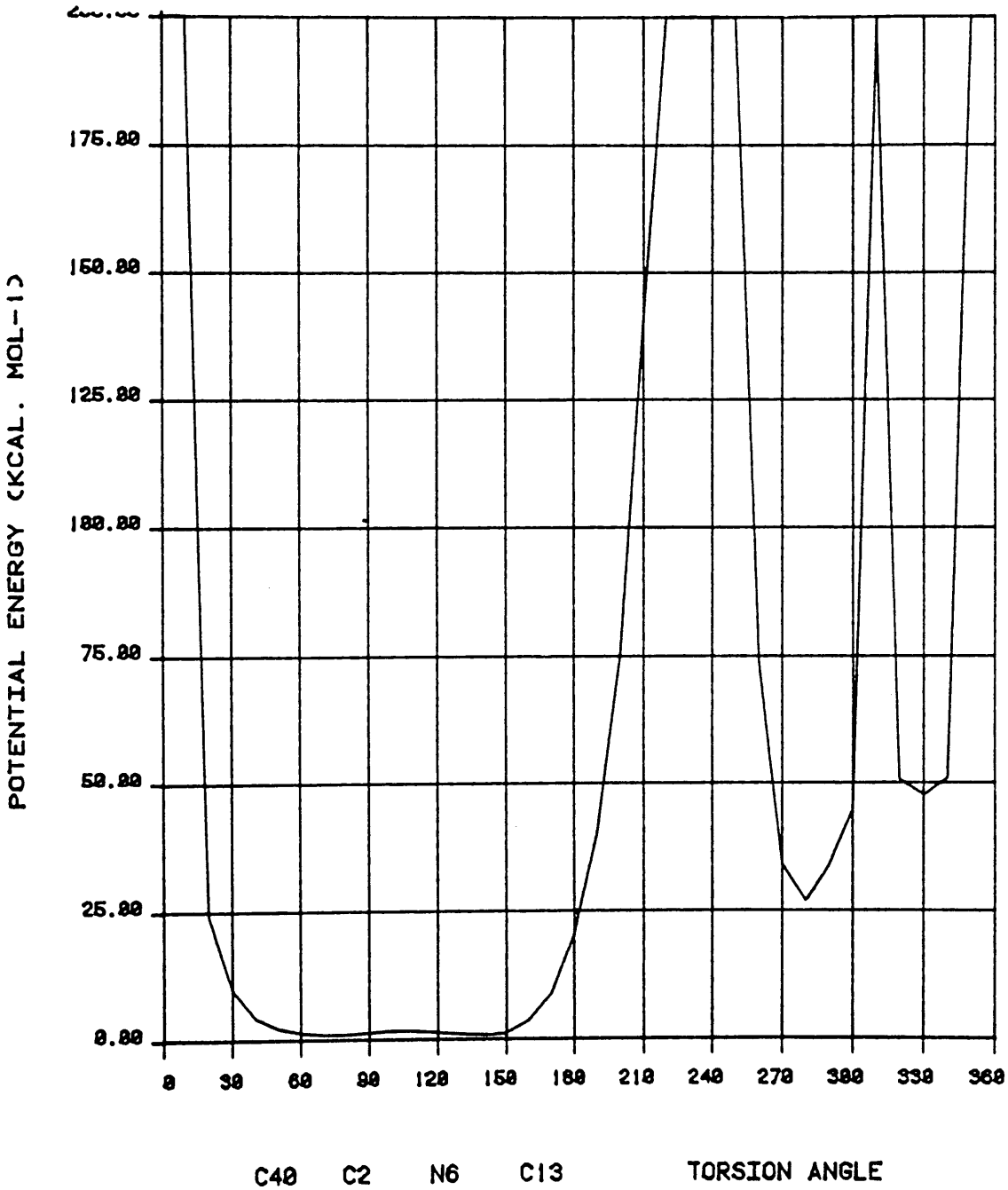

POTENTIAL ENERGY (KCAL. MOL-1)
175.00
150.00
125.00
100.00
75.00
50.00
25.00
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
C40 C2 N6 C13
TORSION ANGLE

Fig. 3.11.

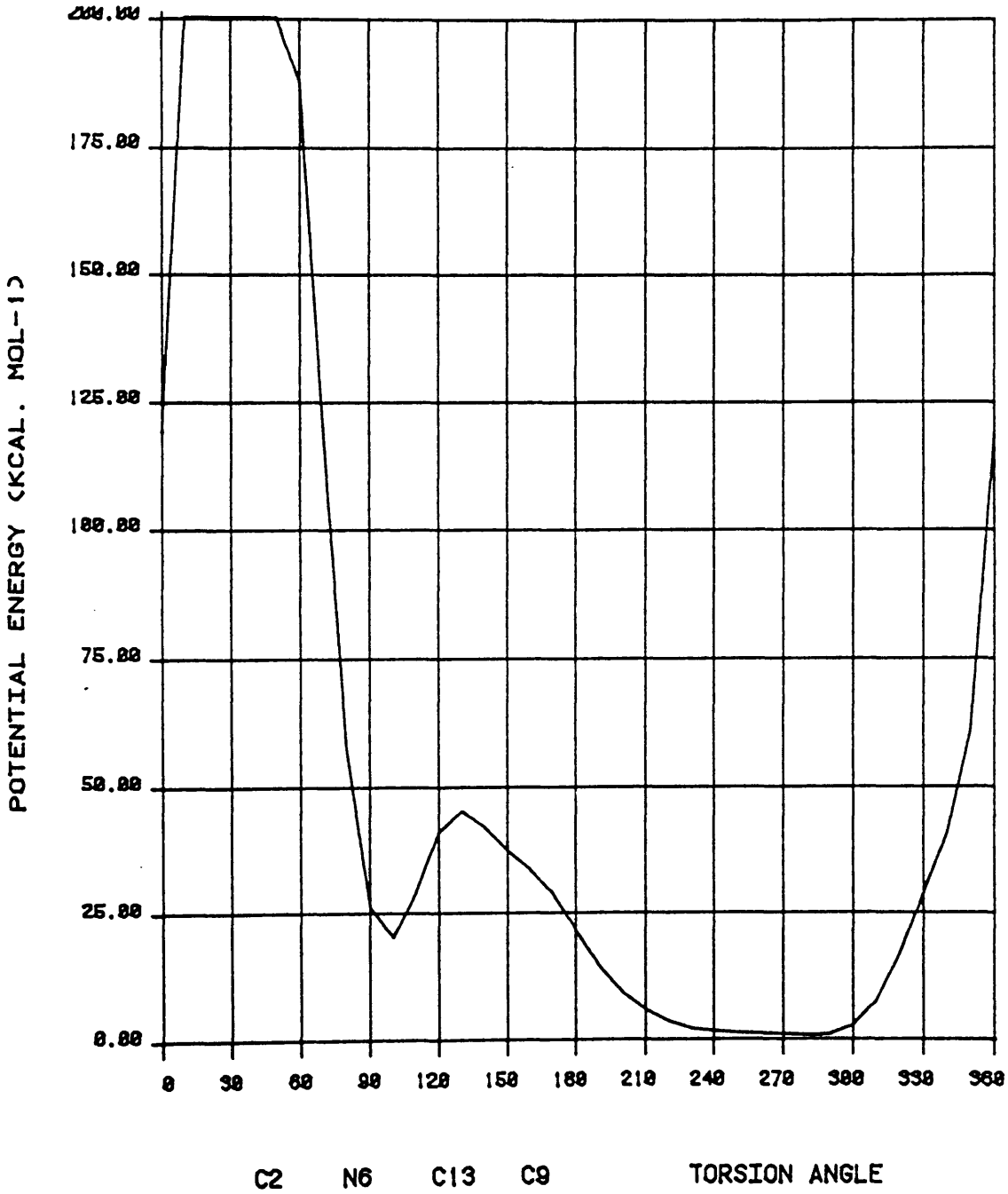

POTENTIAL ENERGY (KCAL. MOL-1)
175.00
150.00
125.00
100.00
75.00
50.00
25.00
0.00
0
30
60
90
120
150
180
210
240
270
300
330
360
C2 N6 C13 C9
TORSION ANGLE

Fig. 3.12.

from which four torsion angle values were chosen namely 60°, 120°, 180° and -70°.

The GLOMIN program accepts as input a single backbone chain (see Fig 3.13), the torsion angles to be altered, specified by four atom names and a set of torsion angle values, or generators. It will then alter the designated torsion angles, in this case 2, 3, 4 and 5, to give all possible permutations of the generators chosen.

As output the program produces a file containing all the conformations produced, in this case 4^4 = 256 conformations. Individual systems format files are then created from this output file. These systems format files are initially in the form of a single unbranched chain and the full molecule can be built using the MOL graphics system, (see Chapter 4) building on any groups or side chains into favourable positions.

The ESCAN program

This program is slightly more sophisticated than the GLOMIN routine. It generates conformations by progressively altering nominated torsion angles in the molecule stepwise until all possible conformations have been generated. As it generates each conformation it roughly calculates the energy, stores the low energy conformations and finally crudely minimises them using torsional parameters. One advantage of this method is that it generates the complete conformation with a crude energy and only saves those conformations which are below a specified energy value. The main disadvantage with the method is that although we would like to rotate each bond through 360° in say 10° steps - this is only feasible for molecules with up to four rotatable bonds, above this number it takes up too much computer

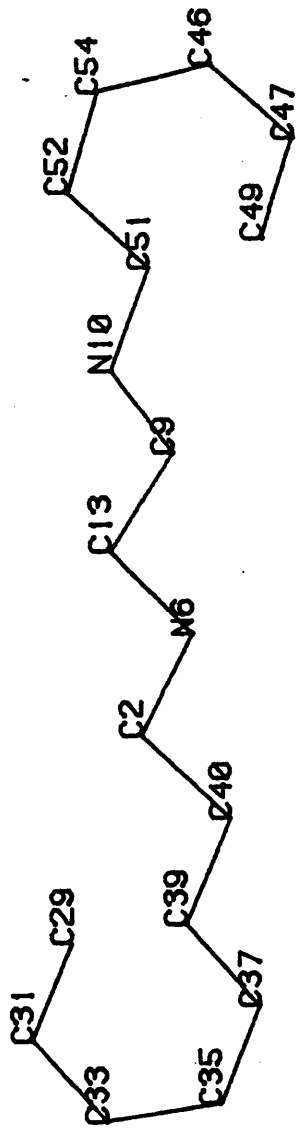

C54
C46
C52
C47
C49
C51
N10
C9
C13
N6
C2
C40
C39
C29
C31
C37
C33
C35

Fig. 3.13.

time. Once again we have to define criteria with which we can restrict the values of certain torsion angles. Once these torsion angles have been set the program gives the user the option of setting those bonds which are being rotated in a particular conformation, or using a random starting point generated by the computer.

Once again it is perhaps easier to describe the procedure with a worked example for one of the inhibitors of Thermolysin, the molecule β-phenylpropionyl-L-Phe (see Fig. 3.14).

The overall conformation of this molecule can be described by eight torsion angle values.

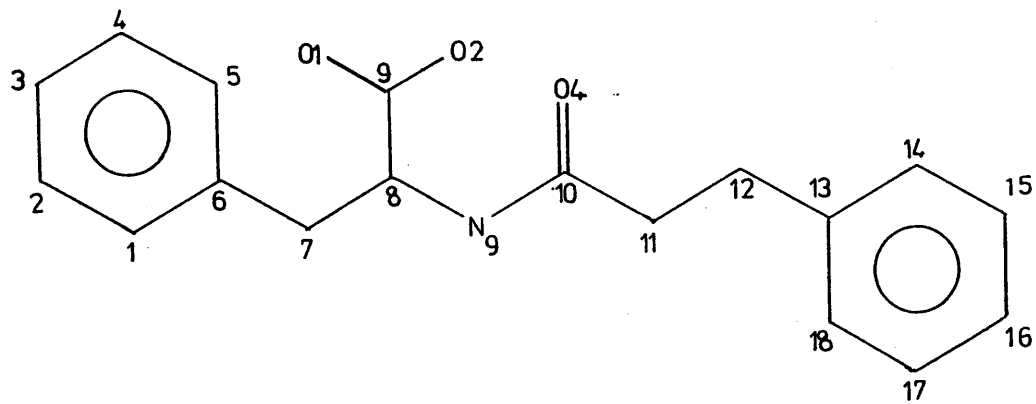
1) C1 C6 C7 C8 2) C6 C7 C8 N9 3) C7 C8 N9 C10

4) C8 N9 C10 C11 5) N9 C10 C11 C12 6) C10 C11 C12 C13

7) C11 C12 C13 C14 8) C7 C8 C9 O1

Once again we must restrict the number of torsion angles we are interested in. For this molecule the amide bond can be set to 180° leaving us with seven rotatable bonds. Ideally we would like to rotate all seven in 10° steps, however since this is not possible the best we can manage is to fix two torsion angles in favourable positions and allow the remaining five to be rotated through 360° in 30° steps.

β-PPP was run in ESCAN twice, on the first occasion torsion angles 7 and 8 were set at values of 90° and 120° since these were found to be favourable, the remaining five torsion angles were rotated through 360° in 30° steps from a randomly generated starting conformation, this gave seven conformations. On the second occasion torsion angles 7 and 8 were set at -110° and 80° respectively - these being the values found from the X-ray studies of the binding



β-Phenylpropionyl-L-Phe

Fig. 3.14.

conformation, again the five remaining bonds were rotated through 360° in 30° steps from a random starting point, this gave a further 13 conformations.

Obviously there are drawbacks with both GLOMIN and ESCAN - the main one being that restricting any torsion angle to a certain value will automatically influence the conformation of the rest of the molecule. This means that although several conformations have been generated there is no certainty that any of them can be minimised to give the Global Minimum Energy Conformer. In fact such are the limitations of both methods when dealing with large molecules that it is likely that several low energy conformations are overlooked. This can be partially overcome by the use of routines ALTER and RAMMAPfrom the MOL graphics system (see Chapter 4 sections 4.4.12 and 4.4.13). These routines can be used to generate more low energy conformations by improving upon the positions of various torsion angles within the molecule. The routines can act as a safety net to locate low energy conformations which would otherwise be ignored.

Section 3.3.

Energy Minimisation

Once we have generated our starting conformations we require to minimise the structures to give the conformers which exist at the minima of the molecule's energy hypersurface.

As previously discussed the energy hypersurface can be represented by a function f(x) the variables for which are the Cartesian coordinates of the atoms. Location of the minima depends upon optimisation of the function and since there are no constraints

imposed upon the values of the atomic coordinates the procedure is one of unconstrained optimisation.

It must be noted that, although the optimisation procedure involves the use of Cartesian coordinates, they must be converted to internal coordinates for the calculation of steric energy since these are the variables involved in the potential energy functions (see Chapter 2).

The optimisation of the function which represents the molecular potential energy hypersurface is a fairly difficult task due to the complexity of the function and the large number of variables involved. Several optimisation procedures have been developed, however none exists which can handle conformational energy minimisation from the trial starting structure to the final minimised conformer. In general two different techniques are used - direct search methods which can cope with trial structures which are far from the minima and gradient methods which can be applied to structures which are closer to the minima.

Direct Search Techniques

With direct search methods the location of minima relies heavily on evaluating the function f(x) at a series of points covering a specified region of the energy hypersurface. This is followed by examination of the function values to locate the minima.

Direct search methods are typically used for functions which are discontinuous or non differentiable, see fig 3.15.

imposed upon the values of the atomic coordinates the procedure is one of unconstrained optimisation.

It must be noted that, although the optimisation procedure involves the use of Cartesian coordinates, they must be converted to internal coordinates for the calculation of steric energy since these are the variables involved in the potential energy functions (see Chapter 2).

The optimisation of the function which represents the molecular potential energy hypersurface is a fairly difficult task due to the complexity of the function and the large number of variables involved. Several optimisation procedures have been developed, however none exists which can handle conformational energy minimisation from the trial starting structure to the final minimised conformer. In general two different techniques are used - direct search methods which can cope with trial structures which are far from the minima and gradient methods which can be applied to structures which are closer to the minima.

Direct Search Techniques

With direct search methods the location of minima relies heavily on evaluating the function f(x) at a series of points covering a specified region of the energy hypersurface. This is followed by examination of the function values to locate the minima.

Direct search methods are typically used for functions which are discontinuous or non differentiable, see fig 3.15.

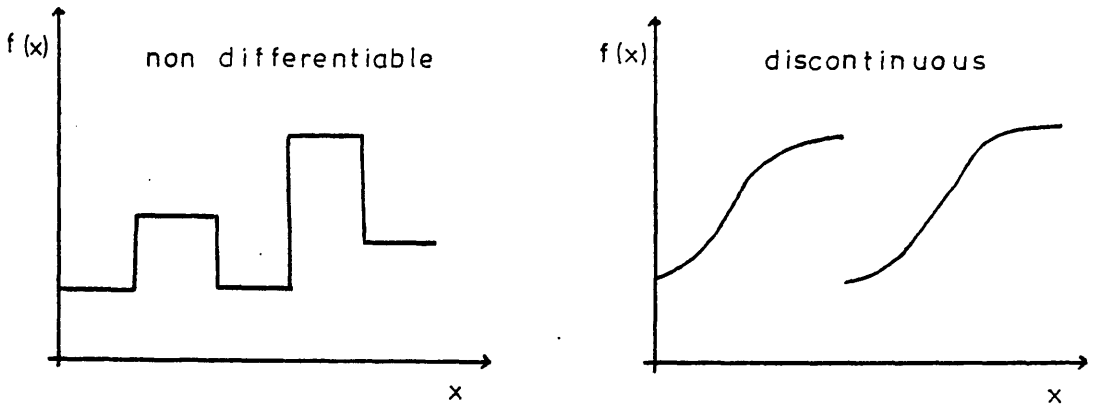


Fig. 3.15.

or where the computational method chosen is dependent upon the amount of computer time available. They can also be used on occasions where a crude answer may be required at any particular instance throughout the calculation.

Direct search methods include the Grid Search method mentioned in the introduction. This is probably the simplest method since it is non-iterative. It has the disadvantages of only being useful in the regions near the minima due to the large number of function values which have to be calculated and also it does not use any information about favourable moves to speed up the calculation.

Other direct search methods include Hooke and Jeeves'[4] method, Rosenbrock's[5] method and the modified simplex method of Nelder and Meade,[6] which is based on the simplex method of Spendley, Hext and Himsworth.[7] All these methods avoid the second disadvantage of the Grid Search method in that they all employ pattern techniques. This involves a series of explanatory moves from which is constructed a table of favourable / unfavourable moves. An attempt to speed up the minimisation procedure is then made by doing a pattern move the direction of which is determined from the table of sucessful moves.

The method used for this work was the Hooke and Jeeves' method.

Hooke and Jeeves' Method

Hooke and Jeeves' method is one of the most straightforward direct search techniques for optimising the function f(**x**) which describes the molecule's energy hypersurface.

The method involves a series of exploratory and pattern moves and requires a starting (or base) point described by the vector $\mathbf{x}_1$ and a set of step lengths h_j for all the variables x_j (i.e. the Cartesian coordinates of the atoms), a consistant step length is frequently used for all the variables.

The exploratory moves provide information about the locality of the starting point $\mathbf{x}_1$ and are carried out as follows:-

E1) Evaluate the function for $\mathbf{x}_1$

E2) Increment x_1 by h_1 and evaluate $f(\mathbf{x}_1)$. If the energy has decreased save the new coordinate, if not decrement x_1 by h_1 and evaluate $f(\mathbf{x}_1)$ save the new coordinate if the energy has decreased if not keep the original coordinate.

E3) Repeat step E2 for all the variables x_j and step lengths h_j. This will give a new base point $\mathbf{x}_2$.

E4) If the new point arrived at, $\mathbf{x}_2$, is different to $\mathbf{x}_1$, proceed to the pattern moves, otherwise halve the steplengths and return to step E2.

Pattern moves decrease calculation time by utilising information obtained about f(**x**) in the region of **x**.

P1) The sucessful exploratory moves which produced $\mathbf{x}_2$ from x_1 are

reapplied to give point p_1.

P2) A series of exploratory moves are then carried out about p_1.

If the energy has decreased during either step P1 or step P2 a point x_2 has been reached and this point is then subjected to more pattern moves.

If the energy has not decreased a further set of exploratory moves are carried out about point x_2. The procedure ends when the step lengths have been halved a set number of times.

Gradient Methods

The position of a minimum on the potential energy hypersurface is that point at which by definition the first derivative of the potential energy function is equal to zero:-

$$\frac{d\,V}{d\,x} = 0$$

Several gradient methods exist for locating this position and they can be divided into methods which:-

a) do not require any derivative calculation.

b) use the gradient of the potential energy surface to locate the minima (1st derivative methods).

c) use information about the gradient and curvature of the potential energy hypersurface to locate the minima (1st and 2nd derivative methods).

Smith's[8] method of conjugate directions and Powell's[9] method of conjugate directions are both descent methods which do not require the

calculation of derivatives.

First derivative techniques include the Steepest Descent method,[10] the Pattern Search method of Schleyer[11] and the Davidon - Fletcher - Powell Variable Matrix method.[12,13]

The Newton - Raphson method involves the calculation of both 1st and 2nd derivatives and is commonly used for conformational energy minimisation calculations.[14,15]

The Newton - Raphson Method

First derivative methods use the gradient of the potential energy function to calculate the size and direction of the correction step required to move from the current base point **x** towards the nearest minimum point. The gradient is, of course, calculated from the 1st derivatives of the function f(**x**). As with all the gradient methods we are unlikely to reach the minimum in one step so this process must be iterative.

The Newton - Raphson method relies heavily on the expansion of the first derivatives $g(\mathbf{x}) = f'(\mathbf{x})$ to a Taylor series:-

$$f(x + \delta x) = f(x) + (\delta x)g + (1/2)(\delta x)^2\mathbf{H} + \text{etc.}$$

where **g** is the 1st derivative (or gradient) vector, **H** is the second derivative matrix, and δx is the increment applied to **x**.

The series is truncated to the quadratic - the simplest function giving a strong minimum. Since the first derivatives are equal to zero at the minimum differentiating the above equation gives:-

$$g + (\delta x)\,\mathbf{H} = 0$$

where H is the Hessian matrix. From this equation we get the iterative formula:-

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}^{-1}\mathbf{g}$$

where $\mathbf{x}_k$ is the current point, $\mathbf{x}_{k+1}$ is the next current point and $\mathbf{H}^{-1}$ is the inverse of the Hessian matrix.

When dealing with Cartesian coordinates problems arise with inversion of the 3N x 3N Hessian matrix. These problems arise because the matrix is sixfold singular due to the three rotational and translational degrees of freedom which do not affect the molecules energy. This problem can be overcome in one of the following ways:-

1) Conversion from Cartesian coordinates to internal coordinates.

2) Rotation and translation of the molecule can be prevented by fixing 6 atomic coordinates. This is generally done by fixing one atom such that x = y = z = 0, a second such that x = y = 0 and a third such that x = 0. This will remove six rows and columns from the 3N x 3N matrix and it will no longer be singular. However, care must be taken that no bond lengths or bond angles are being restricted.

3) The generation of a generalised inverse matrix.[16,17] This can be done by using the six zero eigenvectors and their corresponding eigenvalues which are calculated from the diagonalised matrix.

4) The use of Eckert constraints[18,19] which fix the centre of mass in space and constrain infinitesimal rotations about the molecule.

5) The generation of a non-unique reciprocal matrix using modified Choleski procedures[20,21] such that the resultant matrix is dependent upon the initial orientation of the molecule.

The first two procedures give problems if used alone so various combinations of the above procedures are quite often used.

The Newton - Raphson method converges more rapidly than the 1st derivative methods due to its use of curvature information, however, it has major drawbacks in that the calculation of 1st and 2nd derivatives involves large amounts of computer time as does inversion of the Hessian matrix. This has led to the development of several faster techniques in which different approximations are made to the Hessian matrix.

1) The Block Diagonal Method

With the Block Diagonal method the matrix is split into 3 x 3 submatrices $x_{(i,j)}$ such that i,j=1,3 for each atom. The submatrices along the diagonal contain large values related to the energy changes brought about by movement of an individual atom in the Cartesian direction. The off diagonal submatrices are related to cross terms between pairs of atoms and tend to be small. If these off diagonal submatrices are set to zero far less computer time is required for the calculation.

2) The Pure Diagonal Method.

With the Pure Diagonal method all of the off diagonal elements of the matrix are set to zero. This again speeds up the calculation.

The Full Matrix method is the one discussed previously where all the elements of the matrix are left intact. Obviously it is the

slowest method for each individual iteration, however it converges very rapidly in 5 - 6 iterations, but is very intolerant of poor starting structures and requires starting structures close to the minimum.

The Block Diagonal method is much faster to compute and can tolerate slightly poorer starting structures than the Full Matrix method, however, convergence is slower and elements of symmetry can be lost with this method.

The Pure Diagonal method is the fastest variant of the Newton - Raphson method used for this work. It is very tolerant of poor starting structures,[22] but is slow to converge and again symmetry elements may be lost.

No one minimisation procedure can cope with full minimisation of a trial structure. The general procedure found to be best for the minimisation of trial conformations is:-

1) 2 - 5 cycles of Hooke and Jeeves` pattern search minimisation.

2) Pure Diagonal minimisation until the 1st derivatives are of the order of 0.1, usually 500 or more cycles for the larger molecules.

3) Approximately 5 cycles of Full Matrix minimisation - this takes the derivatives down to 10^{-6} - 10^{-7}, gives the best representation of the molecule and recovers any elements of symmetry which may have been lost during the Pure Diagonal minimisation. With Full Matrix minimisation the energy of the molecule may not be lowered significantly, but the geometry of the molecule can be substantially altered due to torsion angle changes.

CHAPTER FOUR

Section 4.1.

Molecular Modeling and Interactive Computer Graphics

The way in which the majority of people relate to their surroundings is by their perception of three dimensional images which are constantly being processed by the brain. Since people are adapted to respond to visual stimuli it is convenient to illustrate theoretical models pictorially, where possible. Interactive computer graphics is therefore an ideal way in which to present vast amounts of computer generated data, to the user, in a mode which is both easy to interpret and comprehend.

For a long time after the birth of computer graphics in the 1950's[1] the vast resources it required, both for the hardware and high quality output devices, kept it outwith the reach of the average computer user. However, in recent years graphics has benefited from a tremendous reduction in the price/performance ratio for hardware and interactive graphics is now feasible and practical for the vast majority of applications that require only a modest processor and display resources.

In molecular mechanics we are interested in studying the relationship between some of the physical properties of molecules and their three dimensional structures. In particular we are interested in examining low energy conformations of molecules. Interactive computer graphics is an exceptionally useful tool for molecular mechanics as it allows us to observe and investigate how changes in the spatial

arrangement of a molecule will affect its molecular energy.

Section 4.2.

The Hardware

A diagram of the hardware on which our molecular mechanics programs are implemented is shown in Fig. 4.1. The system consists of a DEC PDP 11/40 minicomputer with 128 kbytes of memory operating under RT11 Version IV, two RK05 2.5 Mbyte disc drives, a Megatek 7000 series graphics processor with a calligraphic display terminal, an FPS 100 array processor for number crunching operations, a Thorn EMI printer plotter, a Tektronix electrostatic plotter, a VT100 visual display unit, a high speed paper tape reader, a Diablo daisywheel printer and a 300 baud modem.

There is a second PDP 11/40 with 256 kbytes memory, a 40 Mbyte Plessey disc drive, a Fungus v.d.u., a Systime SYSTAPE II magnetic tape drive, a VT11 graphics terminal and a Decwriter. This system is mainly used as a file storage unit, as on it is kept the source code for the original and the Megatek versions of the molecular mechanics programs.

We also have access to the University's mainframe, an ICL 2988 and, due to their size, a few minimisation routines are implemented on it.

The focal point of the system is the Megatek 7000 series refresh graphics. This has a high resolution (4096 x 4096) monochrome, calligraphic display terminal, the programs for which operate with the Megatek MGS graphics software package. The calligraphic terminal has the advantage of being a good medium with which to display large

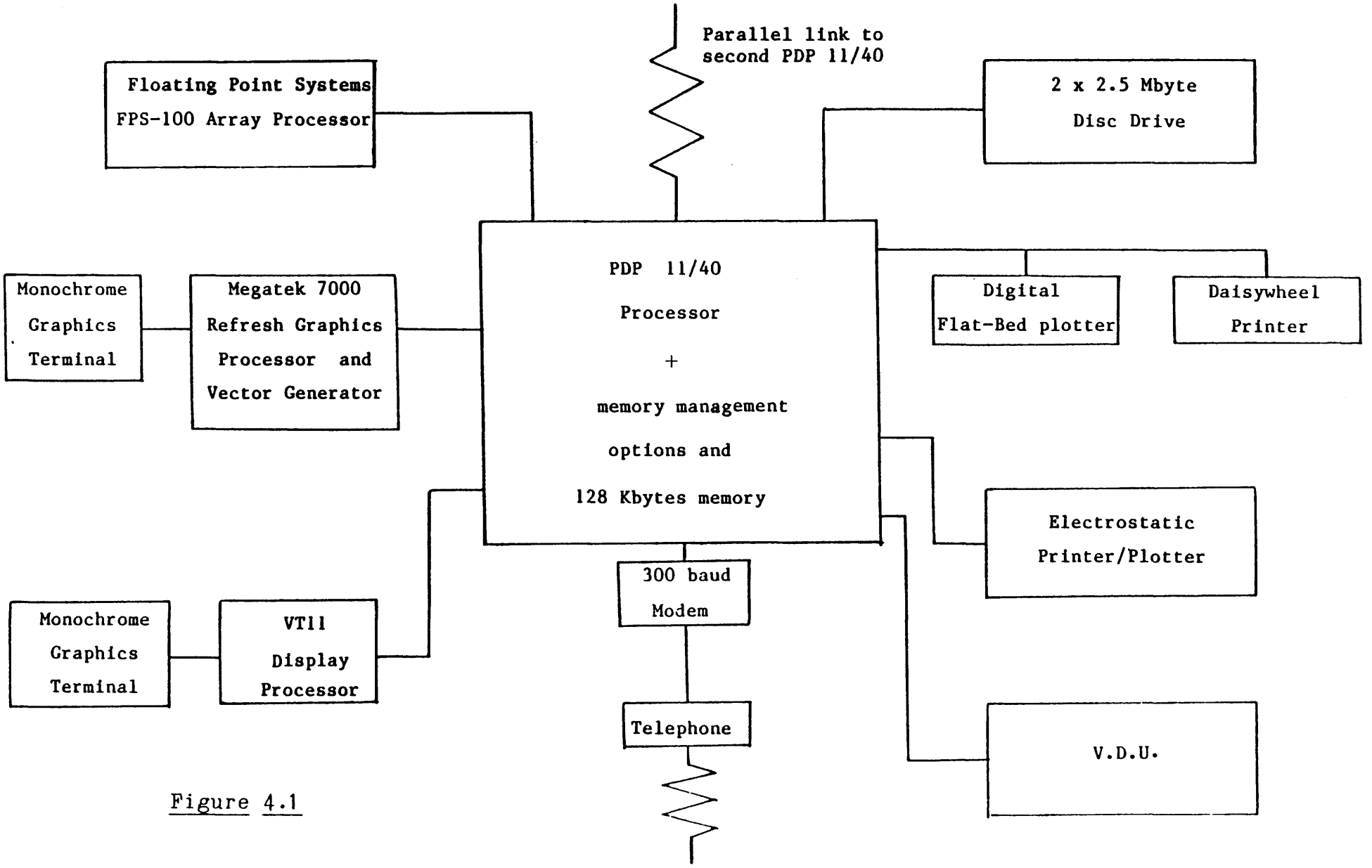

Parallel link to second PDP 11/40
Floating Point Systems FPS-100 Array Processor
2 x 2.5 Mbyte Disc Drive
PDP 11/40 Processor + memory management options and 128 Kbytes memory
Monochrome Graphics Terminal
Megatek 7000 Refresh Graphics Processor and Vector Generator
Digital Flat-Bed plotter
Daisywheel Printer
Electrostatic Printer/Plotter
300 baud Modem
Monochrome Graphics Terminal
VT11 Display Processor
Telephone
V.D.U.

Figure 4.1

molecules at high resolution. However, it has the disadvantages of not having the facility to shade areas of the screen, or use different colours to designate different atom types. More and more we are seeing the use of high resolution colour, calligraphic and raster terminals to display complex molecular structures.

Section 4.3.

The Software

Most theoretical calculations on molecules require as input a set of Cartesian coordinates. Previous to the advent of interactive computer graphics, the user would have had to wade through large amounts of numerical data, with only mechanical models and two dimensional drawings to give him any idea of molecular conformation. As discussed in section 4.1, for certain applications, interactive computer graphics is currently one of the best methods of communicating directly with a computer, especially for the interpretation of output.

The MOL chemical graphics system is an interactive system which has been developed with the average chemist in mind. With MOL the user is able to peform various types of geometry measurements, energy calculations and manipulation to molecules while viewing a three dimensional molecular image. It is a user friendly system with easily understood input/output, the input being interpreted by text handling routines which allow input to be entered in relatively free format and report any input format and subject errors.

The system is built as a series of individual programs controlled by a root segment, which selects any program required by the user and

chains to and from it. Each individual program resides on a disc until it is selected by the root segment. Only the program in current use occupies core memory, but to the user chaining is totally transparent and is analogous to subroutine usage. Thus, once the MOL system has been initiated it behaves as if it were one large program. Chaining therefore gives the advantage of being able to incorporate several large programs into the graphics system without exceeding the memory. The main disadvantage with this method of program management is that chaining to the individual programs is relatively slow. It must also be noted that the individual programs are structured such that most of their subroutines may be overlaid.

The programs make good use of the RT11 memory management facility and some of the programs may be used in stand alone mode.

Section 4.4.

MOL and its Options

The MOL system is capable of handling up to 500 atoms (this does not apply to certain options), either in the form of one molecule, or as several fragments. The system has two modes of operation, SINGLE mode and MULTIPLE mode. In SINGLE mode all atoms in the picture display list are treated as belonging to the same fragment and any calculations (eg. rotation, minimisation) are carried out on all the atoms present. In MULTIPLE mode there may be up to 127 fragments present in the system and various operations may be carried out on each individual fragment.

The atoms in each fragment are basically stored in arrays

connectivity. Most input to the system is therefore in the form of numerical data and/or atom names and error checking is an integral part of the system.

On entering the system the user is presented with a menu of 39 options any of which may be selected using a crosshair cursor controlled by a joystick. Initially the cursor is limited to those options which load a molecule into the system (i.e. the first five).

Once the molecule is in the workspace it is drawn as a stick diagram (see Fig.4.2). An attempt is made to represent three dimensionality by depth cueing the molecule. This is done by drawing bonds to the front of the molecule (+ve z coordinates) at a higher intensity than bonds to the back of the molecule (-ve z coordinates). In addition each bond is composed of two vectors which can be drawn with different intensities. The user gains the best idea of the three dimensional arrangement of the molecule when it is undergoing dynamic rotation (see section 4. 4. 18.).

Section 4.4.1.

GET FILE

This option allows the user to read an ASCII file from a disc into the graphics system. These files are structured in one of two formats - a numeric format and a non-numeric format. Figs 4.3. and 4.4 give examples of these formats for a propanol molecule.

There are ten separate sections within the system format file, making it quite an inefficient way of storing data as a lot of the information is duplicated. It is hoped that in the near future a standard file format will be introduced for use by the Molecular

cyclohexane

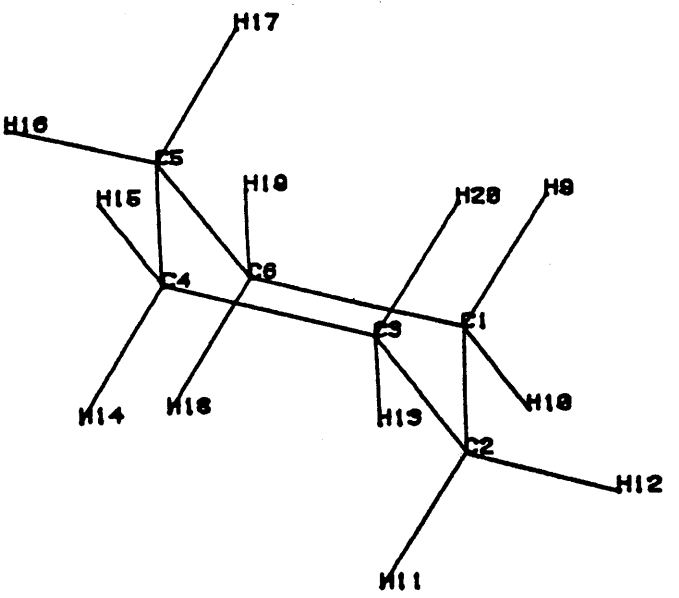


Fig. 4.2.

```
--MOL CHEMICAL GRAPHICS SYSTEM FILE--
TITLE
Propan-1-ol
**
DATE
03-AUG-85
**
COMMENTS
**
UNIT CELL
 1   1   1   90   90   90
**
COORDINATES
C1              1.50000    1.50000    1.50000     1
C2              3.04100    1.50000    1.50000     1
C3              3.55464    2.95288    1.50000     1
O4              4.98464    2.95288    1.50000     1
H1              3.40569    0.98431    2.39326     1
H2              3.40569    0.98431    0.60674     1
H3              3.19000    3.46860    0.60674     1
H4              3.18999    3.46860    2.39325     1
H5              1.13536    2.01572    0.60675     1
H6              1.13536    2.01572    2.39325     1
H7              1.13536    0.46856    1.50000     1
H8              5.30956    2.50652    0.72690     1
**
CONNECTIVITY
C1        C2        H5        H6        H7
C2        C1        C3        H1        H2
C3        C2        O4        H3        H4
O4        C3        H8
H1        C2
H2        C2
H3        C3
H4        C3
H5        C1
H6        C1
H7        C1
H8        O4
**
ATOM TYPES
  1  C1        C2        C3
 10  O4
 15  H1        H2        H3        H4        H5        H6        H7        H8
**
CHARGED ATOMS
**
END
```

Fig. 4.3.

```
--MOL CHEMICAL GRAPHICS SYSTEM FILE--
TITLE
Propan-1-ol
**
DATE
03-AUG-85
**
COMMENTS
**
UNIT CELL
 1  1  1  90  90  90
**
COORDINATES
C1          1.50000   1.50000   1.50000    1
C2          3.04100   1.50000   1.50000    1
C3          3.55464   2.95288   1.50000    1
O4          4.98464   2.95288   1.50000    1
H1          3.40569   0.98431   2.39326    1
H2          3.40569   0.98431   0.60674    1
H3          3.19000   3.46860   0.60674    1
H4          3.18999   3.46860   2.39325    1
H5          1.13536   2.01572   0.60675    1
H6          1.13536   2.01572   2.39325    1
H7          1.13536   0.46856   1.50000    1
H8          5.30956   2.50652   0.72690    1
**
CONNECTIVITY
    1    2    9   10   11
    2    1    3    5    6
    3    2    4    7    8
    4    3   12
    5    2
    6    2
    7    3
    8    3
    9    1
   10    1
   11    1
   12    4
**
ATOM TYPES
  1    1    2    3
 10    4
 15    5    6    7    8    9   10   11   12
**
CHARGED ATOMS
**
END
```

Fig. 4.4.

Graphics fraternity. This format should prove a more efficient method of storing data, while remaining flexible enough for use in the many different areas of computational chemistry.

The first six sections of the file contain the title for the file of upto eighty alphanumeric characters, the date the file was created, any references, comments and space group information and the unit cell parameters

In the numeric format atoms are stored by number and in the non-numeric format they are stored as atom names. Files are more normally stored in the non-numeric format. The coordinate section holds the atoms by name or number and their x, y and z coordinates. From the connectivity section we can determine which atoms are bonded to which. Each atom can have a maximum of four substituents, depending on type and the atoms are listed in lines with their substituents.

In the MOL graphics system we wish to be able to identify different types of atoms. Names are not necessarily enough, for example C1 could be an sp, sp2 or sp3 type of carbon, so numbers are used to designate atom type (a list of these is given in Fig. 4.5). The atom types section holds the atom type of each of the atoms in the molecule.

The final section is the partial charges section. This contains a list of any charged atoms and the charges on them.

Type Number	Description
1	Carbon (sp^3)
2	Carbon (sp^2)
3	Carbon (aromatic)
4	Carbon (sp)
5	-N(+)RR'R"
6	-NRR' (tetrahedral)
7	N (aromatic)
8	N(-) (aromatic)
9	N (trigonal)
10	Oxygen (single)
11	Oxygen (double)
12	Sulphur (diagonal)
13	Sulphur (tetrahedral)
14	Protonated nitrogen
15	Hydrogen
16	Fluorine
17	Chlorine
18	Bromine
19	Iodine
20	Metal

MOL Atom Types

Fig. 4.5.

Section 4.4.2.

NEW FILE

Structures may be input from the keyboard with this option in the form of names, atomic coordinates etc. This is very useful for entering crystallographic data. It is essential that the user inputs a title, a unit cell and the atom names and coordinates. Any remaining information such as connectivity, atom types and partial charges can also be entered in NEW FILE, but these data can be entered via other MOL routines. NEW FILE cannot be used to add to another molecule since any existing structures are deleted as soon as the option is chosen.

Section 4.4.3.

DATABASE

A component of the MOL graphics system is a database, containing the structures of a few small hydrocarbons which may be used as fragments when building larger molecules. To select a structure the user must give the database name and the file name, the program will then search the database for the file and if it is found it is loaded into the system.

Section 4.4.4.

CSSR

The Crystal Structure Search and Retrieval systems database is maintained by the Science and Engineering Research Council on various computers including one in Edinburgh which we are able to access with

a 300 baud telephone modem. Files which are in the CSSR format may then be read into the system using this option.

Section 4.4.5.

QUICK BUILD

It is of prime importance in any interactive molecular modeling system that the user can generate molecules without too much difficulty. The QUICK BUILD routine allows the user quickly to build a new molecule, add on to an existing molecule, or in multiple mode, build a new fragment. If there are no atoms in the workspace, or if a new fragment is to be created, the user is asked for an atom name and type. Once there are atoms present the user is asked for an existing atom name and the name, type and optionally partial charge of the new atom to be connected to it. This new atom is then built on to the existing structure using standard bond lengths, angles etc. Carbon chains are built as open chains (i.e. all trans). The program will check the existing atom to make sure that it does not already have its maximum number of substituents and it will also check the new atom name to make sure that this name is not already in existence.

Section 4.4.6.

PUT FILE

A systems format file of the current display structure may be written to a disc file with this option. The user is prompted for an output file name, the program then checks the disc and if a file of that name already exists the user can either overwrite it or chose a

new file name. The option also allows the user to put comments in the file and dictate whether the file is to be in numeric or non-numeric format.

Section 4.4.7.

JOIN

There are two modes of operation in this option. It can either be used to join all non-bonded atoms within a certain distance of one another, or, the option may be used to form (or break) a bond between specified atom pairs. The option checks carefully the valency of the atoms. Atoms which have their maximum number of substituents cannot form any new bonds. If one atom of a specified pair has a full valency the user is given the option of making a one way join from the other atom. This can be useful for displaying metal-ligand complexes and other inorganic structures. JOIN may also be used to join two atoms in different fragments to make a larger molecule. In this case all existing fragment numbers are updated.

Section 4.4.8.

BUILD

The current display structure can be altered or augmented with this routine. It is similar to QUICK BUILD except it does not use tables of standard bond lengths etc. Instead the user is asked for four atom names (three existing atoms plus a new one), a bond length, a bond angle, a torsion angle, an atom type and optionally a partial charge for the atom. Obviously it is a much slower routine to use than

QUICK BUILD, however it is exceptionally useful for building structures where standard bond lengths, and angles do not apply eg. heterocyclic compounds, or, when building rings since using QUICK BUILD would result in a straight chain which would need to be modified in ALTER.

Section 4.4.9.

FBUILD

With FBUILD the user can combine two fragments. As input the routine requires four atom names in the format:-

atom A - atom H1 atom H2 - atom B

where atom A and atom B are the two atoms to be joined and atom H1 and atom H2 are the two hydrogen atoms which will be eliminated. The routine positions the fragments containing atoms A and B so that the bond to be formed between them will have the correct length and valence angle. The hydrogens are eliminated and a bond is formed, the length is taken from a table of standard bond lengths according to the atom types. This routine is very effective for building large molecules (eg. polymers) from smaller substituents.

Section 4.4.10.

REMOVE

The remove option allows the user to remove one, or more atoms from the current display structure, in MULTIPLE mode whole fragments may be removed. The program deletes atoms, or fragments by editing the relevent vectors from the display list, without rescaling, or

redrawing the picture. The same method is used for deleting bonds in routine JOIN when breaking bonds. All data concerning the deleted atoms is removed from the relevent storage arrays (eg. atom names, coordinates etc). Existing structures may be easily modified by simply removing unwanted atoms and building on new ones with, for example QUICK BUILD or BUILD.

Section 4.4.11.

ADD H'S

Any missing hydrogens may be added to a structure with this facility. The program will add hydrogens to any atom which does not have its full number of substituents determined by the atom type. The hydrogen atoms are added according to a standard geometry, so most hydrogens may be added without any user interaction. However, where these rules are not applicable (eg. in a methyl group) the user is prompted for a torsion angle value.

Section 4.4.12.

ALTER

ALTER can be used in one of two ways. Either interactively to alter the value of a specified torsion angle, while monitoring the variation in distance between two named atoms. Or, alternatively, it can be used to determine which are the most and least energetically favourable positions for a given torsion angle. This is achieved by calculating changes in the torsional, van der Waals and coulombic components of the potential energy, relative to the changing torsion

angle as it proceeds from 0° to 360° in user specified steps (eg. 10°). A scaled plot of the potential energy versus the torsion angle is then drawn and a hardcopy of this is available (see Fig.4.6). The energy scale for the plot has a cut off value of 200 kcal mol^{-1} since we are more interested in the minima than the maxima.

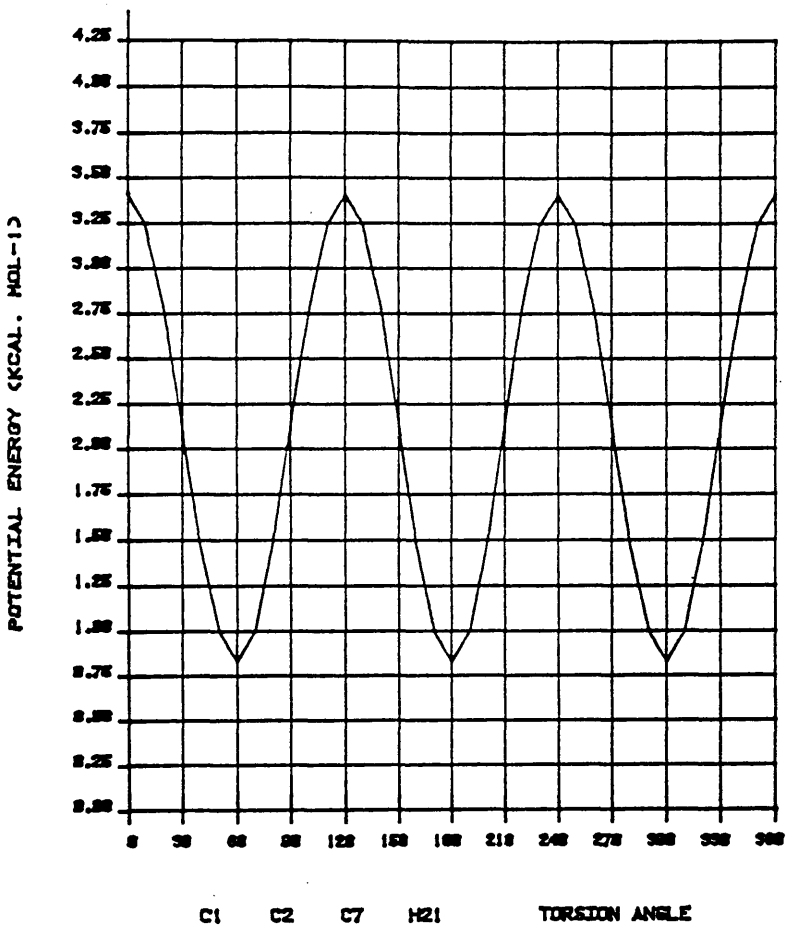
Section 4.4.13.

RAMMAP

With this routine the user can calculate and plot Ramachandran maps.[2,3] The plotting routine may be used in stand alone mode. Ramachandran maps enable the minimum positions of two torsion angles to be determined. The user is prompted for the two torsion angles of interest (by atom names) and a step size for the 0° to 360° rotation of each torsion angle. The program then sets each torsion angle to 0° and holds torsion angle 1 at this while rotating torsion angle two stepwise from 0° to 360°. Torsion angle 1 is then incremented by the value of the step and torsion angle 2 is again rotated stepwise from 0° to 360° until all possible permutations have been covered. With each step the program calculates and stores the potential energy changes involved using the same parameters as ALTER.

Once the routine is finished the user is given the option of plotting a grid, or a contour map of the variation of potential energy against the two torsion angles (see Figs. 4.7 and 4.8). From these plots we can determine the minimum energy positions for the torsion angle pair.

ALTER - Hardcopy from Screen.



Methyl-cyclohexane

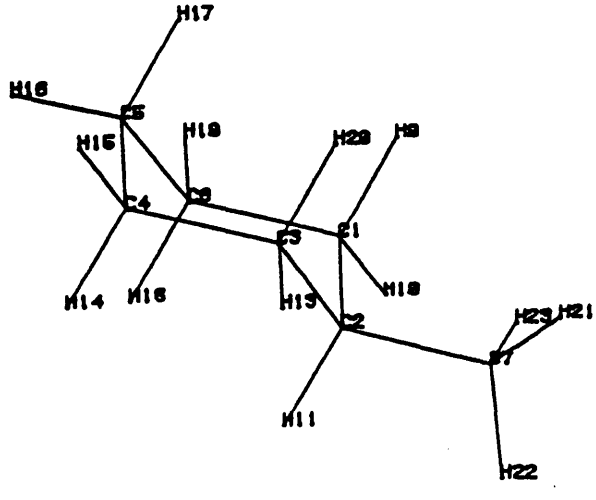
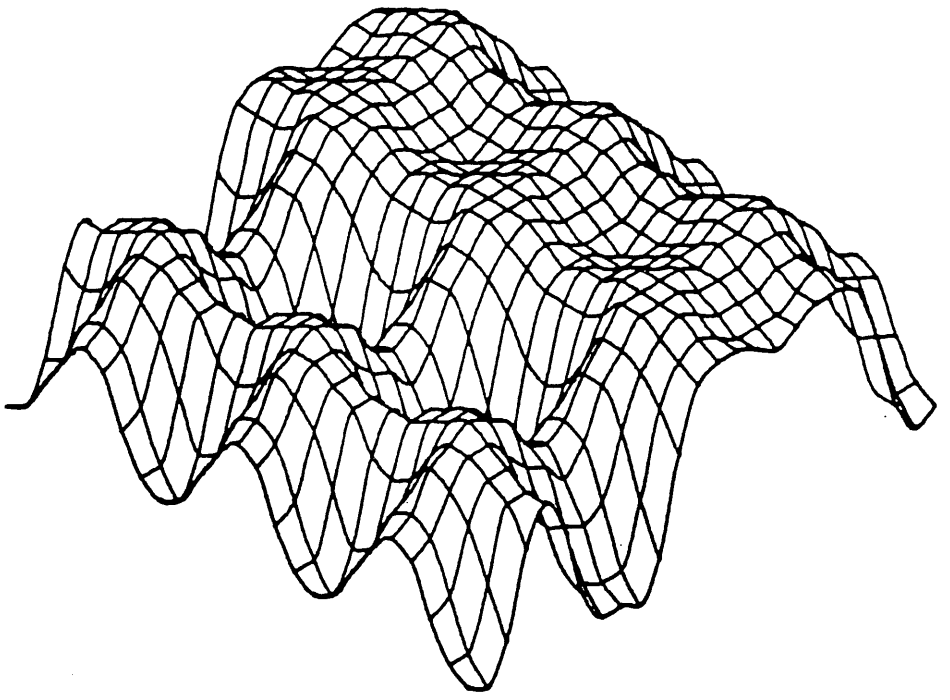


Fig. 4.6.

RAMMAP - Grid Plot.



Ethyl-cyclohexane

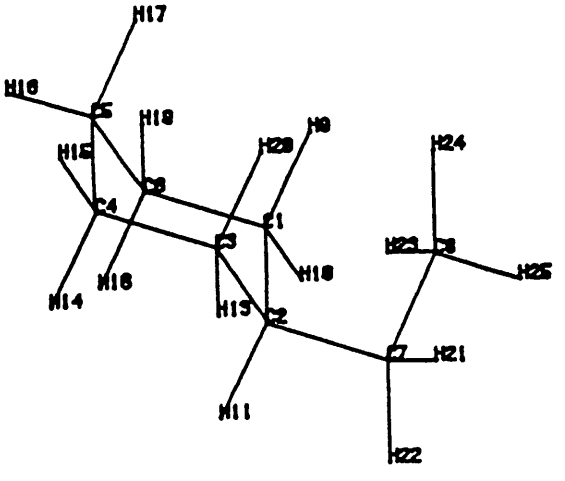


Fig. 4.7.

RAMMAP - Contour Plot.

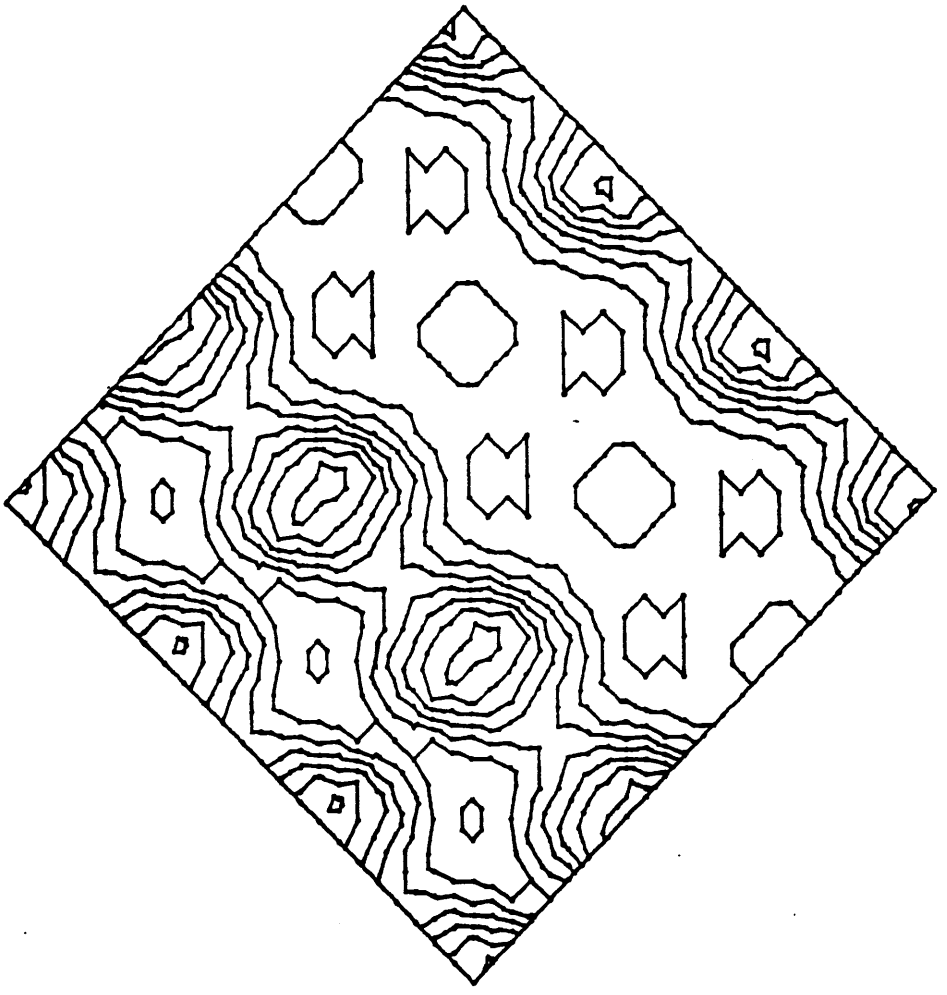


Fig. 4.8.

Section 4.4.14.

CLOSE

One of the most frustrating problems that arises while building rings is that the ends of the chain are invariably the wrong distance apart to make a ring. CLOSE will produce a more favourable end to end distance by tweaking each of the ring torsion angles by a maximum of + or - 10°. Bond lengths and angles are unaffected.

As input the user must enter the two atom names and the target end to end distance. In some cases it is not possible to attain this distance with one cycle of CLOSE and the whole procedure must be repeated.

Section 4.4.15.

DISTANCE

This is the first of three options which perform geometry calculations on the display structure. As its name implies the DISTANCE option calculates distances between specified atom pairs. The input is either in the form of atom names, or if more than one distance is required wild cards may be entered instead of atom names. when wild cards are used the user may specify whether he requires the program to calculate and list only bonded distances, or all interatomic distances. This data can be relayed to the line printer, a disc file, the v.d.u., or the Megatek screen.

Section 4.4.16.

ANGLE

Any angle in the molecule may be calculated with this option. The user is asked to specify three atoms by name and like DISTANCE one or more atoms may be replaced by wild cards. When wild cards are used this routine will only calculate angles between bonded atoms - a listing of which may be sent to a user specified device. Angles which are not bond angles have to be specifically requested.

Section 4.4.17.

TOR ANGLE

This is the third option which performs calculations on the geometry of the structure in the workspace. It enables any torsion angle in the structure to be calculated. The names of the four atoms defining a torsion angle have to be entered. If wild cards are used a list of all torsion angles between bonded atoms (including those containing hydrogen atoms if requested) is again relayed to the user specified device. Torsion angles including non-bonded atoms must be explicitly specified.

Section 4.4.18.

R D FUNC

This option plots a radial distribution function for the current structure. That is to say it produces a plot of the number of non-bonded atoms which are a given distance apart versus that distance and

is useful for comparing molecules which have similar activities. A hardcopy of the plot is available (see Fig. 4.9).

Section 4.4.19.

ROTATE

ROTATE enables the user interactively to rotate a molecule, or fragment, about the X and Y axes using the joystick and the Z axis using the letters P and N for clockwise (positive) and anticlockwise (negative) rotation, any other key will stop the rotation. Although a Megatek graphics sytem is available with a hardware rotate our particular system does not have it and this rotation option uses software routines. The X and Y axes rotation is achieved by calculating a rotation matrix from the position of the joystick and the multiplying the atomic coordinates by it. To prevent a staggering effect during the rotation double buffering is used. This involves using two of the eight picture planes in the Megatek. Each plane can hold a separate picture and may be switched on or off. The picture of the current structure is held in one plane, which is switched on. The structure is rotated and the new picture is drawn in the second plane which is switched off. As the first picture is switched off the second picture is switched on etc.

Depth cueing is maintained during the rotation and this proves very helpful in giving an idea of which part of the structure is rotating towards the user and which parts are moving away.

As part of this project it had been hoped to rewrite this routine to run on the array processor, however, it was found that there was

A Typical Radial Distribution Function

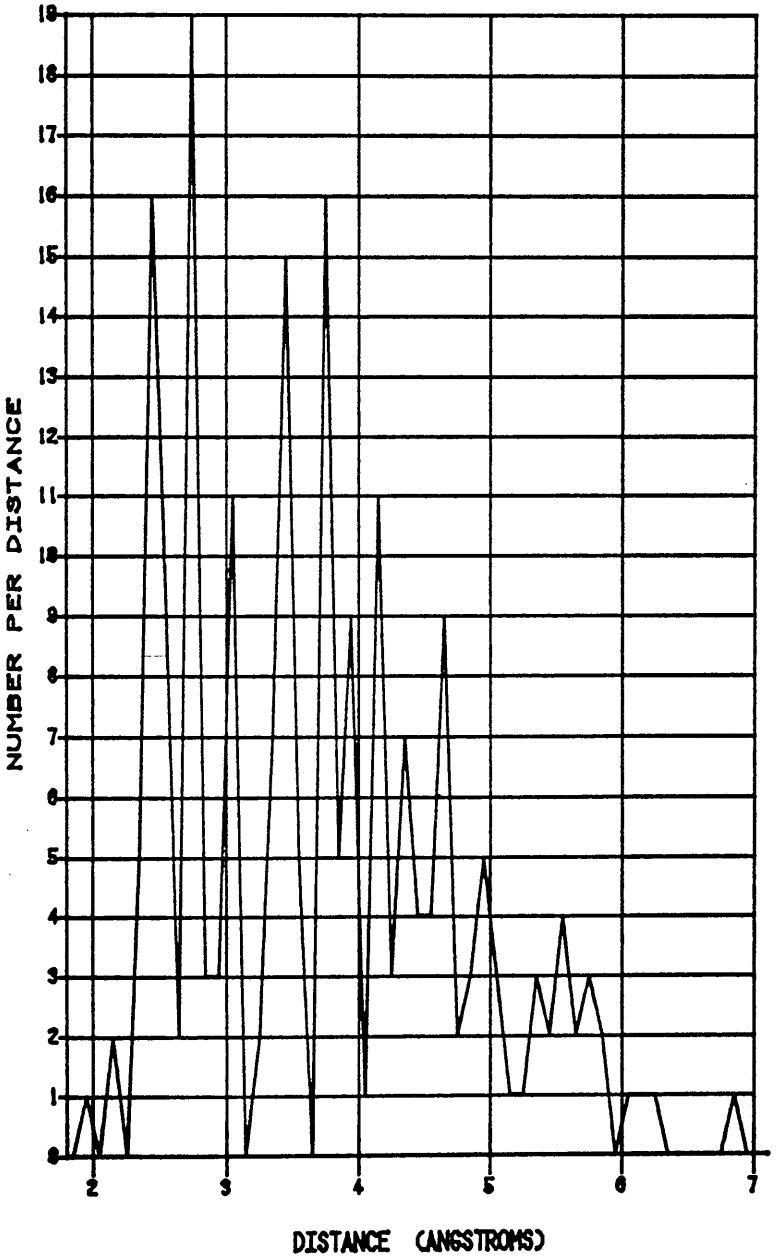


Fig. 4.9.

not enough memory available on the system to link the array processor version with all the library subroutines it required.

The array processor version was however sent to Floating Point Systems in Bracknell where it has been implemented on a Vax 11/780.

Section 4.4.20.

TRANSLATE

There are two parts to TRANSLATE in the MOL graphics system. A hardware translate involving the X-Y plane and a software translate involving the Z direction. With this option one can translate a fragment with respect to the rest of the display and if required the program can monitor the distance between an atom from each fragment as the translation proceeds. This enables the user to maneouvre two fragments to the correct distance apart for, say BUILD. The system must be in MULTIPLE mode for translation

Section 4.4.21.

VPERP

As its name implies this option gives a view perpendicular to a plane which has been specified by entering three atom names.

Section 4.4.22.

ARBROT

This option allows the user interactively to rotate the current structure about an arbitrary axis which has been defined by any two atom names - not necessarily in the same fragment. The rotation is

carried out using the joystick.

Section 4.4.23.

BEST VIEW

The best view of any molecule, or fragment, in the current display may be obtained using this very useful routine. The definition of the best view is the view which has the sum of the squares of the distances of atoms from the plane of the screen as a minimum. Hopefully this gives the user a view of the molecule with as few overlapping atoms as possible.

Section 4.4.24.

S'IMPOS

The superimpose option facilitates the superimposition of one fragment on another. As input the program requires as many pairs of corresponding atoms, one from each fragment, as will define the region of overlap. The program then superimposes the two fragments using a least squares fit technique and then calculates the RMS fit for the two structures. Finally the two superimposed fragments are redrawn on the display. This option can only be used in multiple mode and is mainly useful for comparing regions of structural similarity in different molecules.

Section 4.4.25.

HIDE H'S / SHOW H'S

This is one of four toggle type options (ie. on / off). On

entering the system it is set to show hydrogens, however to clarify a complex structure it may be desired to ignore temporarily the hydrogen atoms and this option can be used to hide or display them at will.

The program does not remove, or add, hydrogens from the structure it simply prevents, or enables their being displayed. While the hydrogens are not being displayed their coordinates, names etc. are still held in the storage arrays and any operations carried out on the structure (eg. rotation) are carried out on the hidden hydrogens. Show H's will simply return the hydrogens to the display picture in their correct positions. As with the next three options only one of the two routines is in the menu at any one time and when it is selected the menu is redrawn with the other option now available.

Section 4.4.26.

NO LABELS / LABELS

Again to clarify a structure NO LABELS can be used to switch off the atom names. Option LABELS restores the atom names to the display picture.

Section 4.4.27.

CUE OFF / CUE ON

In the MOL graphics system an attempt is made to represent a three dimensional image by depth cueing the picture (see section 4.3). This feature can be disabled with CUE OFF and all atom names and bonds are then drawn at the same intensity.

Section 4.4.28.

MULTIPLE / SINGLE

This option acts as a switch for the corresponding modes. In SINGLE mode operations are performed on the whole display structure as if it were one fragment. Whereas in MULTIPLE mode operations can be performed on individual fragments within the display picture. MULTIPLE mode also allows the user to load several molecules into the workspace at once.

Section 4.4.29.

CHANGE

If new molecule has been created from a previously existing file it may be desirable to change the title of the file before storing it on a disc, this option allows the user to do this. It also gives the user the opportunity to change the name, type or partial charge of any atom in the current display.

Section 4.4.30.

ORTEP

A specially modified version of the Oakridge thermal ellipsoidal plot program[4] has been implemented as part of the graphics system. The atoms are plotted as spheres not as ellipsoidal structures and there is no hidden line removal in the routine as this would make the program too large for the PDP 11's memory. Depending on the size of the molecule an Ortep diagram of the current display structure,

adequate for most uses can be produced by the system without user intervention. A hardcopy of the Ortep diagram can be plotted on the Tektronix plotter (see Fig. 4.10).

Section 4.4.31.

SPACEFILL

A spacefilled diagram of the current display picture can be drawn with this option. Since the Megatek has a calligraphic terminal curved lines are represented by many short vectors - a total circle is replaced by a 25 sided polygon. The user has the option of using his own table of van der Waals radii, or a default table of standard van der Waals radii, which has been incorporated into the system. A stick model of the structure may be superimposed on the spacefilled structure and a hardcopy is available (see Fig. 4.11).

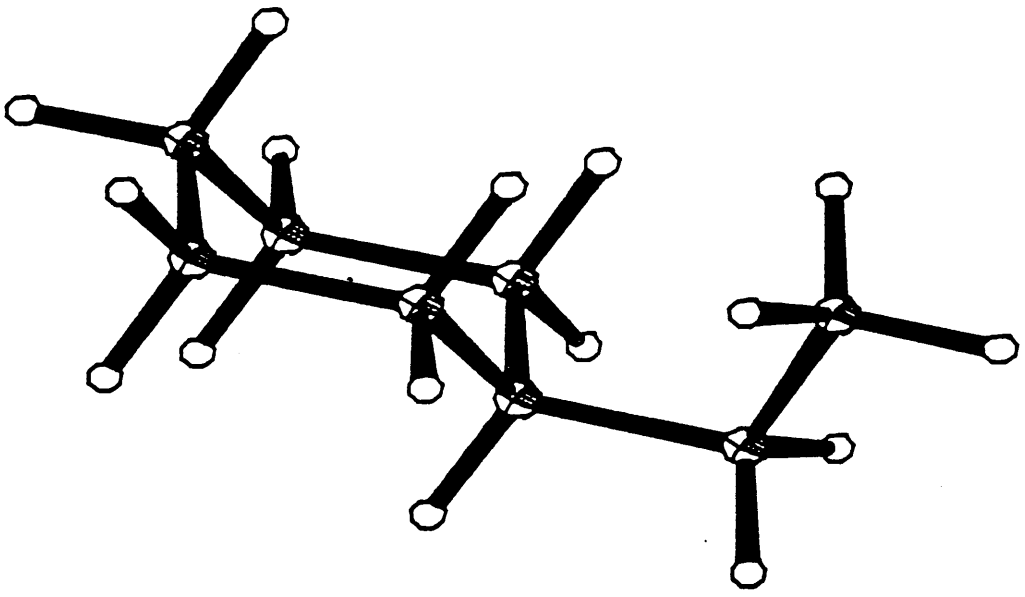
Section 4.4.32.

PD MIN

This option is used to minimise the energy of a molecule. It uses the variant of the Newton Raphson minimisation method in which the Hessian matrix is approximated to a pure diagonal.

There are two versions of this routine, one which runs in the PDP 11 and one which runs in the array processor. They differ in that the one which runs in the PDP 11 is very slow, will not calculate coulombic interactions for partial charges and can take upto 80 atoms, whereas the array processor version is a lot faster (ca. 30 - 40 times), will calculate coulombic interactions for partial charges and can only be used for molecules which have 65 or less atoms.

ORTEP



Ethyl-cyclohexane

Fig. 4.10.

SPACEFILL

Ethyl-cyclohexane

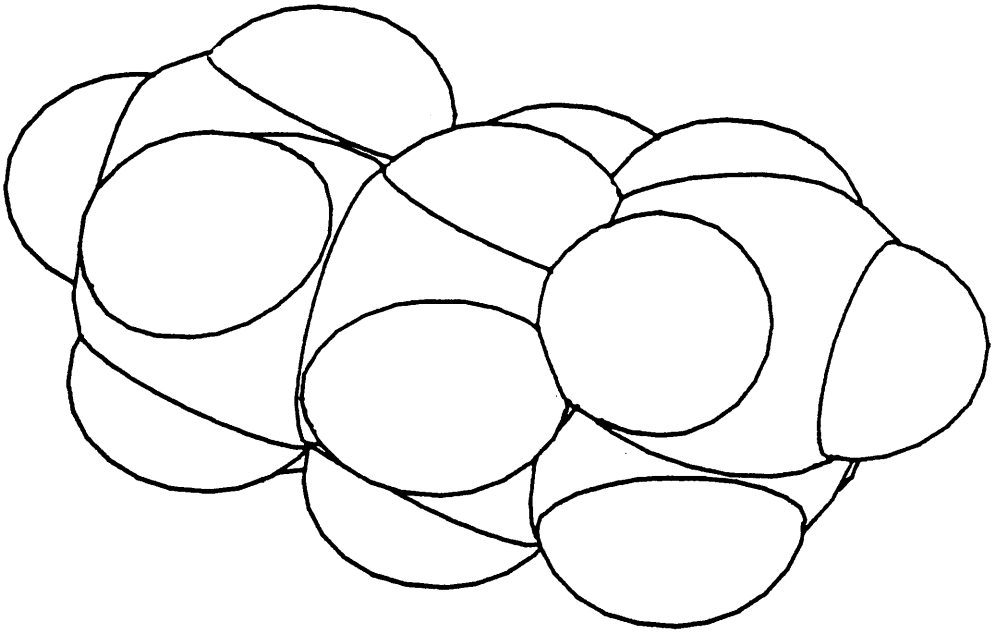


Fig. 4.11.

As input the user must enter a device name for the program output, a force field file name, the number of iterations for the minimisation and whether long ,or short, output is required.

Firstly the program checks the force field to see if it has the correct parameters for the molecule. If any are missing it will report them and the program may either be abandoned, or the interactions for which the parameters are missing will be set to zero. With long output the program list all the individual energy interactions followed by the total contribution of each type of interaction to the total potential energy this is done after each iteration. With short output the program lists the components of the initial and final total energies in the following form.

```
TOTAL E(BONDED) =                          K.CAL PER MOLE
TOTAL E(VAN DER WAALS) =                   K.CAL PER MOLE
TOTAL E(ANGLES) =                          K.CAL PER MOLE
TOTAL E(TORSION) =                         K.CAL PER MOLE
TOTAL E(OUT-OF-PLANE BENDING) =            K.CAL PER MOLE
TOTAL E(COULOMB) =                         K.CAL PER MOLE

TOTAL POTENTIAL ENERGY =                   K.CAL PER MOLE

R.M.S. MAGNITUDE OF FIRST DERIVS=          K.CAL.MOLE-1 Å-1

NUMBER OF CYCLES =
```

The program then lists the RMS derivative for the energy. Structures which are far from a minimum have a high RMS and could cause the program to fail, so these structures should be minimised using QUICK MIN before being entered into PD MIN. Occasionally during

the course of minimisation errors such as a floating point overflow or a floating zero divide may occur. These errors are flagged by the program and the use is informed - the minimsed structure can then be kept or abandoned.

Section 4.4.33.

QUICK MIN

QUICK MIN is based on a modified Hooke and Jeeves pattern search algorithm. It is useful for minimising structures which are far from a minimum energy conformation and may fail when minimised in PD MIN. Again there is a PDP 11 version and an array processor version, which have the same limitations as the PD MIN routines. The input and output for QUICK MIN are exactly the same as for PD MIN since it is called and written by the same subroutines. It does , however, differ in two ways. QUICK MIN requires the user to input a step length for its pattern search (standard values are normally between 0.1Å and 0.001Å) and it does not calculate an RMS derivative for the minimisation.

Section 4.4.34.

CONGEST

A modified version of the algorithm of Wipke and Gund[5,6] is used to predict the outcome of stereoselective attack at trigonal atoms. The names of the two trigonal atoms forming the double bond and the centre at which the attack is to take place must be given. The program then calculates and lists the total steric congestion for each side of the attack centre mean plane and the structure is redrawn so that the

attack centre mean plane and the plane of the screen are coincident, with +ve Z direction towards the viewer.

Section 4.4.35.

ZOOM IN / OUT

When examining large or crowded structures it is useful to be able to consider only a part of the structure at higher resolution. The ZOOM option allows the user to do this by entering an atom name and a sphere radius. The named atom is then redrawn at the centre of a sphere of the given radius - any other atoms within the sphere are also displayed. The atoms in this picture are stored in temporary arrays, modifications may be made to the picture, but when the user zooms out again the original structure is returned. The option may also be used to zoom in on specific fragments when in MULTIPLE mode.

Section 4.4.36.

LOCATE

As its name implies this option can be used to locate any atom in the current display structure. Again this facility is very useful when dealing with large moleclues. On entering an atom name the program locates the atom and causes its label to flash brightly making it immediately identifiable.

Section 4.4.37.

HARDCOPY

Selecting HARDCOPY relays a picture of the current display to the

Tektronix plotter which is attached to the system.

Section 4.4 38.

RESTART

RESTART simply resets the system to its original state by removing any current structure, setting the options to show hydrogens, cue on, single mode, labels on, redrawing the menu and limiting the crosshair cursor to the first five options.

Section 4.4.39.

QUIT

QUIT enables the user to exit tidily from the MOL system.

Chapter Five

Enzymes

Section 5.1.

Introduction

Enzymes are macromolecules which enable living organisms to actively mould and use their environment. They are specific proteins which catalyse the biological reactions which constitute metabolic pathways. As catalysts they are far more efficient than any man made compounds, increasing reaction rates by 10^6 to 10^{20}, at moderate temperatures and pressures, in aqueous solution and at physiological pH.[1,2] Indeed in the absence of enzymes the rates of most biological reactions would be too slow to be observed.

Enzymes are also more efficient than their synthetic counterparts in the area of substrate specificity.[3] Chemical catalysts rarely possess substrate specificity, whereas all enzymes are highly substrate specific and will catalyse a single chemical reaction, or a group of related chemical reactions.

The degree of substrate specificity shown by an enzyme arises from the three dimensional nature of its active site, which is where the reaction being catalysed occurs. The structure of the active site is in turn dependent upon both the amino acid sequence of the protein (the primary structure) and the way in which it is folded (secondary and tertiary structure). In general the active site of an enzyme takes the form of a cleft, or crevice, and some of the amino acid residues forming this cleft are involved in binding to the substrate. These

functional groups, which are required both for binding and for catalytic activity, are usually found in different areas of the linear sequence of amino acids and are brought together by the folding of the chain. For example the binding groups in the active site of lysozome are provided by residues 35, 52, 62, 63 and 101 from a linear sequence of 129 amino acids.

As well as being highly substrate specific many enzymes are also highly stereospecific, catalysing reactions involving say the L-form of a compound, but not the D-form. In such cases the isomer which is not a substrate often acts as an inhibitor to the reaction.

The function of many enzymes is dependent upon the presence of small amounts of chemical agents other than the reactants of the catalysed reaction. These compounds are called cofactors[4] and may simply be metal ions (Mg^{2+}, Fe^{2+} etc.)[5,6], more complex organic compounds,[7] or even other enzymes. Among these cofactors, some may be regarded as integral parts of the enzymes and are known as prosthetic groups. Others may act as substrates in that they combine with the enzyme and leave it again unchanged in the course of a single catalytic cycle - these are called coenzymes.

Section 5.2.

Enzyme Nomenclature

There exists a vast number of enzymes, each catalysing a specific reaction, whose activities cover a whole range of anabolic and catabolic reactions. In order to reduce the confusion and misunderstanding which arose from the trivial naming of these enzymes a Commission on Enzymes was set up in 1955 with the purpose of

devising a method of systematic nomenclature for enzymes. This produced a system in which enzymes are separated into six classes depending upon the type of reaction they catalyse, namely:-

Class	Enzymes (by reaction type)
1.	Oxidoreductases
2.	Transferases
3.	Hydrolases
4.	Lysases
5.	Isomerases
6.	Ligases

Each class is then divided into subclasses and sub-subclasses which further define the nature of the reaction being catalysed. Each enzyme is then given a unique number within its sub-subclass.

The application of this nomenclature system can be demonstrated with the enzyme Creatine kinase, a muscle enzyme which catalyses the reaction:-

Creatine + ATP → Creatine phosphate + ADP

This enzyme has the following unique number:-

EC 2. 7. 3. 2.

Enzyme Comission

Transferase Class

Phosphotransferase subclass

Sub-subclass of phosphotransferases with a nitrogenous group as an acceptor

Enzyme number within sub-subclass

Using this system of nomenclature each enzyme is given a unique number with which it can be identified. This system is flexible enough to allow the naming of newly identified enzymes and it also allows the introduction of new subclasses and sub-subclasses should they be required.

Section 5.3.

Enzyme Regulation

Since enzymes are such efficient biological catalysts a mechanism is often required by which their activity can be regulated. In some cases this is done by synthesising the enzyme in an inactive precursor form which can then be activated at the physiologically appropriate time and place.[8,9]

Another regulatory mechanism involving structural change is that of covalent modification.[10] With this control mechanism a small group is covalently inserted into the enzyme rendering it inactive. This modification can be reversed by hydrolysis.

Other enzymes are under physiological control. For example it has been observed that hormones can exert physiological effects on certain enzymes by altering the specificity of the enzyme.[11]

However the method of control which is probably the most common in metabolic pathways involves inhibition of the enzyme by one or more of the products of the pathway in which it is involved. This type of enzyme regulation is called feedback inhibition.[12-15]

Section 5.4.

Enzymes as Catalysts

For any reversible chemical reaction:-

$$S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} P$$

the rate of the reaction, v, can be given by the rate of appearance of product per unit time, or the rate of disappearance of substrate per unit time:-

$$v = -\frac{d[S]}{dt} = \frac{d[P]}{dt}$$

and once equilibrium has been achieved the equilibrium constant, K, is given by:-

$$K = \frac{[P]}{[S]} = \frac{k_1}{k_2}$$

The variation of the equilibrium constant, K, with temperature can be deduced from classical thermodynamics:-

$$\ln K = -\frac{\Delta G^{\circ}}{RT}$$

where ΔG° is Gibb's free energy for the reaction, R is the gas constant and T is the temperature in degrees Kelvin.

From this can be derived the equation for the rate constant, k.

$$\ln k = -\frac{Ea}{RT} + \text{constant}$$

Another important equation is:-

$$k = Ae^{-(E_a / RT)}$$

where A is a frequency factor and the term $e^{-(E_a / RT)}$ is the Boltzmann expression for the fraction of molecules having an energy in excess of E_a where E_a is the energy of activation for the reaction i.e. the minimum energy which must be acquired by the molecules before the reaction can take place.[16]

The energy involved in the reaction can be plotted as a function of Gibb's free energy versus progress of reaction, see Figure 5.1.

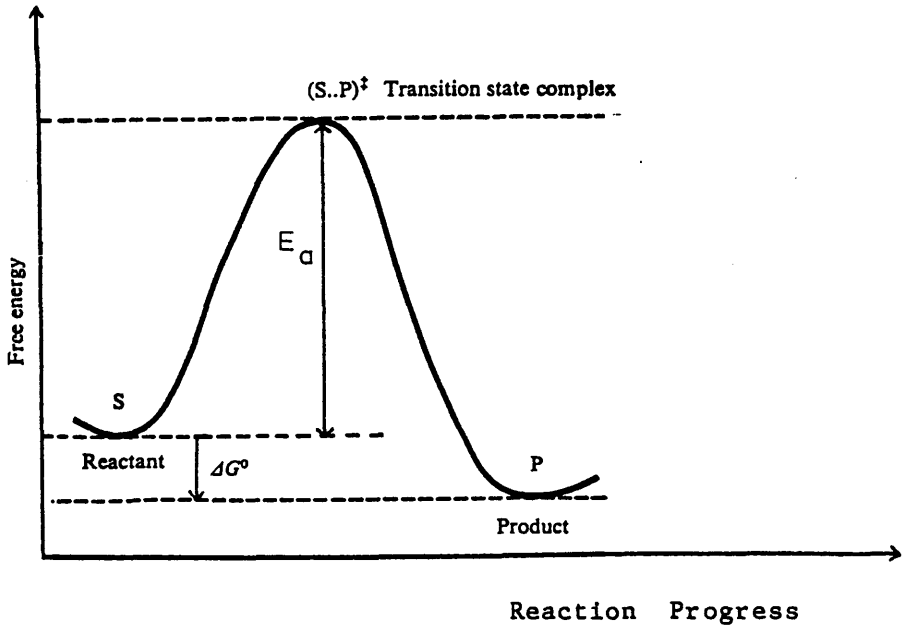


Figure 5.1

From the above figure it can be seen that the reaction proceeds via a transition state and for the reaction to occur the molecule must achieve that energy of activation associated with the transition state.

The reaction rate is proportional to the fraction of molecules having a free energy greater than, or equal to the energy of activation[17] (the proportion of molecules in this position increases with temperature). Consequently if the E_a is very large the reaction

rate will be slow and if E_a is small the reaction rate will be fast.

Catalysts do not alter the position of equilibrium for a reaction, but instead they speed up the rate of attainment of this equilibrium by lowering the activation for the reaction, see Figure 5.2.

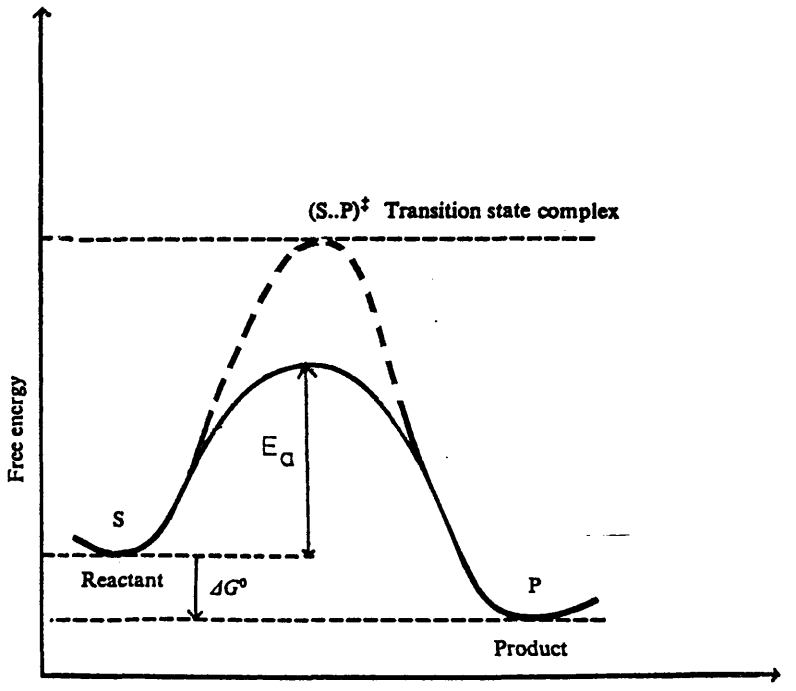


Figure 5.2

Therefore catalysts create a new reaction pathway whose transition state energy is lower than it would be in the absence of the catalyst.

Enzymes act as catalysts, however Figure 5.3 more accurately represents the changes in energy which occur when an enzyme reacts with a single substrate.

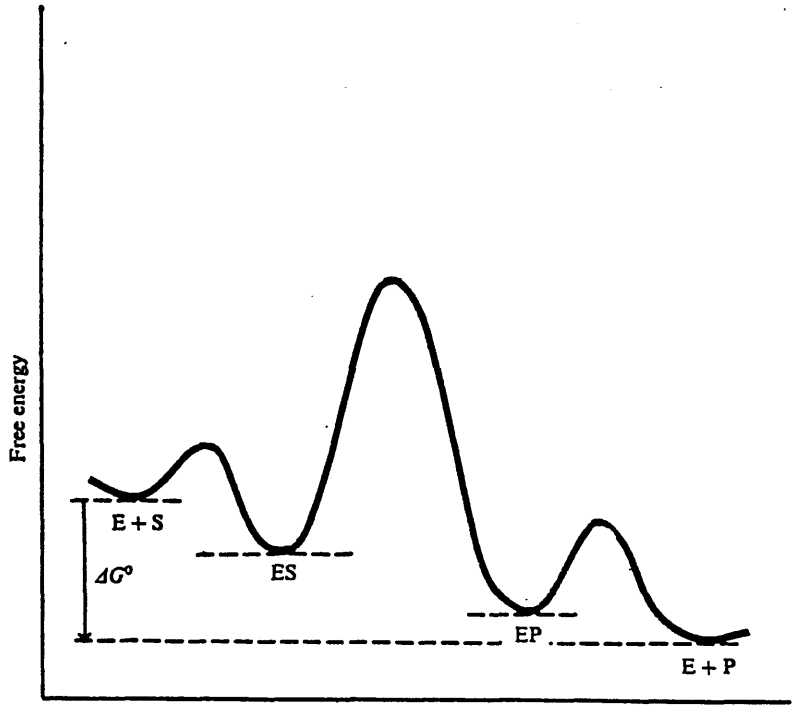


Figure 5.3

With enzymes the reaction rate can be increased in many ways. For example enzymes bind substrates in their optimal positions for reacting.[18,19] This can include a) the precise orientation of binding orbitals to give maximum chance of reaction,[20] b) increasing the effective concentration of the substrates by binding the reacting molecules in close proximity,[21] c) binding the substrate in a strained conformation which is more likely to react.[22]

Section 5.5.

Enzyme Kinetics

The study of the mechanism of enzyme action can be carried out using various techniques. These include X-ray crystallography, electron microscopy, amino acid sequencing, substrate modification and molecular modeling. However the first technique that is usually used is enzyme kinetics.

There is no single kinetic expression which can be used to describe all enzymic reactions. This is due to the many different mechanisms involved which include single substrate reactions, more than one substrate reactions, reactions involving more than one intermediate step and reactions involving allosteric enzymes.

Three mentioned enzyme classes fall into the catagory of single substrate reactions, they are:-

1) Isomerases which catalyse $A \rightarrow B$

2) Lysases which catalyse $A \rightarrow B + C$

3) Hydrolases which catalyse $B - A + H_2O \rightarrow A\text{-}OH + B\text{-}H$

Although hydrolases are only honorary single substrate reactions due to the the fact that water is present at such high concentration that the reaction is presumed to be independent of $[H_2O]$.

Kinetics of single - substrate reactions

Initially it was thought that enzymic reactions could be expressed in terms of the equation:-

$$E + S \xrightarrow{k_3} E + P$$

The rate of this reaction would be proportional to the concentration of each of the reactants E and S.

$$\text{reaction rate} = -\frac{d[S]}{dt} = -\frac{d[E]}{dt} = k_3 [S] [E]$$

and a plot of initial rate against substrate concentration would give a straight line (see Figure 5.4).

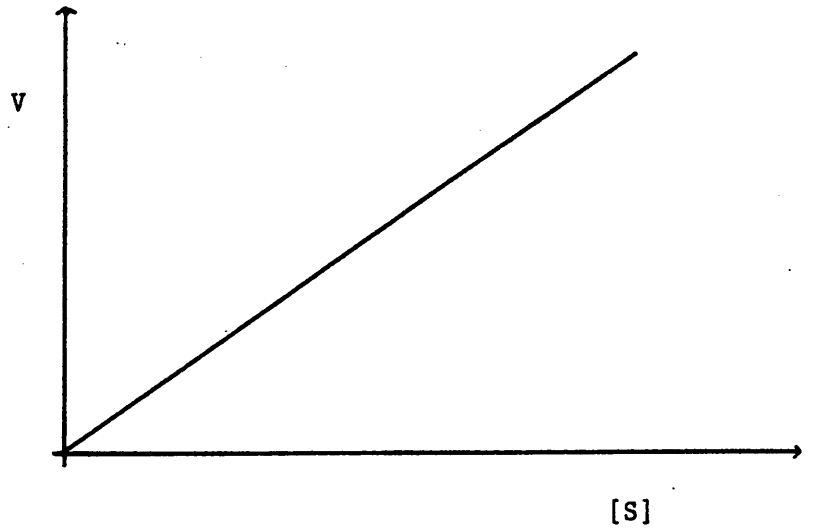


Fig. 5.4

However, it was found that for many enzymes a plot of the reaction rate versus substrate concentration for a set amount of enzyme gives a rectangular hyperbola (see Fig 5.5).

This discrepancy was explained by A.J. Brown in 1902[23] when he proposed the formation of a discrete Enzyme-substrate complex as part of the reaction mechanism:-

$$E + S \rightarrow E S \rightarrow E + P$$

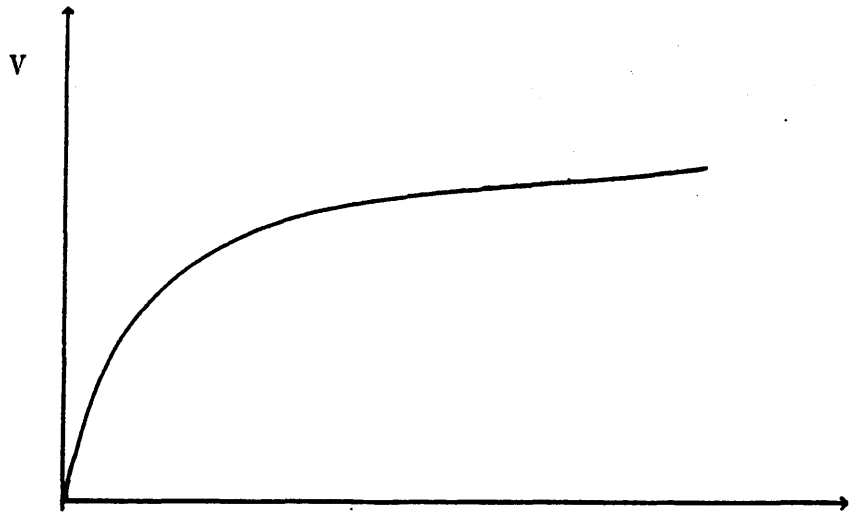


Fig. 5.5

From the above equation the substrate concentration / reaction rate relatioship can now be rationally explained. At small substrate concentrations the reaction rate is almost linearly proportional to the substrate concentration, but at high substrate concentration the enzyme becomes fully saturated and the reaction rate reaches its maximal value.

In 1913 Michaelis and Menten[24] proposed a simple model to explain the kinetic properties of single substrate enzyme systems:-

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E\,S \xrightarrow{k_3} E + P \qquad 5.1$$

Free enzyme, E, and substrate, S, are in rapid equilibrium with the enzyme substrate complex, E S, which can also react to give product, P. Since enzymes are highly substrate specific and P is initially at low concentrations it is assumed that none of the product can be converted back into enzyme substrate complex.

If k_3 is the slow, or rate determining step, the rate of the

reaction is equal to the rate of appearance of product:-

$$v = \frac{d[P]}{dt} = k_3 [ES] \qquad 5.2$$

The concentration of the ES complex is in turn dependent on the rate of formation and dissociation of the ES complex:-

$$\text{rate of formation of ES} = k_1 [E] [S] \qquad 5.3$$

$$\text{rate of breakdown of ES} = k_2 [ES] + k_3 [ES] \qquad 5.4$$

At equilibrium the rate of formation of the ES complex will equal the rate of breakdown, so the concentration of ES will be constant:-

$$\frac{d[ES]}{dt} = k_1 [E] [S] - k_2 [ES] - k_3 [ES] = 0 \qquad 5.5$$

therefore

$$k_1 [E] [S] = k_2 [ES] + k_3 [ES] \qquad 5.6$$

The concentration of free enzyme [E] is given by:-

$$[E] = [E_{total}] - [ES] \qquad 5.7$$

substitution and rearrangement in equation 5.6 then gives:-

$$[ES] = \frac{k_1 [S] [E_{total}]}{(k_2 + k_3) + k_1 [S]} \qquad 5.8$$

and from equation 5.2 the reaction rate then becomes:-

$$v = \frac{k_3 k_1 [S] [E_{total}]}{(k_2 + k_3) + k_1 [S]} \qquad 5.9$$

if K_m is the defined as $\frac{k_2 + k_3}{k_1}$

the reaction rate can be expressed as:-

$$v = \frac{k_3 \ [S] \ [E_{total}]}{K_m + [S]} \qquad 5.11$$

The maximal rate of reaction, V_{max}, occurs when the enzyme is fully saturated with substrate i.e. when [S] is high (see Fig.5.5). At this point [S] / (K_m + [S]) approaches unity and this gives:-

$$V_{max} = k_3 \ [E_{total}] \qquad 5.12$$

substituting this into equation 5.11 gives the Michaelis - Menten equation:-

$$v = V_{max} \ \frac{[S]}{K_m + [S]} \qquad 5.13$$

From the Michaelis - Menten equation it can be seen that for low substrate concentration, i.e. when [S] << K_m :-

$$v = V_{max} \ \frac{[S]}{K_m} \qquad 5.14$$

that is to say the rate is directly proportional to substrate concentration . For high substrate concentration, where [S] >> K_m the rate becomes maximal and therefore independent of the substrate concentration. Finally where [S] = K_m :-

$$v = \frac{V_{max}}{2} \qquad 5.15$$

Thus Km is equal to that concentration at which the reaction rate is half its maximal value.

The equation of a rectangular hyperbola is not a convenient form

for evaluating the kinetic constants K_m and V_{max}. However the Michaelis - Menten equation can be transformed into an equation which gives a straight line using a procedure designed by Lineweaver and Burk:[25]

$$\frac{1}{V} = \frac{1}{V_{max}} + \frac{K_m}{V_{max}} \cdot \frac{1}{[S]}$$

A plot of 1 / V versus 1 / [S] will give a straight line with a gradient K_m / V_{max} and an intercept on the y axis of 1 / V_{max}, see Fig. 5.6.

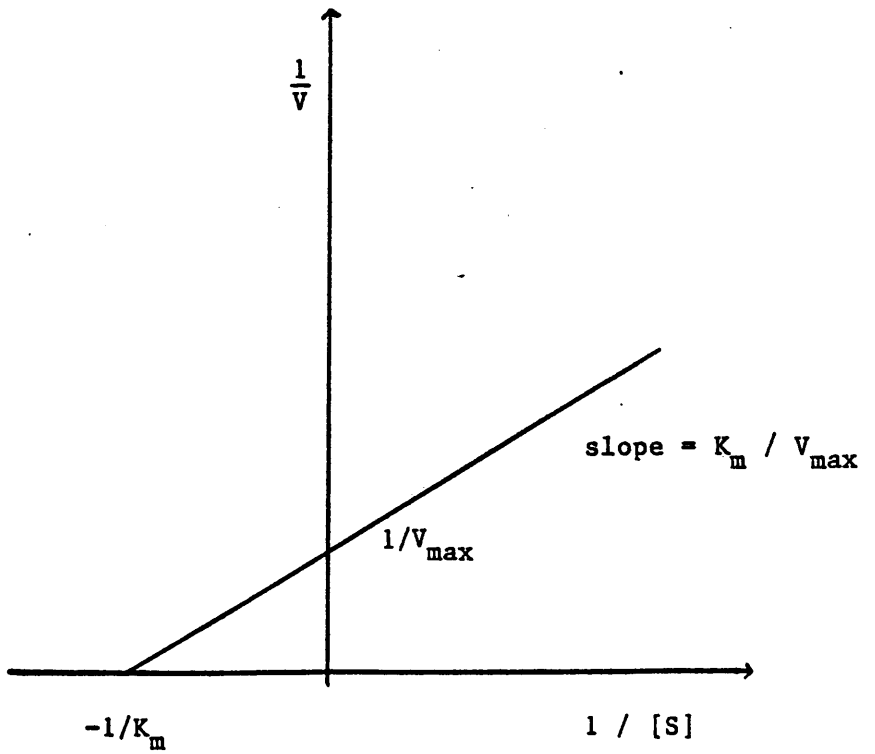


Figure 5.6

For those enzymes following Michaelis - Menten kinetics the K_m value for the enzyme is dependent upon the nature of the substrate and also on reaction conditions, like temperature, pH etc. In general K_m gives two pieces of information about the system. Firstly it is the

concentration at which half of the enzyme's active sites are filled. From this we can calculate the fraction of active sites filled, f_{ES}, for any particular substrate concentration using the equation:-

$$f_{ES} = \frac{V}{V_{max}} = \frac{[S]}{[S] + K_m} \qquad 15.17$$

Secondly K_m is related to the individual rate constants for the reaction:-

$$K_m = \frac{k_2 + k_3}{k_1} \qquad 5.18$$

The apparent dissociation constant for the ES complex is:-

$$K_d = \frac{[E]\ [S]}{[ES]} \qquad 5.19$$

For the circumstance where the formation of product is the rate determining step (which was assumed for the derivation of the Michaelis - Menten equation) k_2 will be much larger than k_3. This gives:-

$$K_m \approx \frac{k_2}{k_1} = K_d \qquad 5.20$$

Thus K_m is approximately equal to K_d when $k_2 >> k_3$. Under these conditions K_m is a measure of the strength of the ES complex, a large K_m indicating weak binding and a small K_m indicating strong binding. It must be noted that the condition $k_2 >> k_3$ does not hold for all single substrate systems.

The maximal rate of reaction, V_{max}, gives the turnover number for the reaction where the concentration of the enzyme is known:-

$$V_{max} = k_3 [E_{total}] \quad 5.21$$

where k_3 is the turnover number , that is the number of substrate molecules converted into product per unit time at full saturation of the enzyme. Turnover numbers are normally in the range 1 - 10,000.

In vivo conditions are rarely the same as in vitro conditions. Most enzymes are not normally fully saturated under physiological conditions and the [S] / K_m ratio is typically between 0.01 and 1.0. This means that the enzymic turnover number in the cell is less than its maximum value, k_3, due to the fact that the active sites of the enzyme molecules are only partially occupied. The kinetics of the enzyme under such conditions can be given by:-

$$V = \frac{k_3}{K_m} [E] [S] \quad 5.22$$

when [S] << K_m [E] is nearly $[E_{total}]$ therefore:-

$$V = \frac{k_3}{K_m} [E_{total}] [S] \quad 5.23$$

The reaction rate is therefore dependent upon k_3 / K_m and [S], substituting for K_m gives:-

$$\frac{k_3}{K_m} = \frac{k_3 \ k_1}{k_2 + k_3} \quad 5.24$$

The reaction rate is therefore set by k_1, the rate of formation of the [ES] complex. The value of k_1 is limited by the rate of diffusion since the rate of formation of the ES complex cannot occur any faster than the diffusion controlled substrate - enzyme encounters. Therefore the maximum value of k_1 as set by diffusion is 10^8 - 10^9 mol^{-1} sec^{-1}. The upper limit upon k_3 / K_m is therefore

between 10^8 and 10^9 mol^{-1} sec^{-1}.

Several enzymes do not obey Michaelis - Menten kinetics due to the fact that their reaction pathways are more complicated than the above example of a single substrate system for which the rate determining step is that of formation of product. For such enzymes the maximal catalytic rate at full substrate saturation is denoted by k_{cat} which may be dependent upon several other rate constants eg, for the three step mechanism:-

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_2}{\underset{P_1}{\swarrow}} ES' \xrightarrow{k_3} E + P_2 \qquad 5.25$$

Assuming that k_2 and k_3 are the rate determining steps.

$$V = k_{cat} [E_{total}] \qquad 5.26$$

where

$$k_{cat} = \frac{k_2 \ k_3}{k_2 + k_3} \qquad 5.27$$

Under such conditions it is impossible to calculate the individual rate constants, but k_{cat} can be readily calculated from equation 5.25 and is in fact equivalent to the turnover number for the reaction. Since K_m gives a measure of how well a substrate binds to the enzyme the k_{cat} / K_m ratio gives a good indication of the effectiveness of the substrate. This ratio is known as the specificity constant and the higher the value of the specificity constant the better is the substrate for the enzyme.

Section 5.6.

Enzyme - Substrate Complexes

The existence of enzyme substrate complexes, first proposed by Brown[23] in 1902, has been shown in a variety of ways such as electron microscopy, X-ray crystallography, changes in physical properties such as heat stability, spectroscopic studies and isolation of pure forms of ES complexes.

The specificity of binding to a substrate will be determined by the shape of the active site. A large number of enzyme - substrate interactions can be envisaged in terms of a lock and key arrangement, for which the the shape of the substrate must match the shape of the active site. With some enzymes, however, the active site may not be totally rigid and in such cases the active site shape only matches the substrate shape, once the substrate has been bound. This is known as induced fit.

The structures of active sites are of prime importance to the pharmaceutical industry since many drugs and toxic compounds inhibit enzymes by binding to the active site, or binding elsewhere on the enzyme causing conformational changes which alter the structure of the active site (allosteric inhibition).

Section 5.7.1.

Enzyme Inhibition

Enzymes can be inactivated by many compounds which will disrupt protein structure (eg. acid, alkali etc.). However there are many selective compounds which will inhibit enzymes by binding at strategic sites without actually disrupting the proteins three dimensional

structure.

Within this category are two different kinds of inhibitor: those which are responsible for reversible inhibition[26] amd those which are responsible for irreversible inhibition[27].

With irreversible inhibition the inhibitor is bound covalently, or else bound so tightly to the enzyme that its dissociation is extremely slow. It is characterised by complete loss of enzyme activity which cannot be regained by physical treatment and is generally quantified in terms of the loss of enzyme activity, or velocity of inhibition.

With reversible inhibition the activity of the enzyme can be regained by the removal of the inhibitor. This is possible due to the rapid equilibrium which exists between the free enzyme, free inhibitor and the enzyme-inhibitor complex. Reversible inhibition can be separated into different types, namely competitive inhibition, non-competitive inhibition, uncompetitive inhibition and mixed inhibition. These types of inhibition can be distinguished kinetically and the type of inhibition involved with EC 3.4.24.11., namely competitive inhibition, is described in more detail in the following section.

Section 5.7.2.

Competitive Inhibition

With competitive inhibition the inhibitor often resembles the natural substrate, or more effectively the substrate transition state. The inhibitor and the substrate are therefore in direct competition

for the same binding site in the molecule. The rate of catalysis is reduced because the binding of a competitive inhibitor decreases the proportion of available active sites.

The simplest mechanism for competitive inhibition can be described by the following equations:

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P$$

$$E + I \underset{k_5}{\overset{k_4}{\rightleftharpoons}} EI$$

$$K_i = \frac{[E]\,[I]}{[EI]} = \frac{k_5}{k_4} \qquad 5.28$$

where [I] is the concentration of inhibitor and K_i is the dissocia-dissociation constant of the enzyme-inhibitor complex.

The reaction rate in the presence of a competitive inhibitor can then be given by:

$$V = \frac{V_{max}\,[S]}{K_m\,(1 + [I]/K_i) + [S]} \qquad 5.29$$

Inversion of this equation gives a Lineweaver-Burk type equation of the form:

$$\frac{1}{V} = \frac{1}{V_{max}} + \frac{K_m\,(1 + [I])\,1}{V_{max}\,K_i\,[S]}$$

A plot of 1/V versus 1/[S] can be compared to the normal plot (see Fig. 5.7). V_{max} is given by the intercept on the ordinal axis and is unaltered by competitive inhibition. However, the gradient is increased by a factor of $(1 + [I]/K_i)$ and K_m is also altered. Competitive inhibition can be overcome at large substrate concentrations

where [S] >> [I] since under these conditions virtually all the active sites are occupied by substrate. The increase in the gradient gives an indication of the affinity with which the inhibitor is binding to the enzyme.

Inhibition constants can be calculated in a similar manner for all other types of reversible inhibition.

COMPETITIVE INHIBITION

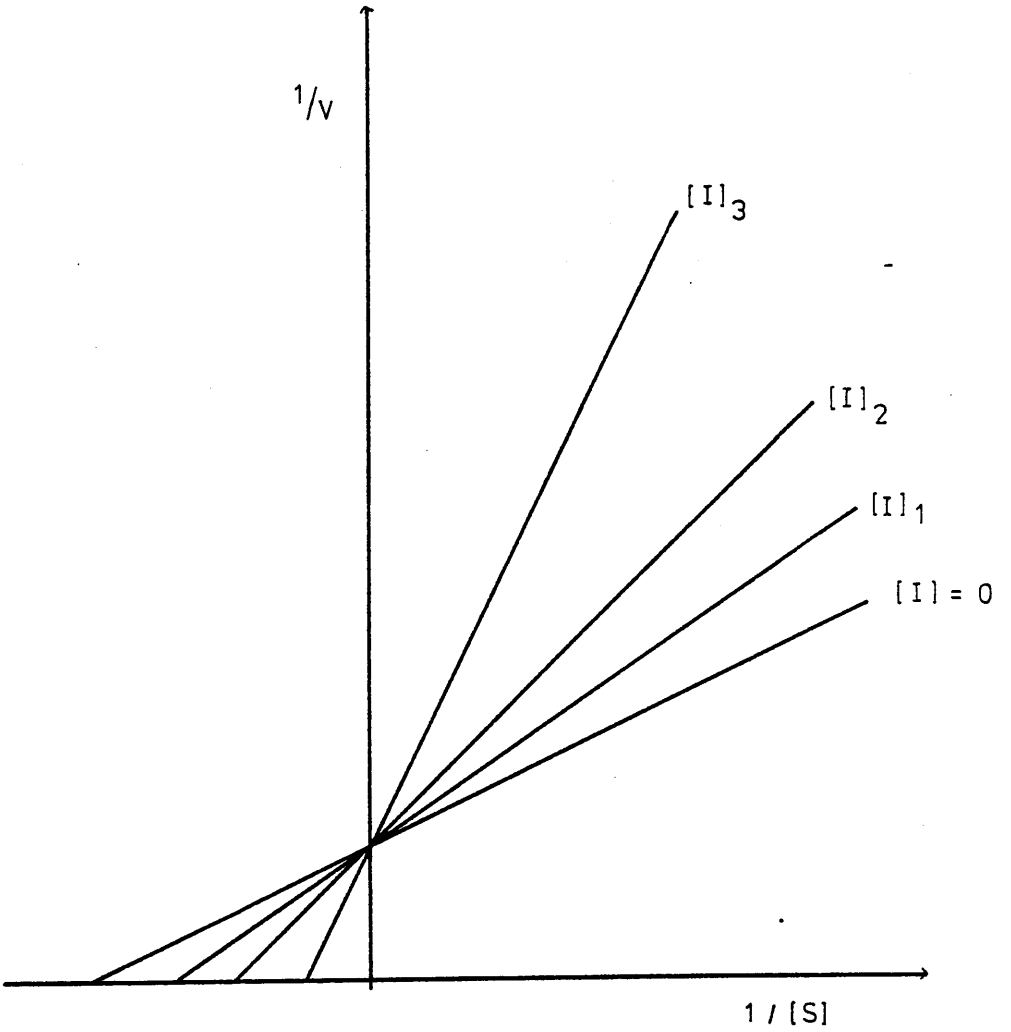


Fig. 5.7.

Section 5.7.3.

Determination of K_i's

K_i values provide a useful measure for comparing the relative potency of different inhibitors. The simplest way of measuring a K_i is to measure the initial reaction rate with various substrate concentrations and then remeasure the rates at the same concentrations with a known concentration of inhibitor. Depending upon the type of inhibition taking place the gradient of the resulting Lineweaver - Burk plots may be affected by a value of $(1 + [I] / K_i)$ and since [I] is known K_i can be calculated. This method would not however detect non-linear inhibition and it is better to repeat the above process with several different inhibitor concentrations. K_i can then be obtained graphically in one of two ways.

a) By plotting slope versus [I] for competitive inhibition, intercepts versus [I] for uncompetitive inhibition, slope, or intercepts vs [I] for non-competitive inhibition and both slope versus [I] (gives K_{ie}) and intercepts versus [I] (gives K_{ies}) for mixed inhibition. All the above plots should result in a straight line plot with an intercept on the abscissal axis equal to $-K_i$, see Fig. 5. 8.

[28]

b) Dixons method can be used for competitive and non - competitive inhibition this involves plotting 1 / V versus [I] for different substrate concentrations, see Fig. 5. 9. K_i can be determined from the intercept of the lines.

The amount of inhibition occuring for any inhibitor is often reported in terms of an IC_{50} value. This is the concentration of an unlabeled inhibitor which causes 50% inhibition of the binding of a labeled substrate and this can be converted into a K_i using the Cheng

Inhibition Constant Determination using Lineweaver-Burk plots.

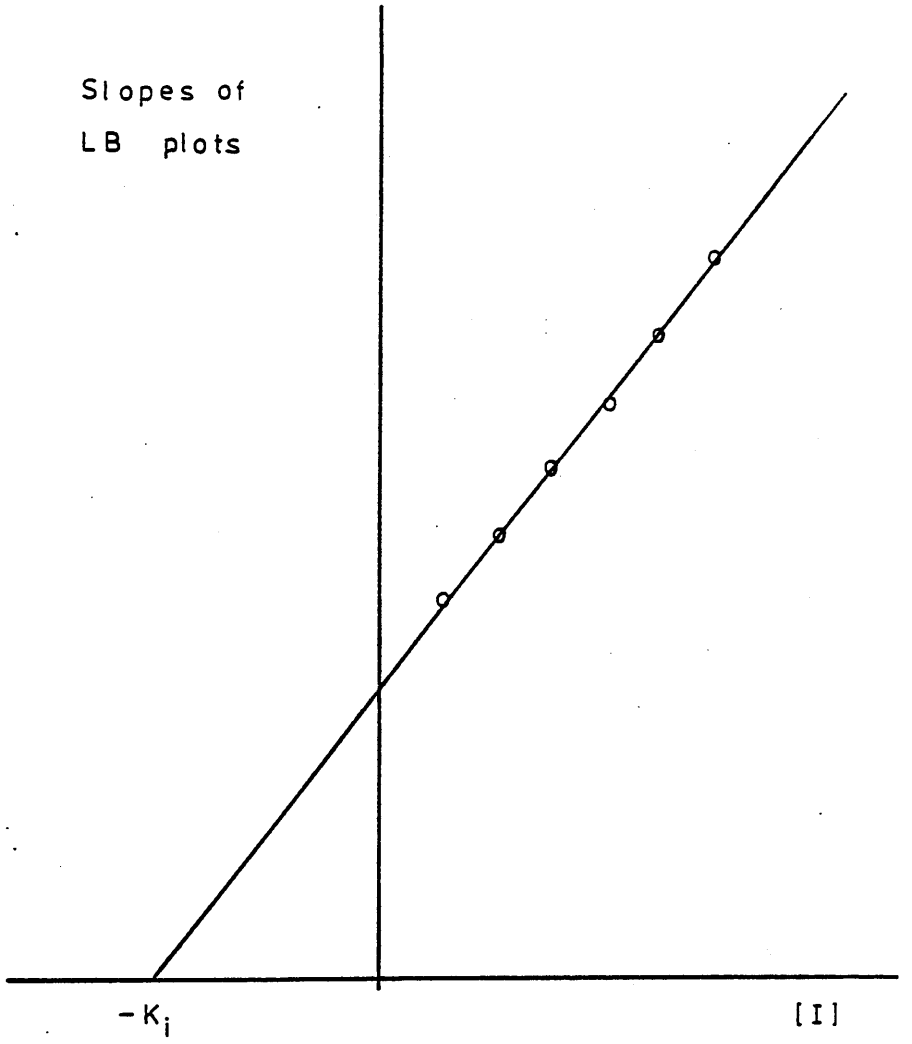


Fig. 5.8.

Inhibition Constant Determination using Dixon's method.

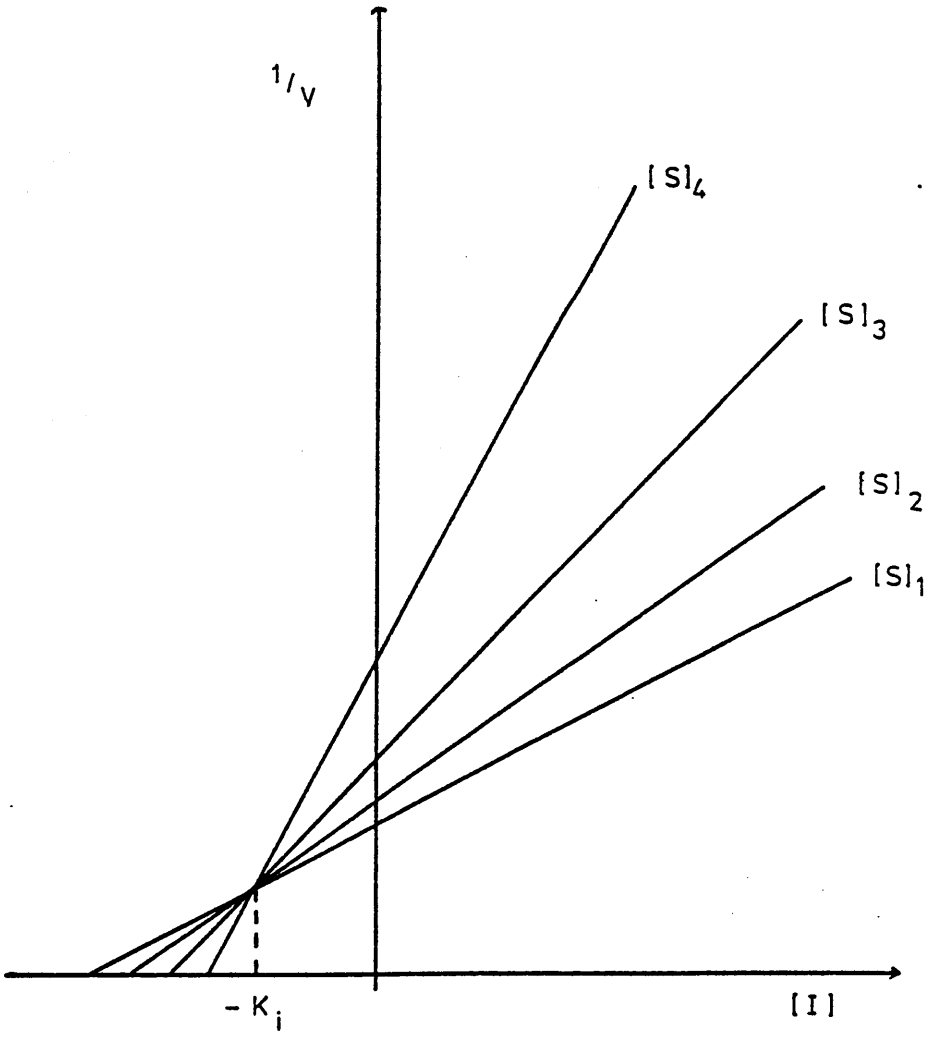


Fig. 5.9.

- Prusoff[29] equation:-

$$K_i = IC_{50} / (1 + D / K_d)$$

where D is the concentration of the labeled substrate and K_d is its dissociation constant.

Section 5.8.

Protein Evolution

From the study of evolution it can be seen that there has been much diversification from the original gene pool giving rise to the vast number of existing species. The evolution of proteins occured in conjunction with the evolution of species and protein diversity is a direct result of this evolution. The most basic evolutionary change occuring in protein structures is the substitution of amino acid residues. Larger structural changes can be brought about by insertion of one or more amino acid residues and major structural modifications occur as a result of gene fusion and duplication. Such structural changes accumulate over time resulting in major differences between the initial and final structures. This causes great difficulties for sequence comparisons of distantly related proteins.[30]

The amino acid sequence of a protein is coded for by certain regions of the DNA of the organism. In general only about 10% of the DNA actually codes for proteins[32,33] - the remainder takes the form of spacer DNA separating the coding regions. Mutations occur with fairly high frequency in DNA and changes in the amino acid sequence can occur as a direct result of mutations in the coding regions. However, due to the degeneracy of the genetic code only about 75% of mutations

occuring in the structural zones actually give rise to changes to amino acid residues. The remaining 25% are silent mutations which do not produce any structural change to the protein being coded for.

Studies of protein evolution can provide an invaluable source of knowledge for analysing biological and evolutionary relationships. Protein evolution can be split into two distinct categories, protein speciation and protein differentiation.

Section 5.8.2.

Protein Speciation

Protein speciation involves the evolution of homologous proteins within different organisms, having a common function. The evolution of such proteins involves a process of random mutation and selection.

In general the amino acid sequence and chainfold are well enough conserved to make direct comparison of the proteins of genetically different organisms. This has shown that the function of the protein is conserved by the natural selection of structures for which the residues involved in the function of the protein are not altered too drastically. In fact it has been shown for families of related enzymes changes in the residues at the protein surface occur much more frequently than changes in internal residues (except where the surface is involved in the active site). This is favoured since internal residues play a more important role in stabilising the folding of the protein and also in the function of the protein. It has also been shown that the secondary structure of the protein tends to be conserved, changes in internal residues which would cause instability are often compensated for by changes in neighbouring residues and

substitution between residues of similar type is found to occur more frequently than with residues possessing different functionalities.

The genetic, or evolutionary distance between two proteins can be measured in terms of the % of accepted point mutations (PAM) per codon per 10^8 years and for human DNA this is of the order of 45% - 50% substitution of codons per 10^8 years. Another unit occasionally used is the unit evolutionary period which is defined as the number of years for one accepted codon change to occur per 100 residues.

Section 5.8.3.

Protein Differentiation

Protein differentiation involves the functional diversification of homologous proteins within say one organism and reflects the evolution of biochemical pathways.

Unlike speciation the function of the protein is not conserved since differentiation involves the adaptation of an enzyme to perform a new function. In general this involves the substitution of one or more amino acid residues which gives rise to a change in substrate specificity.[34] A good example of protein differentiation is the heritable resistance to antibiotics found in bacteria.

Section 5.8.4.

Evolutionary Convergence

At first glance it would seem that the number of possible amino acid sequences and chainfolds is vast and it would be reasonable to assume that proteins which have similar functions are likely to have evolved from a common origin.

However, one factor which could limit the range of protein structures would be selection of favoured types of fold (eg. helices, β sheets) which would impart greater stability on the protein structure.

This view is however too simplistic to explain the discovery of a number of structurally unrelated proteins which have evolved common functions. The active sites of these proteins involve different amino acid residues, however they all provide the optimum binding and catalytic conditions for the substrates.[35] This is done by having similar active site architectures, such that chemically important groups show equivalent functionality and spatial arrangement.

Such evolutionary convergence on function is very common[36] and can be expressed as a natural selection of those proteins whose catalytic sites fulfil all the functional requirements for the catalysis of a specific reaction. Such proteins are analogous rather than homologous. Even proteins derived from a common ancestor can display both convergence and divergence simultaneously and when comparing two proteins which have similar functions it can be difficult to tell the difference between homologous and analogous features.

Chapter Six

Section 6.1

Investigations into the actions of analgesic drugs such as morphine and pethidine led to the discovery of opiate receptors in the CNS[1]. Figure 6.1 shows how these receptors are distributed in the brain and correlates this with the effects of opiate administration[2].

Further binding studies have identified at least three different opiate receptors each having a unique specificity[3]. The exact physiological significance of these multiple receptors is not yet fully understood, but it is known that they have different distributions in various regions of the brain.

Although many detailed studies have been carried out on opiate action the mechanism by which analgesia occurs remains unclear. However, it is known that the binding of opiates to the receptors in some way supresses the transmission of signals across the synapses[4,5]

After the discovery of opiate receptors in the brain it was postulated that compounds should exist within the body which could bind to these receptors. This led to the discovery of a large number of naturally occurring opioid peptides (see Figure 6.2)[6], perhaps the most important of which were [Leu]- and [Met]-enkephalin[7,8]. These naturally occurring opiates were found to be released in response to stressful and painful stimuli, giving rise to a weak analgesia. However, the analgesia produced was short-lived since the enkephalins were rapidly broken down *in vivo*[9-12] by several different

Opioid effect	Location of opioid receptors
Analgesia	
Spinal (body)	Laminae I and II of dorsal horn
Trigeminal (face)	Substantia gelatinosa of trigeminal nerve
Supraspinal	Periaqueductal grey matter, medial thalamic nuclei, intralaminar thalamic nuclei, ? striatum
Autonomic reflexes	
Suppression of cough, orthostatic hypotension, inhibition of gastric secretion	Nuclei tractus solitarius, commisuralis, ambiguous, and locus coeruleus
Respiratory depression	Nucleus tractus solitarius, parabrachial nuclei
Nausea and vomiting	Area postrema
Meiosis	Superior colliculus, pretectal nuclei
Endocrine effects	
Inhibition of vasopressin secretion	Posterior pituitary
Hormonal effects	Hypothalamic infundibulum, hypothalamic nuclei, accessory optic system, ? amygdala
Behavioural and mood effects	Amygdala, nucleus stria terminalis, hippocampus, cortex, medial thalamic nuclei, nucleus accumbens, ? basal ganglia
Motor rigidity	Striatum

Location of opioid receptors proposed to mediate specific opioid effects

Figure 6.1

OPIOD PEPTIDES

[Leu]enkephalin	Tyr-Gly-Gly-Phe-Leu
[Met]enkephalin	Tyr-Gly-Gly-Phe-Met
[Met]enkephalinyl-Arg-Phe	Tyr-Gly-Gly-Phe-Met-Arg-Phe
[Met]enkephalinyl-Arg-Gly-Leu	Tyr-Gly-Gly-Phe-Met-Arg-Gly-Leu
Peptide E	Tyr-Gly-Gly-Phe-Met-Arg-Arg-Val-Gly-Arg-Pro-Glu-Trp-Trp-Met-Asp-Tyr-Gln-Lys-Arg-Tyr-Gly-Gly-Phe-Leu 1 5 10 15 20 25
Dynorphin	Tyr-Gly-Gly-Phe-Leu-Arg-Arg-Ile-Arg-Pro-Lys-Leu-Lys-Trp-Asp-Asn-Gln 1 5 10 15 17
α-Neoendorphin	Tyr-Gly-Gly-Phe-Leu-Arg-Lys-Tyr-Pro-Lys 1 5 10
β-Neoendorphin	Tyr-Gly-Gly-Phe-Leu-Arg-Lys-Tyr-Pro 1 5 9
PH-8P	Tyr-Gly-Gly-Phe-Leu-Arg-Arg-Ile 1 5 8
β_h-Endorphin	Tyr-Gly-Gly-Phe-Met-Thr-Ser-Glu-Lys-Ser-Gln-Thr-Pro-Leu-Val-Thr-Leu-Phe-Lys-Asn-Ala-Ile-Ile-Lys-Asn-Ala-Tyr-Lys-Lys-Gly-Glu 1 5 10 15 20 25 30 31
α-Endorphin	1-16 sequence
γ-Endorphin	1-17 sequence
δ-Endorphin	1-27 sequence
β-Casomorphin-7	Tyr-Pro-Phe-Pro-Gly-Pro-Ile
β-Casomorphin-5	Tyr-Pro-Phe-Pro-Gly
Kyotorphin	Tyr-Arg

Figure 6.2

enzymes[13-16]. It was therefore apparent that inhibition of this degradation would give rise to a prolonged and perhaps increased analgesia and might therefore provide new analgesic compounds which would be free from the side effects associated with the administration of opiates, eg, dependency.

At least three enzymes have been shown to be involved in the inactivation of opioid peptides[17]. These are angiotensin converting enzyme (ACE), a neutral endopeptidase (EC 3.4.24.11.) and an amino peptidase which has not yet been fully characterised. These enzymes appear to complement each other's actions, not only in the inactivation of opioid peptides, but also in the degradation of neuropeptides (see Figure 6.3)[18].

As a target for drug action an enzyme requires to be well characterised in terms of function, substrate specificity and role in metabolism. This is the case for both ACE and the neutral endopeptidase. Enzyme kinetic studies have shown that *in vitro* both ACE and EC 3.4.24.11. seem to be implicated in the early stages of enkephalin activation and it would therefore appear that both these enzymes could be useful targets[19]. However, ACE is unsuitable for two reasons. Firstly, it is an exopeptidase which plays an important role in blood pressure regulation[20] by generating the potent vasoconstrictor, Angiotensin II, from Angiotensin I and by the degradation of the vasodilator, bradykinin. Secondly, *in vivo* studies have shown that inhibition of ACE makes very little difference to the striatial levels of enkephalins[17].

It appears that EC 3.4.24.11. plays a more major role in enkephalin inactivation *in vivo*. This may be due to differences in

Compound	Site of Cleavage
[Met] - Enkephalin	Tyr - Gly - Gly - Phe - Met (AP ↓ Gly-Gly; ACE EP ↓↓ Gly-Phe)
[Leu] - Enkephalin	Tyr - Gly - Gly - Phe - Leu (AP ↓ Gly-Gly; ACE EP ↓↓ Gly-Phe)
Bradykinin	Arg - Pro - Pro - Gly - Phe - Ser - Pro - Phe - Arg (EP ↓ Gly-Phe; ACE EP ↓↓ Pro-Phe)
Neurotensin	pGlu - Leu - Tyr - Glu - Asn - Lys - Pro - Arg - Pro - Tyr - Ile - Leu (ACE ↓ Pro-Arg; EP ↓ Arg-Pro; EP ↓ Tyr-Ile)

ACE - Angiotensin converting enzyme

AP - Aminopeptidase (as yet uncharacterised)

EP - Endopeptidase 24.11.

Figure 6.3

substrate specificity between the two enzymes; however, EC 3.4.24.11. does not have a high specificity for either [Leu]- or [Met]-enkephalin[21]. A more likely explanation is that the distribution of EC 3.4.24.11. closely parallels that of the opiate receptors[15,22] which ACE does not, ie, EC 3.4.24.11. is more highly localised in those regions where opiod peptides will be found.

Therefore EC 3.4.24.11. has become a target in the search for novel analgesics and several inhibitors of this enzyme have been shown to possess analgesic properties[23-25].

Section 6.2.

Endopeptidase 24.11.

EC 3.4.24.11. is a neutral endopeptidase first isolated from the brush border of rabbit kidney[26] and thought to be one of the principal enzymes involved in the degradation of naturally occurring, biologically active peptides. The enzyme has a substrate specificity very similar to that of the bacterial enzyme Thermolysin in that it specifically cleaves peptide bonds for which the amino group has been provided by a hydrophobic residue[21,27,28] such as phenylalanine, valine, leucine, etc.

Enzymes with an identical substrate specificity have been isolated from a range of tissues including brain[29], pituitary[30], lung[31] and spleen[31] and a similar enzyme has also been identified at high concentrations in the serum of patients with sarcoidosis[31].

The majority of experiments have been carried out on enzyme isolated from brain and kidney due to their relative abundance.

However, there is probably more interest in the enzyme isolated from the brain where it is thought to be involved in the degradation of neuropeptides[18,21].

These two enzymes have been found to be almost identical with regard to molecular weight (app. 95,000), substrate specificity and pH optimum[32,33]. However, they are inhibited at slightly different rates by the same inhibitors[34,35] and autocatalytic immunoinhibition studies[36] have shown that the kidney enzyme has different antigenic determinants from the brain enzyme. One explanation for this could be that in both organs the enzyme is a membrane bound glycoprotein and as such its carbohydrate component could differ. This may result in small changes in the protein's conformation which, although not affecting substrate specificity, could alter the binding of inhibitors. However, differences in the amino acid sequences of the two enzymes cannot be ruled out.

As mentioned in the previous section, EC 3.4.24.11. is of interest to the pharmaceutical industry due to the likelihood of inhibitors finding applications as analgesics. However, Pfizer are interested in the enzyme for other reasons which cannot be disclosed in this thesis.

The design of potent inhibitors of this enzyme has been hampered by the fact that very little is known about the structure of its active site, nor is the binding mode of inhibitors understood. This is due to the fact that EC 3.4.24.11. is a glycoprotein and as such is resistant to crystallisation. However, EC 3.4.24.11. belongs to a group of metalloproteases and has a very similar activity to the

bacterial enzyme Thermolysin, whose structure is well known from X-ray crystallography[37].

As discussed in Chapter 5, convergence on function is quite common and many enzymes possessing similar function and substrate specificities have very similar active sites. Superimpositions of the active sites of Thermolysin, carbonic anhydrase, liver alcohol dehydrogenase and carboxypeptidase-A have shown that there is a marked similarity in the structures of their active sites, especially in the regions of the zinc atoms[38].

Therefore, the main objective of this project was to try to predict some of the features of the active site of EC 3.4.24.11. from the active site of Thermolysin using information from substrate and inhibitor studies, molecular mechanics calculations and quantitative structure activity relationships.

The first stage of the project involved the comparison of low energy conformers of inhibitors with the binding conformations and investigating whether or not this could be related to the potency of the inhibitor.

Section 6.3.

Conformational Studies of β-phenylpropionyl-L-phenylalanine and Carbobenzoxy-L-phenylalanine.

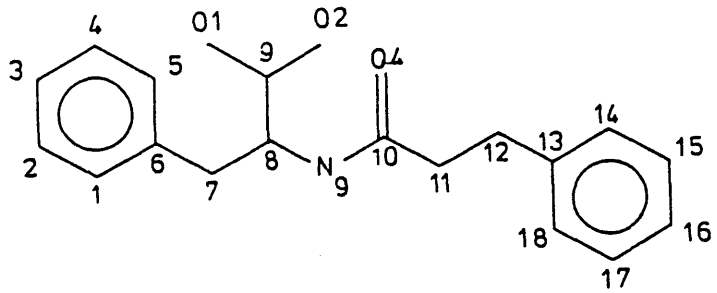
For any molecule to bind in an active site it must be able to arrange its spatial geometry to fit the active site. If in doing this the molecule has to adopt a high energy conformation it is likely that it will not be as potent as expected, conversely, if it can maximise binding with the active site while remaining in a low energy conformation it is likely to be a potent inhibitor.

Inhibition constants cannot solely be relied upon as a measure of how well a drug binds to an active site, unless they can be compared with how well it would be expected to bind. Dr.P.Andrews has devised a method[39], for calculating the average binding energy for any molecule, by assuming that each group in a molecule will contribute on average a certain amount of energy to the total binding energy. Summation of these contributions gives an average binding energy for the molecule from which can be derived an average K_I, or K_D for the molecule. If the observed K_I is larger than expected we could predict that the molecule is not able to utilise all of its groups in binding to the active site and may be binding in a higher energy conformation. Conversely if the observed K_I is much smaller than the average K_I the molecule is binding more strongly than expected and is able to utilise most or all of the binding potential of its functional groups and may be binding in a low energy conformation.

The two molecules studied are β-phenylpropionyl-L-phenylalanine and carbobenzoxy-L-phenylalanine, which are both inhibitors of the

bacterial enzyme, Thermolysin. Although these molecules are very similar (isosteric), X-ray studies have shown their binding conformations to be very different[40] and this can be explained using molecular mechanics studies.

β-Phenylpropionyl-L-phenylalanine



Using Andrew's method it is calculated that β-PPP has an average binding energy of 5.5 kCal mol^{-1} and an average K_I of 9.265×10^{-5}. Since the observed K_I is 1.6×10^{-3} it is apparent that the molecule is not binding to the active site as well as expected,but since it is known from X-ray crystallography that the majority of the functional groups are being used it is possible that it is binding in a high energy conformation. In β-PPP there are eight torsion angles which determine the overall conformation of the molecule, namely:-

TA1 C1 C6 C7 C8 TA2 C6 C7 C8 N9 TA3 C7 C8 N9 C10

TA4 C8 N9 C10 C11 TA5 N9 C10 C11 C12 TA6 C10 C11 C12 C13

TA7 C11 C12 C13 C14 TA8 C7 C8 C9 01

Since the amide bond is relatively rigid there are only seven bonds that need to be considered around which restricted rotation can occur.

Several low energy conformations were generated for β-PPP (for a

description of this procedure see chapter 3). These conformations were then minimised and the torsion angles of the resulting structures are shown in Table 6.1. File PPP.TLN is the minimised form of the binding conformation from the X-ray study.

How then does the binding conformation differ from the other low energy conformations found ? The first obvious difference is that the binding conformation even after minimisation is much higher in energy than the other conformers. Table 6.1 shows that for the binding conformation torsion angles 1 and 2 have values in the same range as conformations of similar energies. Torsion angles 4, 5, 7 and 8 have values in the same range as conformations with a spread of energies. However torsion angles 3 and 6 have values unlike anything found for the other low energy conformations.

The reason for this is apparent on examining the binding scheme for β-PPP and Thermolysin (see Fig. 6.4). From this diagram it can be seen that the differences occuring between β-PPP and the other low energy conformations arise because:-

1) Torsion angles 1 and 2 adopt values associated with higher energy conformations in order to allow the phenyl ring access to the hydrophobic pocket.

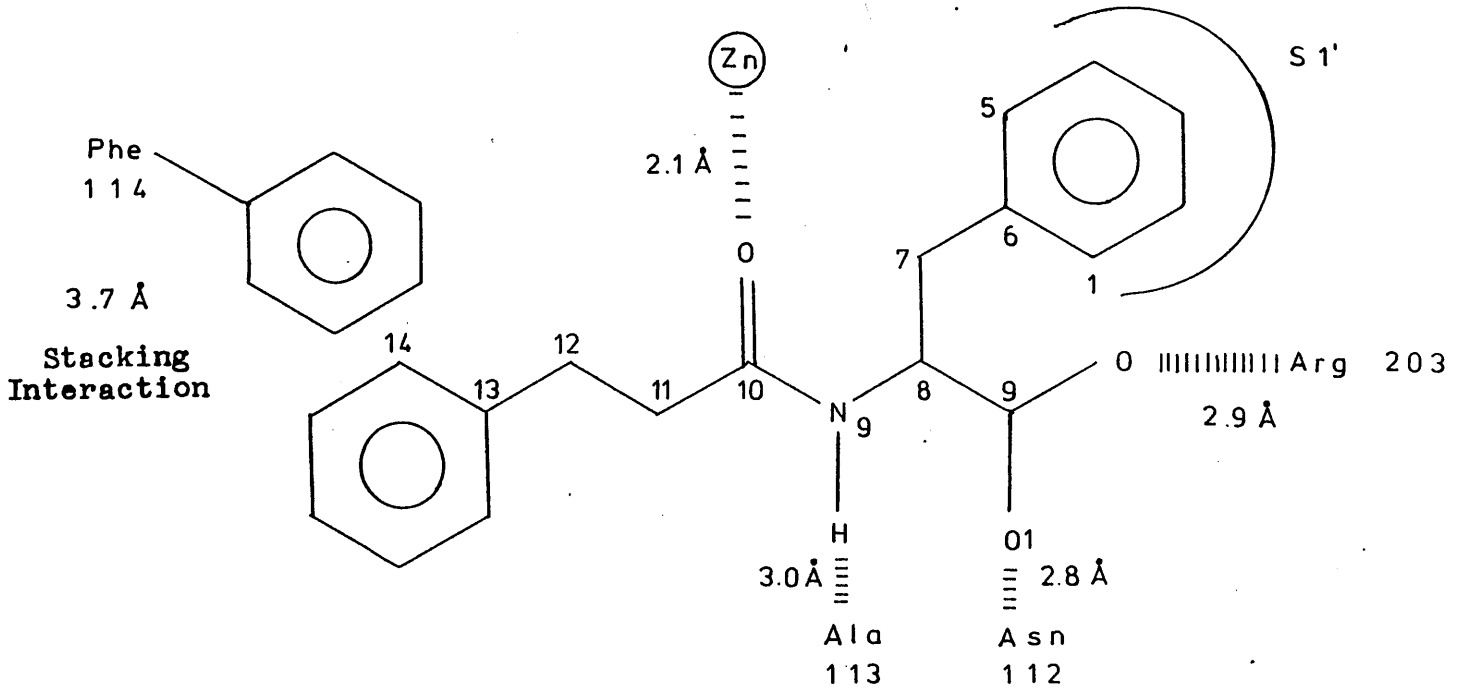
2) Torsion angle 3 has a value completely different from those found in other low energy conformations since its value determines the spatial relationship between the carboxylate group and the amide group and the value it adopts in the binding conformation is such that it maximises binding of these groups with the active site. Torsion angle values commonly found in lower energy conformations would not do this.

3) The position of the phenylalanine phenyl ring is determined by

TABLE 6.1

	TORSION ANGLES								
FILE	1	2	3	4	5	6	7	8	ENERGY (kCal mol^{-1})
PPMN3	-83	-54	100	-178	-72	57	-108	59	6.949
PPMN1	-80	58	92	-178	-68	59	-110	64	7.020
PPMIN2	-83	-60	102	177	-95	57	-105	66	7.114
PPMIN9	-60	-72	94	173	66	65	-120	62	7.335
PPMN4	-95	-169	91	179	-72	58	-108	61	7.461
PPMN13	-48	-64	101	171	60	62	-119	112	7.840
PPMN7	-76	-49	106	-177	64	176	-107	63	7.906
PPMIN3	-89	56	75	178	-75	59	-101	102	7.988
PPMIN5	-77	-48	107	-177	65	178	-91	112	8.225
PPMN5	-71	-62	98	-177	67	-91	-103	62	8.579
PPMN10	-82	-55	95	-176	-70	-178	-108	62	8.648
PPMIN7	-84	-54	96	-176	-69	177	-94	113	8.852
PPMN9	-96	-168	89	-177	62	178	-104	62	9.008
PPMN11	-96	-167	82	-177	-68	-179	-105	62	9.258
PPMN12	-88	54	72	-176	-68	179	-109	73	9.497
PPMN8	-98	45	72	-178	74	-96	-113	79	9.534
PPMN6	-95	28	75	179	-63	-61	-86	65	9.799
PPMN2	-130	-175	69	-172	55	16	-112	61	10.392
PPP.TLN	-118	-177	141	-178	-62	143	-102	65	10.710

Binding Scheme for β-PPP and Thermolysin



TORSION ANGLES

1) C1 C6 C7 C8 2) C6 C7 C8 N9 3) C7 C8 N9 C10 4) C8 N9 C10 C11

5) N9 C10 C11 C12 6) C10 C11 C12 C13 7) C11 C12 C13 C14

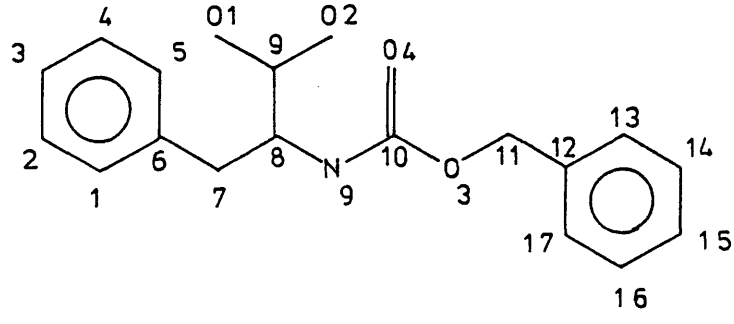
8) C7 C8 C9 01

Figure 6.4

the value of torsion angle 6. We see that instead of taking up a more favourable value torsion angle 6 adopts a value which places the phenyl ring in a better position for stacking with Phe 114.

In summary it can be said that β-PPP is much less active than expected (it has only .057 of its expected activity) since it is binding in a relatively high energy conformation. This is due to the fact that it is unable to bind successfully to the active site in one of its lowest energy forms.

Carbobenzoxy-L-phenylalanine



Comparing diagrams it can be seen that carbobenzoxy-L-phenylalanine is isosteric to β-PPP, however X-ray crystallography studies have shown that it binds to Thermolysin in a completely different orientation to β-PPP. Using Andrew's method we calculate an average binding energy for carbobenzoxy-L-phenylalanine of 5.8 kCal mol^{-1} which is equivalent to an average K_I of 5.6×10^{-5}. The observed K_I for the molecule is 5.1×10^{-4}. So the molecule is only 0.109 times as potent as expected (β-PPP was only .057 times as potent as expected).

Again there are eight torsion angles of interest, but particularly the values associated with the seven rotatable bonds.

These torsion angles are:-

TA1 C1 C6 C7 C8 TA2 C6 C7 C8 N9 TA3 C7 C8 N9 C10

TA4 C8 N9 C10 O3 TA5 N9 C10 O3 C11 TA6 C10 O3 C11 C12

TA7 O3 C11 C12 C13 TA8 C7 C8 C9 O1.

Low energy conformers of CBZ were generated using the same method as described for β-PPP. The torsion angles for the resulting low energy conformers are shown in Table 6.2. File CBZTLN gives the torsion angles for the minimised form of the bound structure from the X-ray data.

Table 6.2 shows that there is only an energy difference of 1.2 kCal mol^{-1} between the lowest energy conformation found and the minimised form of the binding conformation. Examination of this table shows the torsional values in the binding conformation agree well with torsion angle values found for other low energy minima. From this it can be seen that although carbobenzoxy -L Phe is less potent than predicted it is relatively more potent than β-PPP since it is able to bind in the active site in a relatively low energy conformation. The binding of carbobenzoxy-L-Phe is shown in Fig 6.5.

If the torsion angles in β-PPP are set to the values found in the binding conformation of CBZ and the structure is minimised a β-PPP conformation is obtained with an energy of 9.153 kCal mol^{-1} which is lower than the β-PPP binding conformation. However, β-PPP is unable to bind to Thermolysin in the same manner as CBZ since there is only 3.0Å separating the ester oxygen of CBZ and Asn 112 of Thermolysin and there is not enough room for a methylene group in this position.

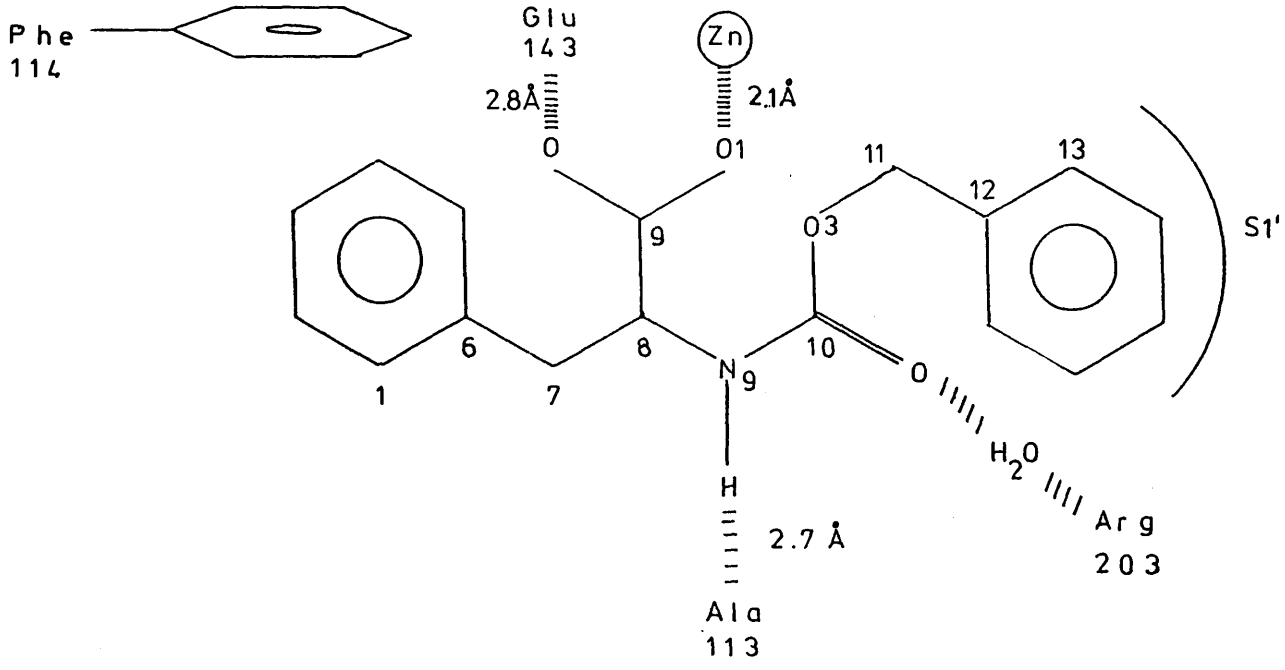
This result is in agreement with NMR studies carried out on the two molecules in that it demonstrates that the enzyme does not

TABLE 6.2

TORSION ANGLES

FILE	1	2	3	4	5	6	7	8	ENERGY (kCal mol^{-1})
CRMN7	-95	52	129	178	0	-79	113	115	8.729
CMIN9	-96	51	141	178	0	-89	97	113	9.272
CMIN12	-99	67	85	174	-4	88	75	112	9.532
CBZTLN	-77	-77	97	178	-2	-90	104	-119	9.960
CRMN3	-103	-173	79	173	-2	-63	126	-121	10.099
CMIN1	-78	-61	153	174	-169	75	66	115	10.234
CMIN6	-82	71	90	172	0	-64	155	113	10.353
CMIN5	-64	-74	125	177	8	101	74	118	10.386
CRMN4	-90	62	67	174	-3	-68	129	114	10.466
CRMN19	-89	63	67	174	-4	-67	129	114	10.519
CRMN20	-80	-59	149	174	-169	99	115	115	10.622
CMIN2	-110	-147	141	179	8	-57	112	112	10.688
CRMN15	-98	59	159	174	-172	95	105	113	10.762
CRMN8	-90	62	138	179	1	110	115	115	10.797
CRMN10	-101	-166	121	-177	3	-175	129	-115	10.991
CRMN17	-93	45	109	-177	3	-168	132	116	11.031
CMIN4	-57	70	61	-163	157	-82	91	115	11.072
CRMN12	-84	59	-134	173	-170	100	130	-115	11.100
CMIN10	-92	62	59	173	-6	-75	116	112	11.117
CRMN11	-101	-166	127	-175	4	98	108	-115	11.305
CRMN2	-65	81	-133	-179	4	-70	121	-117	11.495
CRMN5	-104	-175	73	179	-3	141	75	-119	11.530
CRMN13	-103	-165	140	169	-165	93	112	110	11.544
CMIN7	-64	-73	77	175	-11	-82	106	102	11.896
CRMN6	-93	63	53	175	-7	141	72	114	12.520

Binding Scheme for Carbobenzoxy -L- Phe and Thermolysin



TORSION ANGLES

1) C1 C6 C7 C8 2) C6 C7 C8 N9 3) C7 C8 N9 C10

4) C8 N9 C10 O3 5) N9 C10 O3 C11 6) C10 O3 C11 C12

7) O3 C11 C12 C13 8) C7 C8 C9 O1

Figure 6.5

necessarily bind to the conformations of lowest energy, but rather it binds those conformations which give the best fit and this results in a drop in the potency of the compounds.

Section 6.4.

Thermolysin and its Active Site

Thermolysin is a thermostable endopeptidase which has been isolated from the bacterium Bacillus Thermoproteolyticus. As previously discussed it belongs to a group of zinc dependent proteases and both its amino acid sequence and its three dimensional structure are known.

Substrate studies have shown that Thermolysin specifically cleaves peptide bonds on the amino side of a hydrophobic residue i.e. a hydrophobic residue in the P_1` position. The size of the hydrophobic group is crucial and Phe is the most favoured residue , the pocket is too small for larger residues such as Trp and Tyr and they cannot be accomodated as well. Conversely smaller residues do not maximise binding with the pocket and are therefore poorer substrates.

The enzyme structure is not thought to alter radically on crystallisation and indeed the crystal structure of the enzyme has been shown to hydrolyse substrates.[44]

X-ray studies on inhibitor complexes have given a major insight into the dimensions and natures of the functional groups which make up the S_1, S_1` and $S_2$` subsites of the active site and Figures 6.4 to 6.11 show details of the way in which the inhibitors β-phenylpropionyl-L-Phe (β-PPP)[40], carbobenzoxy-L-Phe (CBZ), phosphoramidon[41], L-benzylsuccinic acid[42], (2-benzyl-3-mercaptopropanyl)-L-Ala-glycinamide (BAG)[43], L-Leu-NHOH[44], HONH-benzylmalonyl-L-Ala-Gly-p-nitroanilide (Bzm-AGNA)[44] and N-(1-carboxy-3-phenylpropyl)-L-Leu-L-Trp (CLT)[45] bind to the active site.

Binding Scheme for Phosphoramidon and Thermolysin

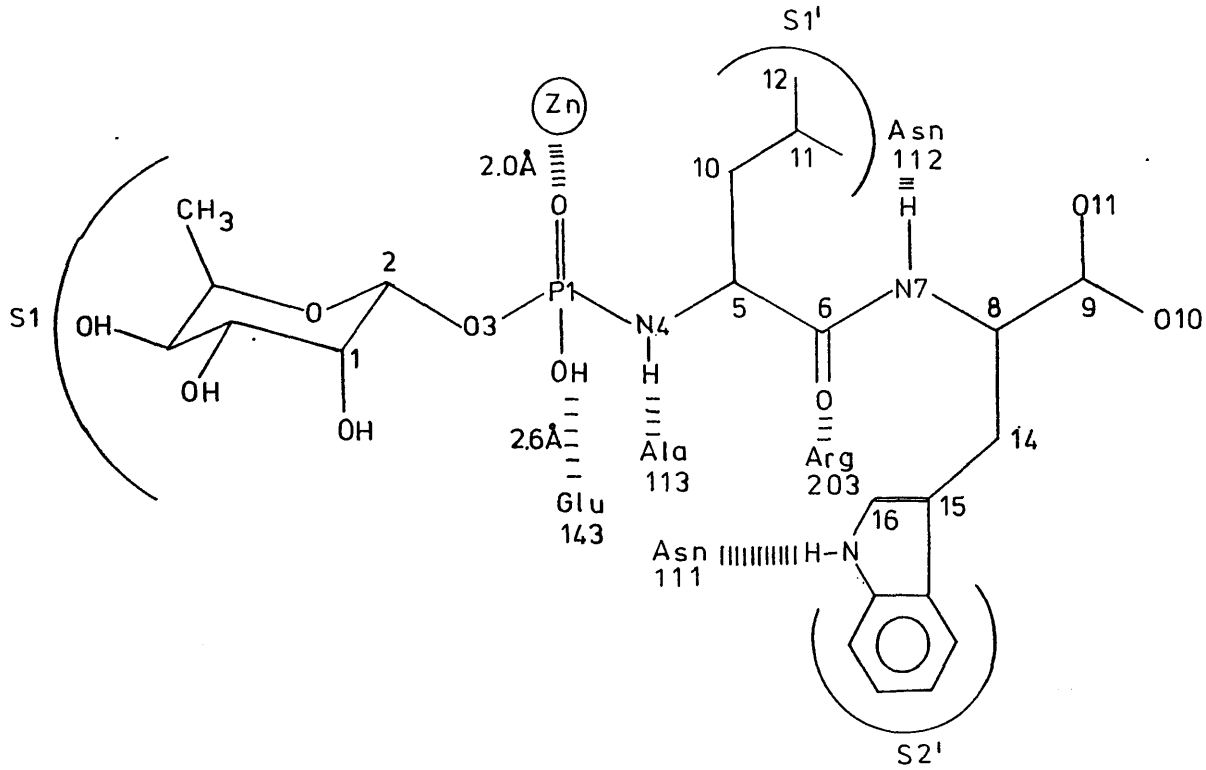
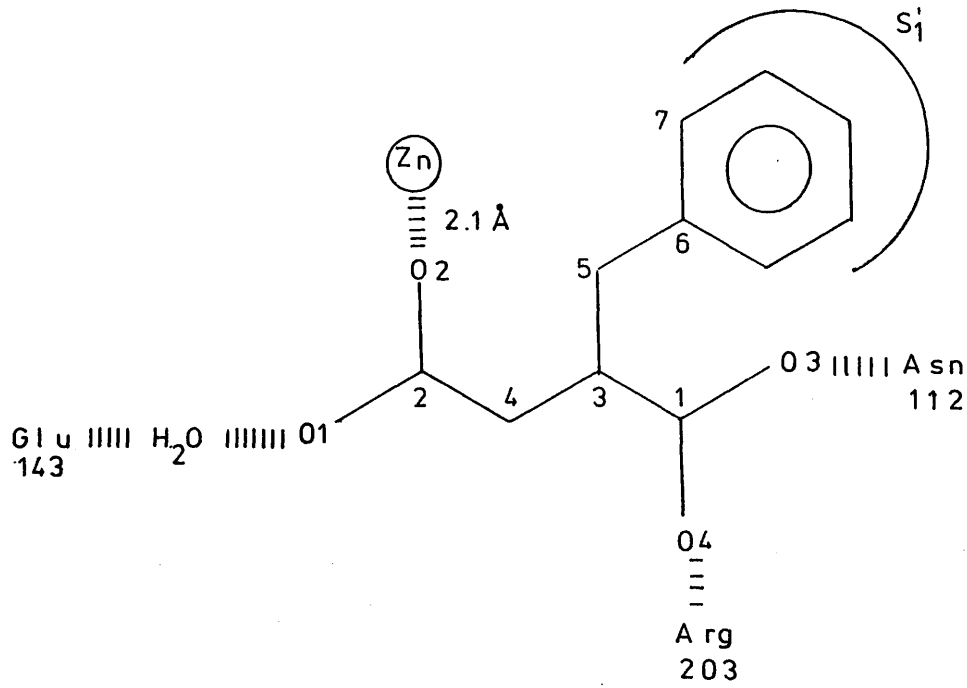


Figure 6.6

Torsion Angles

1) C1 C2 O3 P1
2) C2 O3 P1 N4
3) O3 P1 N4 C5
4) P1 N4 C5 C6
5) N4 C5 C6 N7
6) C5 C6 N7 C8
7) C6 N7 C8 C9
8) N7 C8 C9 O10
9) N4 C5 C10 C11
10) C5 C10 C11 C12
11) N7 C8 C14 C15
12) C8 C14 C15 C16

Binding Scheme for L-benzylsuccinic acid and Thermolysin



Torsion Angles

1) C2 C4 C3 C1

2) C4 C3 C1 O3

3) C4 C3 C5 C6

4) C3 C5 C6 C7

5) C3 C4 C2 O1

Figure 6.7

Binding Scheme for (2-benzyl-3-mercaptopropanoyl)-L-Ala-glycinamide and Thermolysin

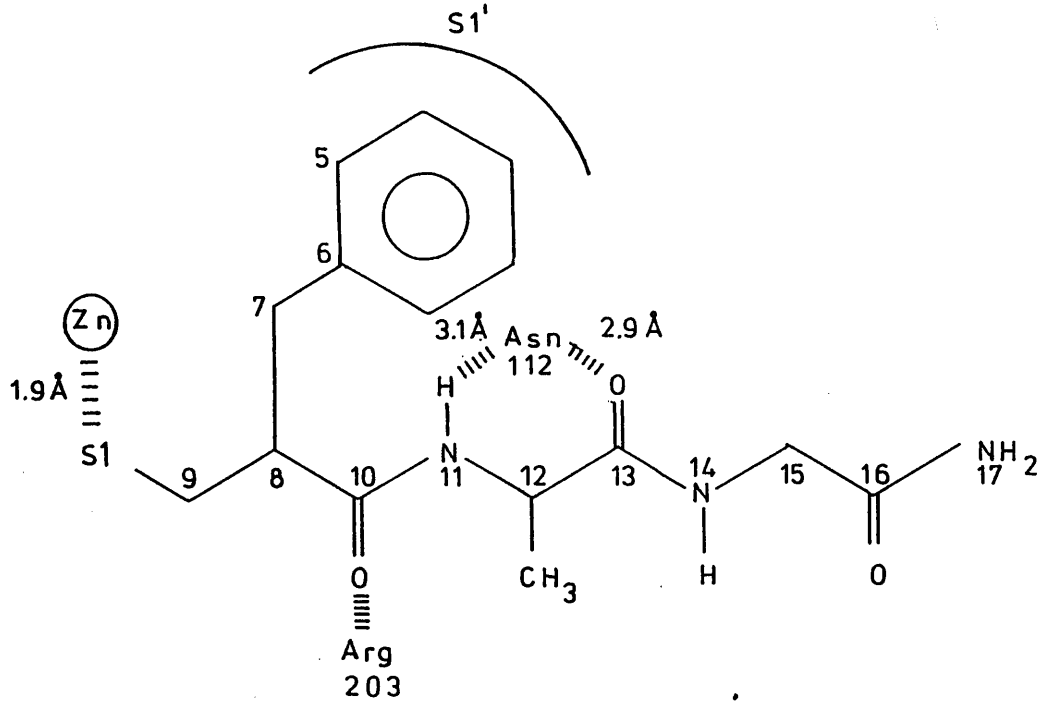


Figure 6.8

Torsion Angles

1) S1 C9 C8 C10

2) C9 C8 C10 N11

3) C8 C10 N11 C12

4) C10 N11 C12 C13

5) N11 C12 C13 N14

6) C12 C13 N14 C15

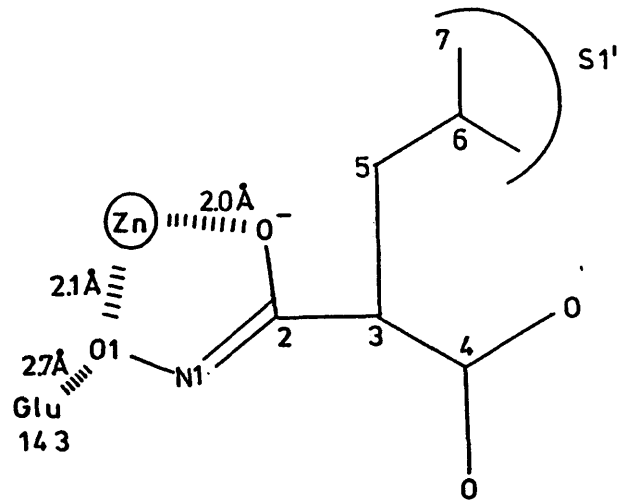
7) C13 N14 C15 C16

8) N14 C15 C16 N17

9) C9 C8 C7 C6

10) C8 C7 C6 C5

Binding Scheme for L-Leu-NHOH and Thermolysin



Torsion Angles

1) O1 N1 C2 C3

2) N1 C2 C3 C4

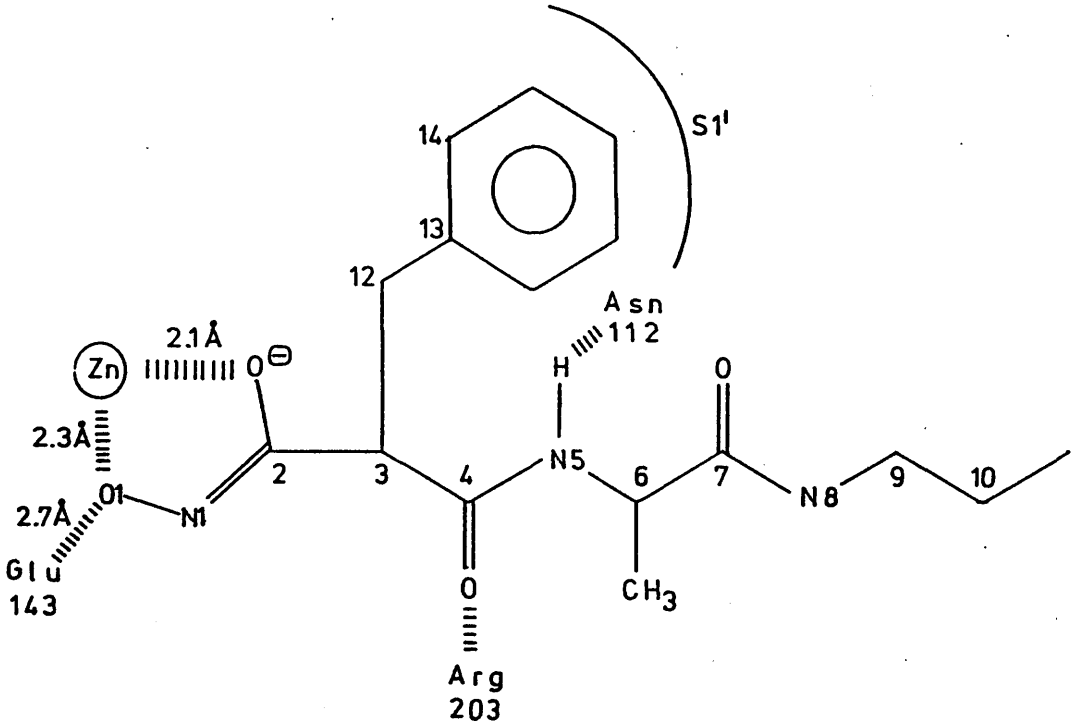
3) C2 C3 C4 O5

4) C2 C3 C5 C6

5) C3 C5 C6 C7

Figure 6.9

Binding Scheme for NHOH-benzylmalonyl-L-Ala-Gly-p-nitroanilide and Thermolysin



Torsion Angles

1) O1 N1 C2 C3
2) N1 C2 C3 C4
3) C2 C3 C4 N5
4) C3 C4 N5 C6
5) C4 N5 C6 C7
6) N5 C6 C7 N8
7) C6 C7 N8 C9
8) C7 N8 C9 C10
9) C2 C3 C12 C13
10) C3 C12 C13 C14

Figure 6.10

Binding Scheme for N-(1-carboxy-3-phenylpropyl)-L-Leu-L-Trp and Thermolysin

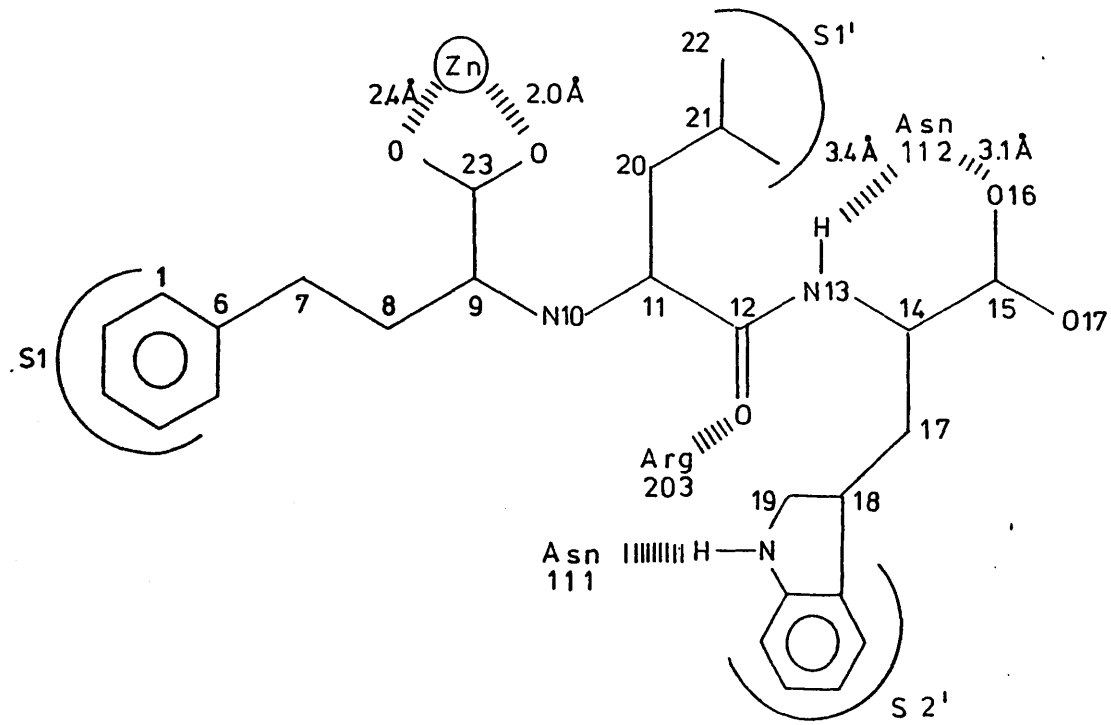


Figure 6.11

Torsion Angles

1) C1 C6 C7 C8

2) C6 C7 C8 C9

3) C7 C8 C9 N10

4) C8 C9 N10 C11

5) C9 N10 C11 C12

6) N10 C11 C12 N13

7) C11 C12 N13 C14

8) C12 N13 C14 C15

9) N13 C14 C15 O16

10) N10 C11 C20 C21

11) C11 C20 C21 C22

12) N13 C14 C17 C18

13) C14 C17 C18 C19

14) C8 C9 C23 O1

These diagrams show the presence of a hydrophobic binding region in the S_1 position along side Phe 114, a well defined hydrophobic pocket in the S_1` subsite (as predicted by the substrate studies) surrounded by residues Val 139, Phe 130, Leu 133, Ile 188, Gly 189, Val 192 and Leu 202 and a further nonspecific hydrophobic region in the $S_2$` position which can accomodate large hydrophobic groups. The diagrams also show the positions of the residues which would be expected to bind to the backbone of the peptide chain i.e. Ala 113 hydrogen bonding to the nitrogen of the scissile bond, Asn 111 possibly H-bonding to the protonated nitrogen of the scissile bond, Arg 203 hydrogen bonding to the P_1` carbonyl (or forming a salt linkeage to the terminal carboxylate) and Asn 112 hydrogen bonding to both the nitrogen and the carbonyl (or free terminal carboxylate) of the $P_2$` residue. When the P_2` position is occupied by a tryptophan residue the indole nitrogen of the side chain appears to be able to H-bond to Asn 111.

The first five of the inhibitors listed above all provide one ligand to bind to the zinc and since inhibitors tend to resemble the transition state of the substrate it was originally felt that the transition state intermediate would also involve a tetrahedral complex and a mechanism for catalysis was postulated from this [40,41] (see Fig. 6.12). Three groups were presumed to play a major role in hydrolysis namely Glu 143, His 231 and the zinc atom. This mechanism involved the carbonyl group of the scissile bond binding to the zinc. Glu 143 then acts as a base to promote attack on the carbonyl carbon by a water molecule, at the same time His 231 donates a proton to the peptide nitrogen giving an intermediate which is tetrahedral at both the carbon and nitrogen of the scissile bond, which then breaks to yield

Proposed Mechanism for the Cleavage of Peptides by Thermolysin Involving a Tetrahedrally Coordinated Zinc

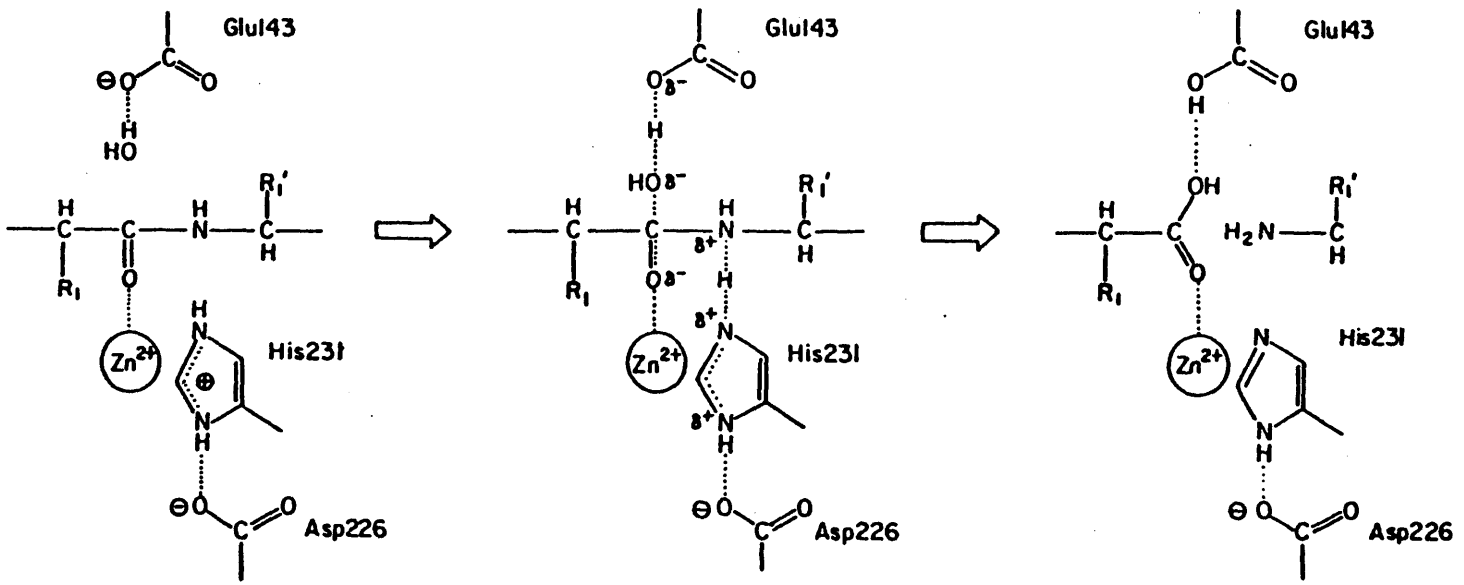


Figure 6.12

the products.

However, with the discovery of the pentavalent zinc coordination found for the hydroxamic inhibitors and for the N-(1carboxy-3-phenylpropyl)-L-Leu-L-Trp[44,46,47], a new mechanism has been proposed involving the pentacoordination of the zinc by a hydrated peptide bond and a diagram of this is shown in Fig. 6.13. This mechanism involves the addition of a water molecule across the petide bond, this tetrahedral structure then ligands to the zinc with both oxygens of the hydrated carbonyl followed by cleavage of the scissile bond. Glu 143 is again involved in hydrating the peptide bond, stabilising the tetrahedral intermediate and may also act as a proton shuttle - i.e. delivering back the proton to the N of the scissile bond.

Regardless of the mechanism of peptide cleavage the torsion angles associated with the binding inhibitors provide a vital source of information as to how the substrates would be expected to bind to Thermolysin. The structures of the inhibitors vary greatly, however they all have one thing in common and that is a hydrophobic residue at the $P_1`$ position . Several of the inhibitors also have a peptide chain which extends as far as the $P_3`$ position and a few also have groups which are binding in the S_1 subsite. From these structures comparisons can be made of the torsion angles involved in binding to the active site.

Figures 6.4 to 6.11 give details of the torsion angle numbering schemes of the inhibitors and Table 6.3 shows a diagram of a model substrate molecule binding to the $S_1`$, $S_2`$ and $S_3`$ active site with numbered torsion angles and a listing of the comparable torsion angles of the inhibitors as they bind to Thermolysin. The first obvious

Proposed Mechanism for the Cleavage of Peptides by Thermolysin Involving Pentacoordination of the Zinc

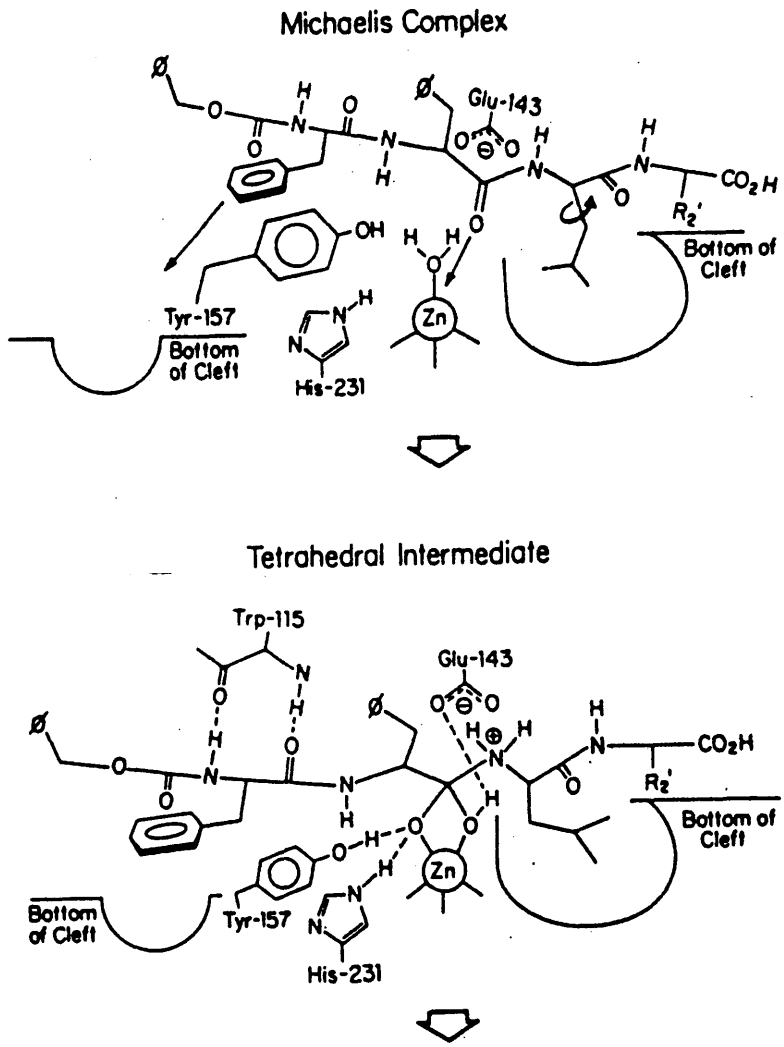


Figure 6.13

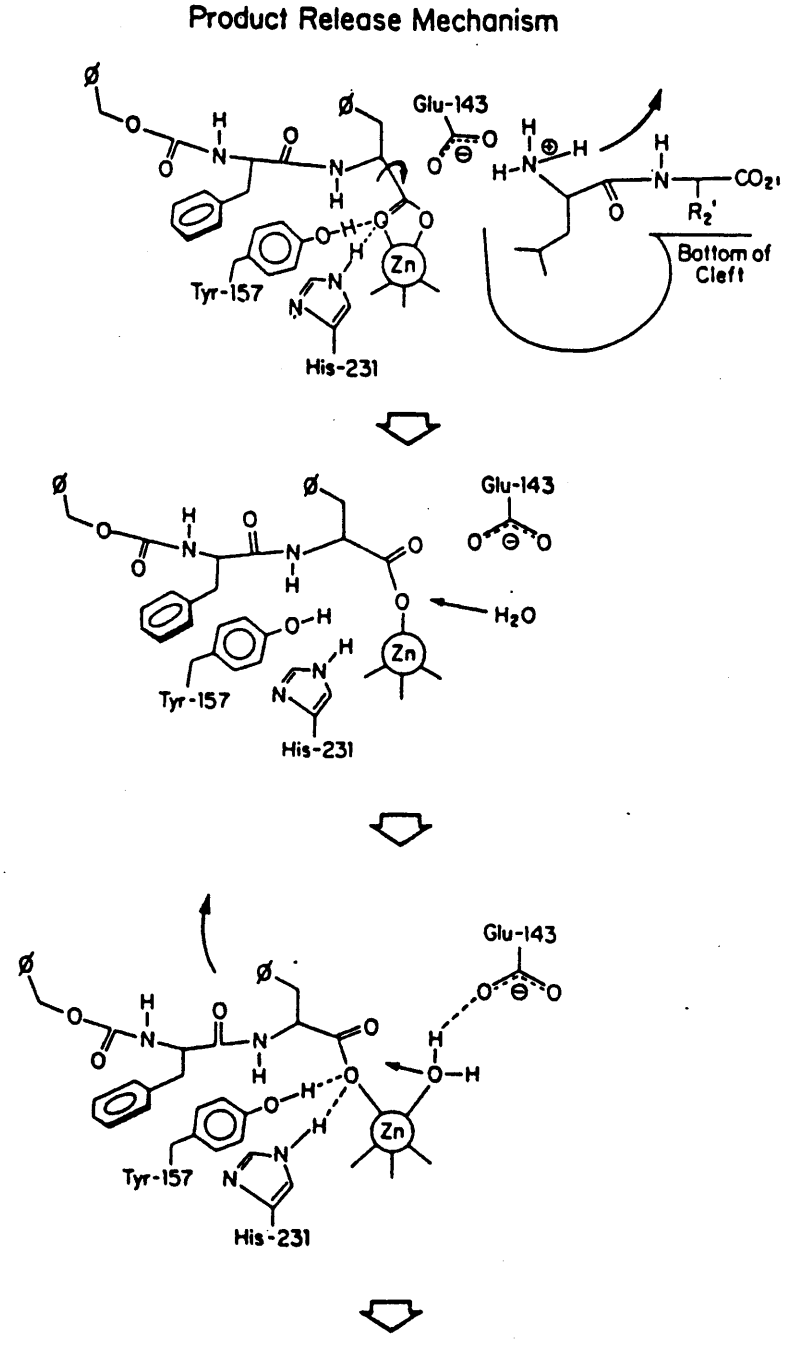
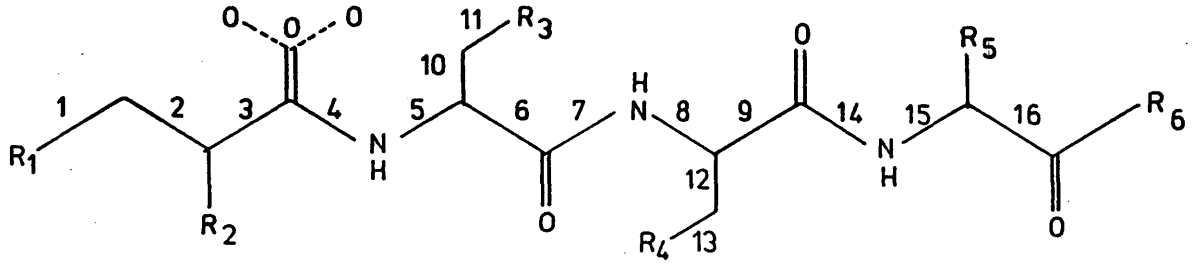


Figure 6.13 (cont.)



Torsion Angles	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Phosphoramidon		175	-34	141	-97	-55	-179	-142	-24	-165	35	62	-66			
C L T	-81	-177	-90	176	-50	-43	-176	-146	8	-171	60	76	-49	79		
β- P P P	-110	133	-48	-178	-96	-54				179	61					
C B Z		91	-59	95	180	-106	88	107								
B A G					-76	-69	164	-115	-23	-163	35			177	-146	125
B S A										-160	56					
Leu - NHOH						-162				-119	57					
Bzm - AGN						-100	179	-106	-18	-124	41			-176	-164	57

▨ non-comparable values

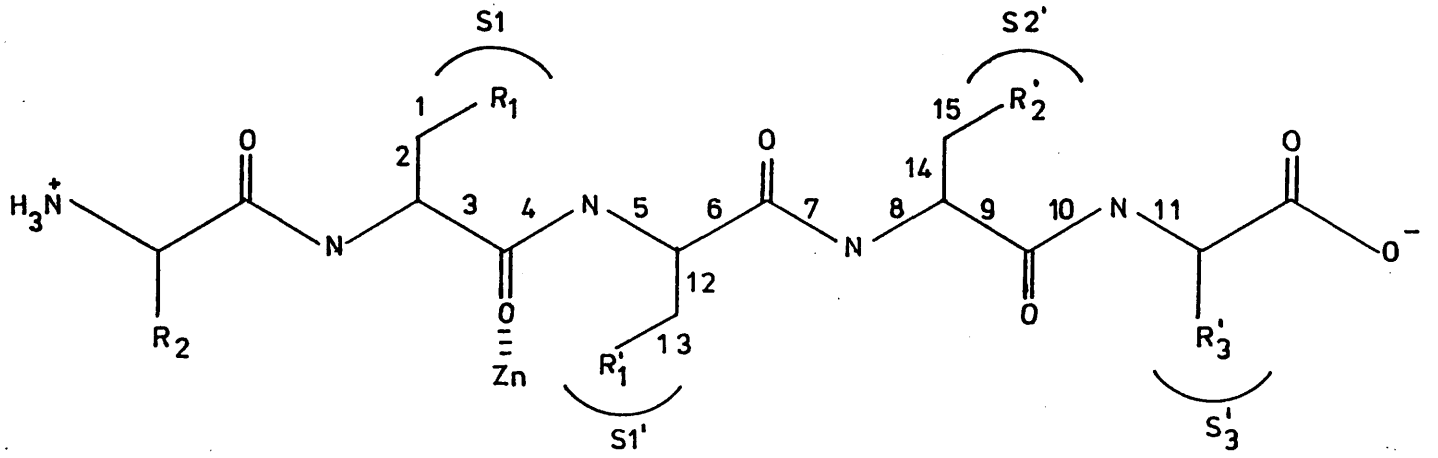
Table 6.3

similarity is in the torsion angles associated with the groups which are binding in the hydrophobic pocket of the S_1` subsite (columns 10 and 11), excluding the hydroxamic acid inhibitors the two relevant torsion angles appear to have values in the regions -160° to 180° and 35° to 60° respectivly.

Columns 6, 7 and 8 contain the backbone torsion angles as the inhibitors bind from the S_1` subsite to the $S_2$` and these (again excluding the torsion angles associated with hydroxamic acid inhibitors) are also in good agreement being in the regions -40° to -70°, -176° to 164° and -106° to -146°. Column 9 contains the torsion angle which determines the position of the P_2` carbonyl or carboxylate which binds to Asn 112 and again for those molecules which contain this group the torsion angles are in fairly good agreement being in the order of 8° to -24°

Columns 12 and 13 contain the torsion angles associated with the positioning of the tryptophan residue of the two inhibitors which contain this group and these torsion angles have values of 62° to 76° and -49° to -66° respectively. Finally the torsion angles found for the extended chain of Bzm-AGNA and BAG , columns 14 and 15, are found to be in the region -176° to 177° and -146° to -164° respectively. None of the torsion angles involving the binding of the hydrophobic residue in the S_1 subsite can be directly compared since they are so dissimilar.

From the similarities found in the torsion angle values for the inhibitors binding in the S_1`, $S_2$` and S_3` subsites it is possible to predict ranges for the torsion angles found in the substrate for the $P_1$`, P_2` and $P_3$` residues and these are given in Table 6.4.



Torsion Angles

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
-110	133	-48	-178	-50	-43	-176	-106	-24	177	-146	-160	35	62	-49
		-90	176	-96	-69	164	-146	8	176	-164	179	60	76	-66

TABLE 6.4

Section 6.5.

The Active Site of EC 3.4.24.11.

All the information currently known about the active site of EC 3.4.24.11 has come from substrate and inhibitor studies. The enzyme obeys Michaelis - Menten kinetics and substrate specificities have shown much about subsites within the enzyme.

Inhibitor studies have shown that the enzyme is neither a serine, or a thiol protease, nor is it a carboxyl protease. The enzyme is however inhibited by various metal chelating agents and loses all activity on centrifugation with EDTA. Since activity can be regained with zinc the enzyme is infact a member of the class of zinc proteases.[30]

As yet the residues preferred for substrate binding and catalysis have only been determined for subsites S_2, S_1 and S_1`. Table 6.5 shows the activity of the enzyme with naturally occuring peptides. From this it can be seen that, like Thermolysin the enzyme is highly specific for a bulky, hydrophobic residue at the $P_1$` position with Phenylalanine being preferred over Leucine, Isoleucine, Valine and Histidine. With this in mind several synthetic substrates have been made[35] with a Phe in position P_1` (see Table 6.6) and by varying those residues in the P_1 and P_2 positions it has been found that the best catalysis is obtained with an Ala residue at P_1 and high rates of hydrolysis are also observed with a Phe in P_2 coupled to a Gly, or an Ala in P_1. The presence of a Phe residue in the P_1 position seems to give better binding (lower K_i`s) but not better catalysis. Arg-Arg and Ala-Ala in the P_1 and P_2 also give good hydrolysis and although the

Table 6.5

Cleavage of Naturally Occuring Peptides by EC 3.4.24.11.

Compound	Structure
Met-Enkephalin	Tyr-Gly-Gly-Phe-Met
Leu-Enkephalin	Tyr-Gly-Gly-Phe-Leu
Met-Enkephalinamide	Tyr-Gly-Gly-Phe-Met-NH_2
Leu-Enkephalinamide	Tyr-Gly-Gly-Phe-Leu-NH_2
Angiotensin I	Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu
Angiotensin II	Asp-Arg-Val-Tyr-Ile-His-Pro-Phe
Bradykinin	Arg-Pro-Phe-Gly-Phe-Ser-Pro-Phe-Arg
Renin substrate tetradecapeptide	Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu-Leu-Val-Tyr-Ser

v indicates a primary site of cleavage

v indicates a secondary site of cleavage

TABLE 6.6

SUBSTRATES					Km (mM)		kcat (s-1)	
P3	P2	P1	P1'	P2'	X=pAb	X=2NA	X=pAB	X=2NA
Glt -	Gly -	Gly -	Phe -	X	0.071 (2.9)	0.092	3.3 (0.72)	2.3
Glt -	Gly -	Ala -	Phe -	X	0.19 (3.6)	0.49	9.3 (6.1)	104
Glt -	Gly -	Phe -	Phe -	X	0.19 (0.94)	0.042	6.9 (36.2)	2.0
Glt -	Ala -	Gly -	Phe -	X	0.15 (1.5)	0.074	11.7 (22.7)	41.6
Glt -	Ala -	Ala -	Phe -	X	0.13 (2.6)	0.20	12.6 (132.0)	139
Glt -	Ala -	Phe -	Phe -	X	0.21 (0.92)	0.036	13.1 (367.0)	14.1
Glt -	Phe -	Gly -	Phe -	X	0.14 (0.71)	0.01	14.0 (23.0)	42.3
Glt -	Phe -	Ala -	Phe -	X	0.19 (1.1)	0.058	17.0 (161.0)	173
Glt -	Phe -	Phe -	Phe -	X	0.09 (0.44)	0.021	11.0 (339.0)	2.6

Figures show in parenthesis are for Thermolysin

rate of hydrolysis increases with longer chains implying the presence of an extended binding site it would appear that the enzyme can accommodate a large range of residues in the P_1 and P_2 subsites without losing too much activity. The residue which plays the most important role in the function of the enzyme is the one occupying the S_1' subsite .

Initially it was thought that although the enzyme could act as an endopeptidase its function was primarily that of a carboxy dipeptidase, similar to ACE . This was proposed for three reasons.

1) Studies of the substrate specificity of the brain enzyme showed that it cleaved Leu- and Met-enkephalin at the Gly^3-Phe^4 bond (see Table 6.5)[21]

2) It had been noted that a free carboxyl in the P_2' position enhanced the rate of hydrolysis since amidated derivatives were hydrolysed much more slowly than the natural substrates (see Table 6.7).[21,48]

3) Inhibition studies with butanedione, a compound which binds strongly to arginine residues, indicated the presence of an arginine residue at the S_2' subsite since the enzyme lost about 80% of its activity with natural substrates, only 50% of its activity with amidated derivates and 20% of its activity with napthylamide derivitives (see Table 6.8).[48,49] Therefore it was proposed that the free carboxylate of the natural substrate was binding to this arginine (perhaps as a salt linkage) giving an increase in catalysis and blocking this arginine residue results in a loss in activity.

It is now accepted that the primary function of the enzyme is that of an endopeptidase and that the carboxy dipeptidase activity is only apparent when the penultimate amino acid of the substrate is a hydrophobic residue.

Table 6.7

Reactivities of Various Peptides with EC 3.4.24.11.

Compound	K_m (um)	k_{cat} (min^{-1})	k_{cat}/K_m
[Leu^5]-Enkephalin	86	3781	43.9
[Leu^5]-Enkephalinamide	680	1129	1.7
[D.Ala^2,Met^5]-Enkephalin	13	1044	80
[D.Ala^2,Met^5]-Enkephalinamide	320	2512	8
Substance P	31	5062	158
Substance P deaminated	37	7766	202

Table 6.8 (a)

Comparison of the Degradation of Various Enkephalin Analogs per unit time by EC 3.4.24.11. from Brain and Kidney.

Compound	% degraded	
	kidney	brain
Tyr-Gly-Gly-Phe-Leu	20	25
Tyr-Gly-Gly-Phe-Leu-NH_2	2	4
Tyr-D.Ala-Gly-Phe-Leu-NH_2	3	3
Tyr-Gly-Gly-Phe-Met	22	31
Tyr-Gly-Gly-Phe-Met-NH_2	12	18
Tyr-D.Ala-Gly-Phe-Met-NH_2	10	14
Tyr-Gly-Gly-Phe-Met-Arg-Phe	18	26
Tyr-Gly-Gly-Phe-Met-Arg-Phe-NH_2	7	11
Phe-Met-Arg-Phe-NH_2	2	1
D.Ala-Gly-Phe-Met-NH_2	0	10
Tyr-Gly-Gly-Phe-Met-Arg-Arg	5	9
Tyr-Gly-Gly-Phe-Leu-Arg	5	8
N-acetyl-Tyr-Gly-Gly-Phe-Leu	15	5

(continued overleaf)

Table 6.8 (b)

Activity of Free Enzyme and Butanedione Treated Enzyme with Various Substrates

Substrate	Activity (umole/mg/min)		
	Control	Butanedione treated	%decreased
Tyr-Gly-Gly-Phe-Met	4.9	0.83	83
Tyr-Gly-Gly-Phe-Leu	2.2	0.46	79
Tyr-D.Ala-Gly-Phe-Leu	1.6	0.27	83
Tyr-D.Ala-Gly-Phe-Met-NH_2	.64	0.3	53
Hip-Arg-Leu-Leu-βNA	1720	1324	23

Section 6.6.

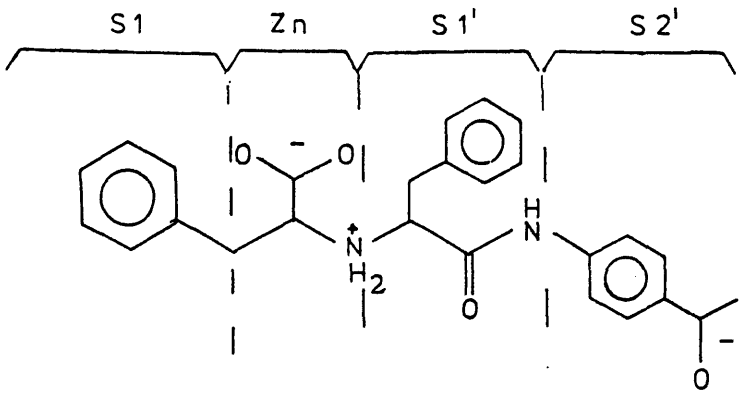
Inhibitor Studies

The substrate studies have lead to predictions as to the binding modes of the inhibitors and Fig. 6.14 shows the predicted binding of several of the inhibitors. However, this is only a two dimensional scheme and the design of a three dimensional active site requires more information.

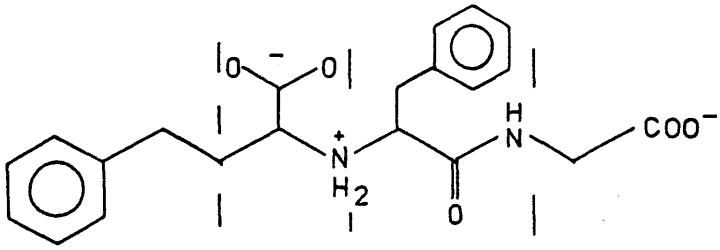
Many of the inhibitors of the enzyme closely resemble the natural substrates (or their transition states) and the three dimensional structures of these inhibitors can be used to gain information about the active site. However, it is not enough to generate low energy conformers of the inhibitors since it was shown in section 6.3. that the enzyme does not necessarily bind the lowest energy conformation of the molecule instead it binds the conformation which gives the best fit.

The first stage in elucidating the active site must therefore involve predicting which compounds are binding in low energy forms by comparing their expected, or average K_i`s with their actual $K_i$`s.

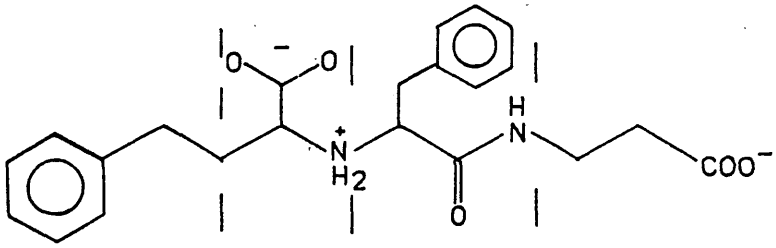
Tables 6.9 and 6.10 list several of the more potent inhibitors with their average and observed Ki`s (the average Ki`s being calculated in the same manner as in section 6.3). It must be noted that the average K_i for the compound Phe-N-L-Phe-pAB was calculated without including the contribution made to binding by the terminal carboxylate group (see Fig. 6.14). This contribution was omitted because the para-aminobenzoate group is totally rigid group and modifications to the position and functionality of the carboxylate



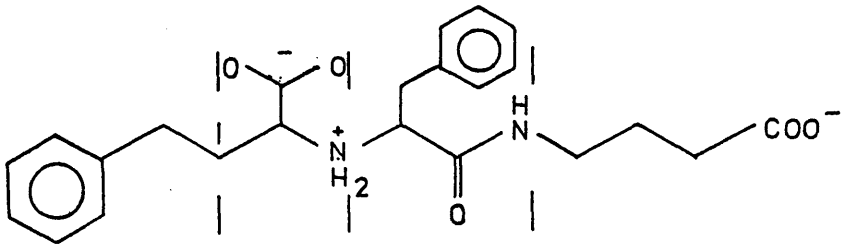
Phe-N-L-Phe-p-AB



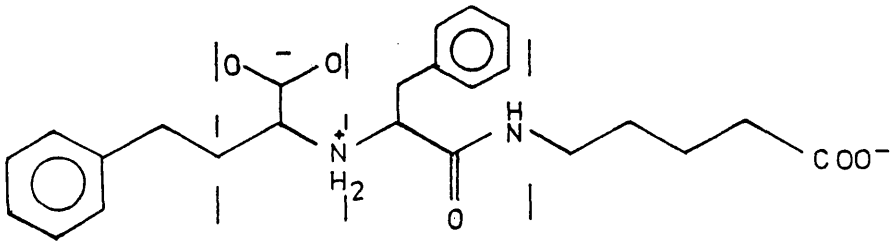
HomoPhe-N-L-Phe-L-Gly



HomoPhe-N-L-Phe-βAla

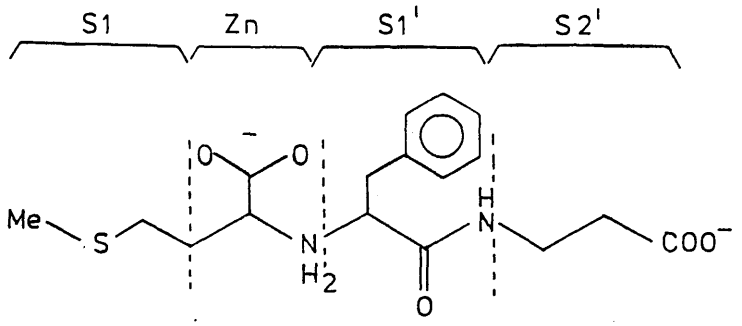


HomoPhe-N-L-Phe-γABA

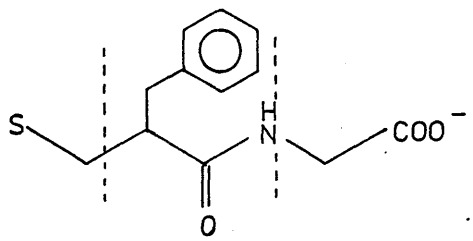


HomoPhe-N-L-Phe-δAVA

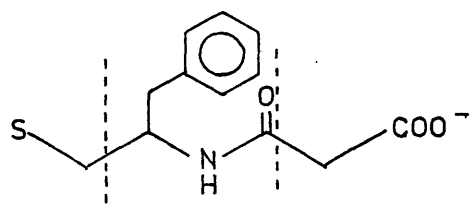
Figure 6.14



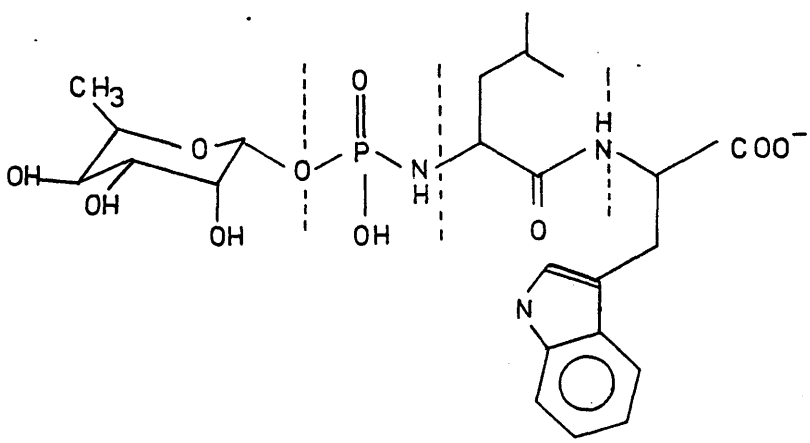
Met - N-L-Phe - βAla



Thiorphan



Retrothiorphan



Phosphoramidon

Figure 6.14 (cont.)

made little difference to the potency of the compounds (Pfizer's unpublished results) therefore the carboxylate is not thought to be involved in binding to the active site. This does not agree with published data which suggest that the carboxylate of the Phe-[N]-L-Phe-pAB is binding to an arginine at the S_2' subsite. However, the compounds Phe-pAB and Phe-Ala both inhibit with similar potencies (see table 6.10), with K_i's of 1.9×10^{-5}M and 2.1×10^{-5}M respectively. Assuming that for each inhibitor the phenylalanine is occupying the S_1' binding site then the terminal carboxylate of Phe-Ala should be in the correct position to bind to the residue (possibly arginine) at the S_2' subsite, but it is highly unlikely that the terminal carboxylate of the Phe-pAb is able to bind to the same residue since the para-aminobenzoate group is totally rigid and in terms of a substrate backbone chain (nitrogen, carbon, carbonyl repeat unit) is separated from the P_1' carbonyl by six atoms corresponding to the P_3' carbonyl position of a substrate.

Table 6.9 shows a comparison of the average K_i's and observed ID_{50}'s for a series of Phe-[N]-L-Phe type inhibitors. As for the Thermolysin inhibitors β-PPP and CBZ (see section 6.3) the average K_i's were calculated using Andrew's method. It is interesting to note that the expected potencies are far larger than the observed potencies. This is probably due to the fact that all the molecules were treated as being fully dissociated and it is likely that Andrew's method overestimates the contributions made by the charge groups under such circumstances. However, all the compounds contain the same number of charge groups and although the average K_i's do not compare in any way with the observed potencies, they may be useful for comparing the relative potencies of the compounds.

Table 6.11 lists the expected order of potencies of the Phe-[N]-L-Phe type compounds and their observed ID_{50}'s.

TABLE 6.9

	ID50	Ki (average)
1)HomoPhe-N-L-Phe-Gly	4.5×10^{-7}	2.4×10^{-18}
2)HomoPhe-N-L-Phe- βAla	3.2×10^{-7}	2.1×10^{-18}
3)HomoPhe-N-L-Phe- γABA	2.0×10^{-7}	1.7×10^{-18}
4)HomoPhe-N-L-Phe- δAVA	7.0×10^{-7}	1.5×10^{-18}
5)HomoTyr-N-L-Phe- βAla	2.7×10^{-7}	1.0×10^{-19}
6)HomoPhe-N-L-Phe-IsoAsn	4.5×10^{-7}	9.5×10^{-21}
7)HomoPhe-N-L-pClPhe- βAla	2.5×10^{-7}	2.4×10^{-19}
8)Met-N-L-Phe- βAla	7.5×10^{-7}	1.0×10^{-16}
9)Phe-N-L-Phe- βAla	6.5×10^{-7}	2.5×10^{-18}

(all inhibitors have S,S geometry)

ID50 = Inhibitory dosage giving 50% inhibition.

TABLE 6.10

	Ki (observed)	Ki (expected)
(S)-Phe-N-L-Phe-pAB	2.9×10^{-8}	9.7×10^{-15}
(R)-Phe-N-L-Phe-pAB	2.4×10^{-7}	9.7×10^{-15}
	IC50	
(S)-Thiorphan	2.1×10^{-9}	2.4×10^{-10}
(R)-Thiorphan	3.6×10^{-9}	" "
Retrothiorphan	6.0×10^{-9}	" "
	Ki (observed)	
Phe-Ala	2.1×10^{-5}	
Phe-pAB	1.9×10^{-5}	

	Expected Potency	Observed Potency
Homo-Phe-[N]-L-Phe-IsoAsn	9.48×10^{-21}	4.5×10^{-7}
Homo-Tyr-[N]-L-Phe-βAla	1.0×10^{-19}	2.7×10^{-7}
Homo-Phe-[N]-L-pCl-Phe-βAla	2.35×10^{-19}	2.5×10^{-7}
Homo-Phe-[N]-L-Phe-δAVA	1.5×10^{-18}	7.0×10^{-7}
Homo-Phe-[N]-L-Phe-γABA	1.78×10^{-18}	2.0×10^{-7}
Homo-Phe-[N]-L-Phe-βAla	2.1×10^{-18}	3.2×10^{-7}
Homo-Phe-[N]-L-Phe-Gly	2.49×10^{-18}	4.5×10^{-7}
Phe-[N]-L-Phe-βAla	2.49×10^{-18}	6.5×10^{-7}
Met-[N]-L-Phe-βAla	1.0×10^{-16}	7.5×10^{-7}

Table 6.11

The two most noticable discrepancies between the expected inhibition constants and the observed ID_{50}`s are for Homo-Phe-[N]-L-Phe-IsoAsn and Met-[N]-L-Phe-βAla. Homo-Phe-[N]-L-Phe-IsoAsn is expected to be the most potent inhibitor, however instead of increasing the potency of the compound the presence of the IsoAsn side chain actually decreases potency. This implies that the side chain of the IsoAsn is not actually involved in binding to the active site and could be lying in a hydrophobic area similar to the S_2` site in Thermolysin. The other major difference noted with this series of inhibitors is that Met-[N]-L-Phe-βAla although still the poorest inhibitor is binding relatively better than the rest of the compounds.

Compounds 1 to 4 of Table 6.9 belong to a series of compounds of the type Homo-Phe-[N]-L-$(CH_2)_n$-CO_2^- in which the carbon chain length increases from 1 to 4. From the calculated K_i`s compound 4 would be expected to be the most potent with a stepwise decrease in activity to compound 1. However from the Table 6.12 it can be seen that this does not happen.

Expected Order of Activity		Observed Order of Activity	
	x 10^{-18}		x 10^{-7}
4)	1.5	3)	2.0
3)	1.78	2)	3.2
2)	2.10	1)	4.5
1)	2.49	4)	7.0

Table 6.12

With each increase in the chain length the potency of the compound is expected to increase by a factor of roughly 1.2 .The observed values closely mirror this fact upto n=3. Increasing the chain length from 1 to 2 gives an increase of the order of 1.4, increasing the chain length from 2 to 3 gives an increase in potency of the order of 1.6, however increasing the chain length to 4 gives a dramatic decrease in the potency. This can be explained in one of two ways, either the loss of a good interaction (i.e. carboxylate binding) which, with corresponding gain of entropy leads to a loss in binding energy, or the introduction of a bad interaction causing rearrangement of the carbon chain and a resulting loss in binding energy.

There is also a series of β-Ala derivatives whose activities can be compared.

Expected Order of potency		Observed Order of Potency	
5)	1.0×10^{-19}	7)	2.5×10^{-7}
7)	2.35×10^{-19}	5)	2.7×10^{-7}
2)	2.10×10^{-18}	2)	3.2×10^{-7}
9)	2.49×10^{-18}	9)	6.5×10^{-7}
8)	1.02×10^{-16}	8)	7.5×10^{-7}

Table 6.13

The order does not differ significantly, but the relative activities are quite different. Compounds 5 and 7 are expected to be an order of magnitude more potent than compound 2. This is not the case and it can therefore be assumed that neither the hydroxyl of the Tyr (compound 5) or the chlorine of the pCl-Phe (compound 7) are actually contributing much to the binding energies. The slight increase observed in the potencies of these two compounds when compared with the Homo-Phe-[N]-L-Phe- βAla could be explained by the fact that both the S_1 and the S_1` hydrophobic pockets are of limited size. The introduction of a Tyr, or a pCl-Phe into either pocket may prevent the phenyl fully penetrating into the pocket, causing a slight alteration in the orientation of the other binding groups giving rise to an increase in the binding energy. Alternatively the S_1 site may be not so much a pocket as a patch of limited surface. Compound 2 is expected to be approximately 1.2 times as potent as compound 9 and infact there is a twofold increase in the observed potencies of these inhibitors. Finally it can be seen that compound 8 is indeed relatively more potent than the other compounds since it is expected to have 0.024 times the activity of compound 9 and in fact it has 0.8

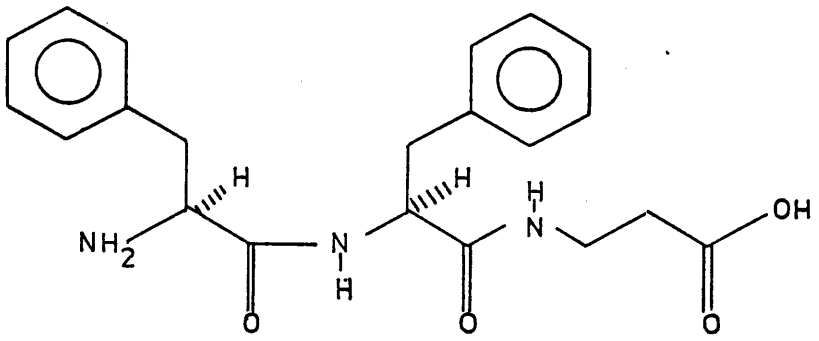
times the activity of compound 9.

From Table 6.10 it can be seen that the para-Aminobenzoate derivatives of the Phe-[N]-X type compounds, although not as potent as expected are relatively more potent than the compounds of Table 6.9.

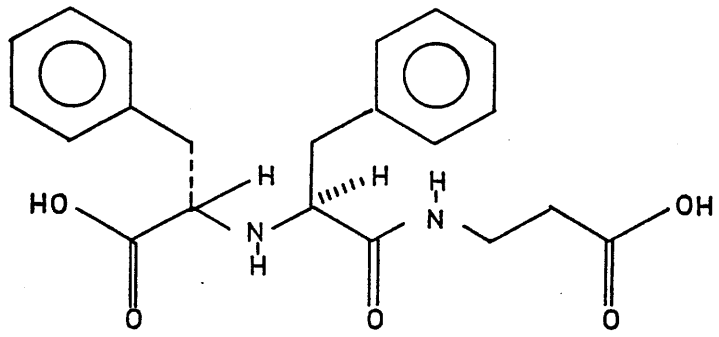
It can also be seen from this table that Thiorphan appears to be the most potent inhibitor relative to its expected potency.

The Phe-[N]-L-X type compounds are thought to be good inhibitors since they mimic the transition state of the substrates[50] (see Fig 6.15). It is expected that these inhibitors will bind to EC 3.4.24.11. in similar conformations to CLT binding in Thermolysin.[45] The Homo-Phe derivatives are preferred since the N linkage shortens the chain length to the P_1 Phenyl group which means that it is not so well placed for binding in the S_1 hydrophobic pocket (or patch) as the natural substrate would be (see Fig. 6.16). It is also interesting to note that the N linkage gives the P_1 an apparent R configuration with respect to the the natural substrate[50] (see Fig. 6.16). Thermolysin and EC 3.4.24.11. have been shown to have almost identical substrate specificities and are inhibited at similar rates by various inhibitors (see Tables 6.6 and 6.11) and even although Thermolysin is not inhibited at the same rates by the para-aminobenzoate inhibitors it shows more or less the same inhibition rates across this series of inhibitors.[35]

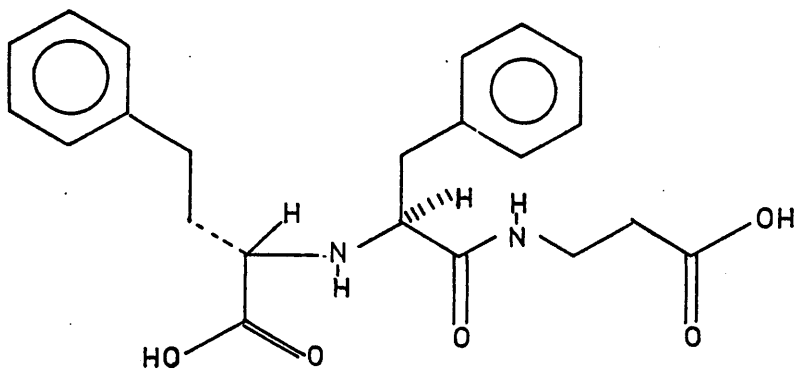
Therefore when designing the active site of EC 3.4.24.11. it was assumed that the S_1 and S_1` hydrophobic pockets of the two enzymes were in the same three-dimensional arrangements. It was also assumed that the $S_1$` pocket was the same size in Thermolysin and EC 3.4.24.11. since both enzymes have a maximum action when a Phe group occupies



L - Phe - L - Phe - βAla a natural substrate

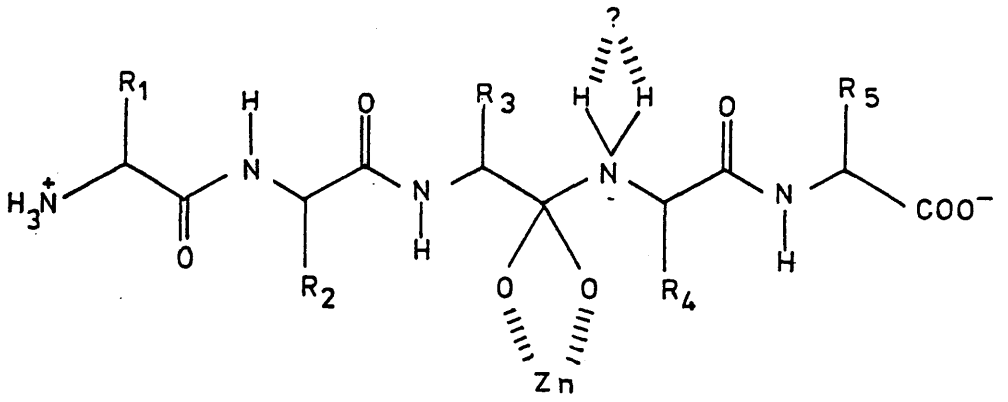


L - Phe - [N] -L - Phe - βAla

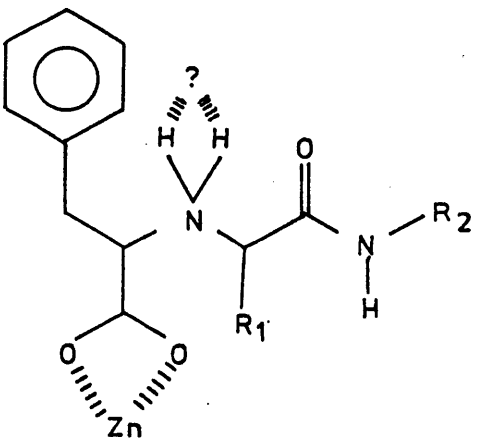


HomoPhe - [N] - L - Phe - βAla

Figure 6.16



proposed transition state for substrate



Phe-[N]-L-X type compounds mimic the transition state

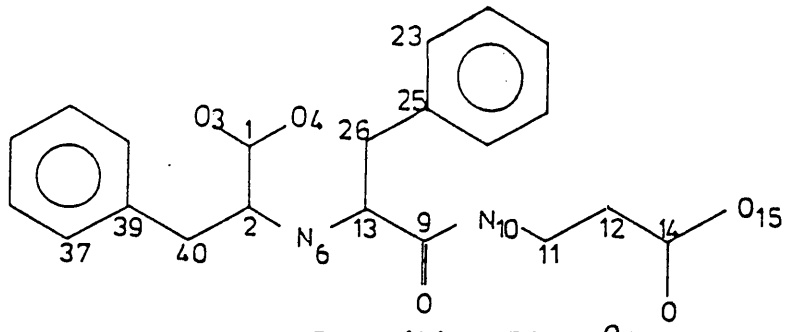
Figure 6.15

this pocket and activity is lost when larger residues such as Tyr and Trp are introduced in this position. The S_1 pocket of EC 3.4.24.11. would appear to be slightly smaller than that of Thermolysin since substrates having a Phe residue in the P_1 position are hydrolysed well by Thermolysin, however in EC 3.4.24.11 such substrates bind very well, but tend to cause a decrease in catalytic rate. Since phosphoramidon is a potent inhibitor of both compounds there must also be a hydrophobic binding site at the S_2' subsite of EC 3.4.24.11. in which the Trp residue side chain can be accomodated. The other obvious difference between the two enzymes is that Thermolysin has an arginine residue (Arg-203) in its active site which, depending upon the substrate involved, binds to the P_1' terminal carboxylate, or carbonyl group.[42] Thermolysin can therefore act as an exopeptidase, whereas EC 3.4.24.11. is unable to cleave substrates at the final peptide bond. This implies that EC 3.4.24.11. requires residues in the S_1' and S_2' subsites before cleavage can occur and it is unlikely that there is a functional group in EC 3.4.24.11. similar to Arg-203 of Thermolysin.

EC 3.4.24.11. has however shown dipeptidase action and it has been proposed that this enzyme has an arginine residue at the S_2' subsite. This may well be the case however it is possible that previous researchers have been reading too much into their results. The presence of the arginine residue was proposed for two reasons a) amidated compounds are hydrolysed at lower rates than those with a free terminal carboxylate and b) butanedione studies showed that inhibition with butandedione caused a loss of 80% activity with substrates which had a free carboxylate terminal, 50% loss with amidated derivatives and only 20% loss with naphthlyamide derivatives.

However on closer examination of these data (see Table 6.8) it can be seen that [Met^5]-enkephalin-Arg-Phe-NH_2 which has its amide group in the S_4` subsite is also hydrolysed much more slowly than the natural substrate which has a free terminal carboxylate. This could indicate that amidated compounds are in general poorer substrates and this could be due to the fact that the entrance to the active site cleft is more accessible to anionic, or neutral groups. The butanedione study is not conclusive either since, although the loss of activity with [Leu^5, D-Ala^2]-enkephalin is shown the loss of activity with [Met^5,D-Ala^2]-enkephalin is not mentioned and it is not good practise to cite the loss of activity which occurs with the amidated form [Met^5 D-Ala^2]-enkephalin when no mention is made of the loss in activity with [Met^5, D-Ala^2]-enkephalin. Also although there may be a 20% loss in activity with the napthylamide derivative no mention is made of the drop in activity for Hip-Arg-Arg-Leu and this substrate appears to have a much higher activity than the enkephalin derivatives and it is possible that it is binding in a completely different orientation to these compounds. However it is expected that there will be a functional group at the S_2` subsite capable of binding to carbonyl, or carboxlate groups and for the sake of argument this will be considered to be a arginine residue.

Low energy conformers of several of the inhibitors of EC 3.4.24.11. were generated using the procedures described in chapter 3 (see Tables 6.14 to 6.24). Since the Phe-[N]-L-X type inhibitors are expected to bind in a similar manner to CLT in Thermolysin, comparing the low energy conformers with the binding conformation of CLT may give some insight into the differences between the two enzymes. Table



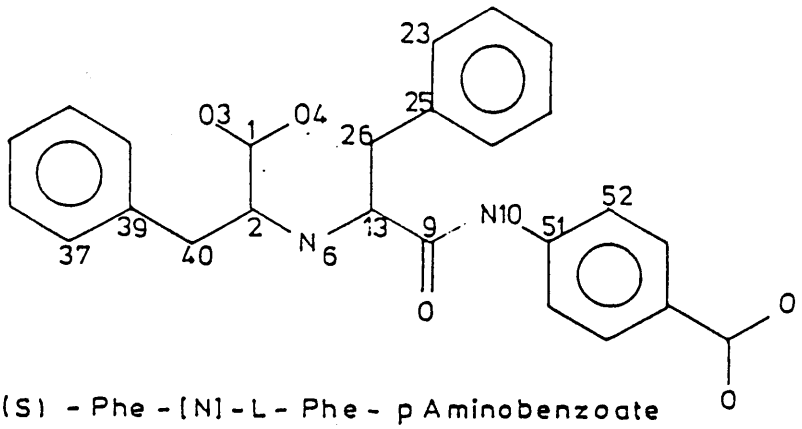
Phe - [N]-L-Phe - βAla

TORSION ANGLES

1) C37 C39 C40 C2 2) C39 C40 C2 N6 3) C40 C2 N6 C13
4) C2 N6 C13 C9 5) N6 C13 C9 N10 6) C13 C9 N10 C11
7) C9 N10 C11 C12 8) N10 C11 C12 C14 9) C11 C12 C14 O15
10) C40 C2 C1 O3 11) N6 C13 C26 C25 12) C13 C26 C25 C23

MINIMA	ENERGY kCAL/MOL	TORSION ANGLES 1	2	3	4	5	6	7	8	9	10	11	12
1	5.250	85	-68	159	-64	-53	178	131	63	114	-118	-168	75
2	5.415	84	-69	158	-63	-53	179	131	63	114	-118	-70	93
3	5.626	83	-75	168	-84	118	-178	103	-56	120	-118	-167	-104
4	6.100	95	-69	159	-78	118	-177	-124	67	116	-118	-179	-115
5	6.464	84	-68	159	-65	119	-178	111	62	115	-118	-69	92
6	6.600	89	-66	146	-83	116	-178	113	59	117	-120	62	-68
7	6.605	94	-74	161	-76	118	-176	109	61	116	-119	-178	-115
8	6.683	83	-74	166	-84	-52	-178	130	-57	120	-118	-166	-105
9	6.846	94	-74	164	-82	115	-175	100	65	117	-118	-165	-106
10	6.936	94	-70	163	-83	115	-175	-120	68	116	-118	-165	-107
11	7.041	74	174	159	79	116	-177	-124	681	116	-119	-179	-115
12	7.569	74	174	159	-79	116	-177	113	68	116	-119	179	-115
13	7.672	95	-62	165	-85	-115	-177	-117	61	115	-118	-164	77
14	7.692	75	176	162	-84	114	-175	-119	69	116	-118	-165	-107
15	7.711	73	175	164	-83	114	-175	101	66	117	-118	-165	-106
16	7.802	99	50	140	-81	3	178	133	76	116	-116	57	94
17	7.849	73	179	155	-77	115	-174	112	63	116	-188	57	-79
18	7.907	77	-179	156	-82	-6	178	-120	70	117	-119	-135	75
19	7.994	81	-171	157	-62	-53	179	131	65	117	-118	-70	93
20	8.998	89	-63	144	-83	119	-179	75	61	117	-119	62	-67
21	9.299	75	179	153	-79	108	-175	-148	64	116	-118	56	-78
22	9.375	72	173	154	-76	120	178	-121	66	117	-120	-68	172
23	9.687	92	-62	163	-86	-117	-178	110	64	116	-118	-161	-7
24	9.935	119	-46	-70	-90	127	176	175	178	-179	-171	-173	78
25	9.984	123	-83	109	-92	130	173	173	-178	-179	-175	-174	77
26	10.551	121	-47	-71	-103	129	174	174	178	-179	-172	-171	5

TABLE 6.14



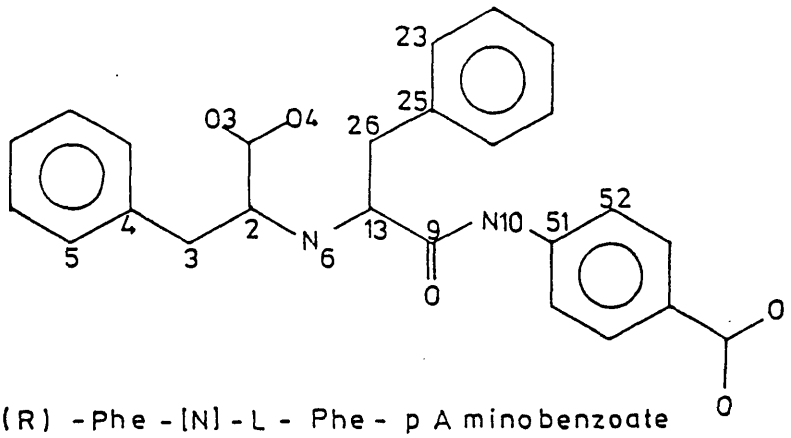
(S) - Phe - [N] - L - Phe - p Aminobenzoate

TORSION ANGLES

1) C37 C39 C40 C2 2) C39 C40 C2 N6 3) C40 C2 N6 C13
4) C2 N6 C13 C9 5) N6 C13 C9 N10 6) C13 C9 N10 C51
7) C9 N10 C51 C52 8) C40 C2 C1 O3 9) N6 C13 C26 C25
10) C13 C26 C25 C23

MINIMA	ENERGY kCAL/MOL	TORSION ANGLES 1	2	3	4	5	6	7	8	9	10
1	12.889	94	-67	162	-68	-56	-178	22	61	174	64
2	12.928	92	-67	163	-69	-54	179	21	62	-72	98
3	13.833	93	-66	161	-67	125	179	21	61	175	64
4	13.977	125	-84	107	-88	128	175	176	60	-175	77
5	14.131	92	-67	163	-70	127	179	21	62	-71	97
6	14.567	112	-82	119	-86	-86	-174	21	61	177	61
7	14.684	126	-81	111	-101	123	-177	10	60	-167	81
8	14.721	126	-85	108	-88	128	175	176	60	-175	77
9	14.930	116	-87	118	-83	-80	-174	34	-178	179	63
10	14.953	133	-88	107	-94	122	-175	5	60	-172	74
11	15.341	-167	50	74	-86	-57	179	-161	124	-173	75
12	15.726	-170	51	74	-86	-59	179	-162	-176	-172	74
13	15.893	131	-82	114	-101	120	-175	17	-124	-168	80
14	16.043	154	-83	70	-99	136	179	178	-118	-173	79
15	16.165	140	67	73	-90	-61	179	-167	-116	-178	71
16	16.227	135	-89	111	-97	117	-173	10	-175	-174	73
17	16.390	179	54	62	-93	112	-172	-173	-175	-171	-95
18	16.686	103	59	107	-82	112	-178	-164	-118	-164	-100
19	16.692	174	58	60	-90	122	-177	-158	-176	-70	-77
20	16.785	-92	66	-93	-87	120	-178	-171	58	-176	77
21	16.957	110	80	124	-77	-65	-175	-159	119	175	62
22	16.959	125	-66	100	-146	61	177	149	-118	-165	83
23	17.292	95	71	-100	-80	129	-178	-149	-125	174	67
24	17.324	158	-83	68	-97	137	-179	174	-173	175	78
25	17.446	139	68	72	-90	-61	-179	-169	-174	177	71
26	17.451	108	81	127	-80	-62	-176	-161	119	175	62
27	17.638	121	65	86	-83	-65	179	-156	-176	-79	105

TABLE 6.15



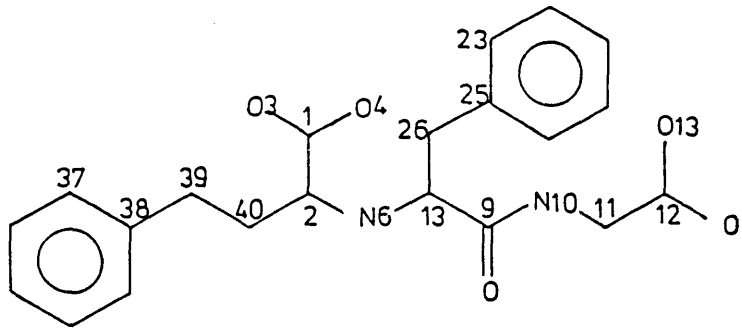
(R) -Phe -[N]- L - Phe - p Aminobenzoate

TORSION ANGLES

1) C5 C4 C3 C2 2) C4 C3 C2 N6 3) C3 C2 N6 C13
4) C2 N6 C13 C9 5) N6 C13 C9 N10 6) C13 C9 N10 C51
7) C9 N10 C51 C52 8) C3 C2 C1 O3 9) N6 C13 C26 C25
10) C13 C26 C25 C23

MINIMA	ENERGY kCAL/MOL	TORSION ANGLES 1	2	3	4	5	6	7	8	9	10
1	14.155	91	-48	-177	-91	-20	178	-175	-63	59	86
2	15.002	81	38	-177	-176	-24	-175	174	-61	57	79
3	15.595	97	168	-168	-91	-16	179	-179	-57	61	83
4	15.630	86	51	-174	-101	175	173	173	-59	59	80
5	15.755	97	168	-164	-152	-13	-179	176	-61	88	88
6	15.964	94	65	-139	-63	-58	-174	-167	-62	150	72
7	16.016	90	-49	-177	-93	169	174	177	-64	61	85
8	16.086	98	170	-165	-156	172	176	172	-61	60	89
9	16.173	100	66	-157	-170	-56	-176	-171	-61	145	72
10	16.175	101	70	179	-100	169	175	176	-54	120	110
11	16.428	86	63	-153	-151	-7	-172	-179	-62	86	136
12	16.468	97	169	-169	-93	-172	174	175	-58	62	83
13	16.772	88	39	-174	-179	173	171	171	-61	58	79
14	17.882	87	64	-146	-156	-172	164	-175	-61	95	129
15	18.046	100	63	-178	-87	-44	179	-166	-59	129	92
16	18.540	83	-71	175	-169	175	174	179	-73	-145	81
17	18.996	101	65	-175	-169	168	169	178	-61	126	101
18	19.115	94	166	-178	-95	167	169	179	-57	125	100
19	19.394	101	160	-119	99	-19	175	176	54	120	110
20	19.495	95	160	-170	-170	-36	-176	-177	-61	129	58
21	19.889	96	158	-173	-173	168	168	176	-61	125	104
22	20.564	101	159	-124	-124	-21	-174	174	-61	115	109
23	20.573	98	162	-122	-122	17	169	179	-60	128	100
24	20.884	101	158	-121	-121	173	168	178	-60	117	109

TABLE 6.16



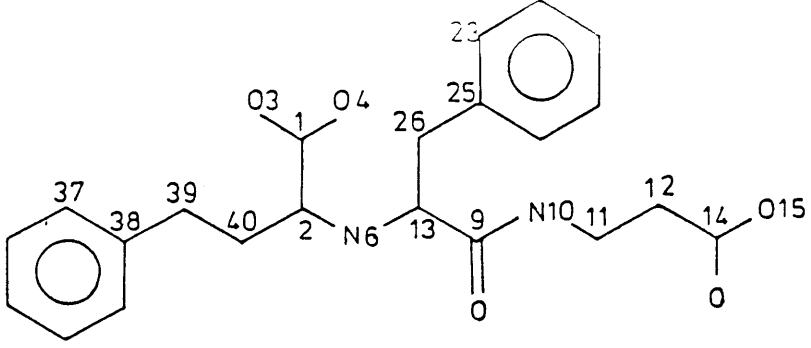
Homo - Phe - [N] - L - Phe - Gly

TORSION ANGLES

1) C37 C38 C39 C40 2) C38 C39 C40 C2 3) C39 C40 C2 N6
4) C40 C2 N6 C13 5) C2 N6 C13 C9 6) N6 C13 C9 N10
7) C13 C9 N10 C11 8) C9 N10 C11 C12 9) N10 C11 C12 O13
10) C40 C2 C1 O3 11) N6 C13 C26 C25 12)C13 C26 C25 C23

MINIMA	ENERGY kCAL/MOL	TORSION ANGLES 1	2	3	4	5	6	7	8	9	10	11	12
1	5.568	110	-62	-69	172	-70	110	-172	113	179	-118	-162	-103
2	6.359	68	65	-71	174	-64	111	-171	116	178	-118	-166	78
3	7.032	80	65	-176	174	-66	120	-179	154	179	60	-169	76
4	7.144	62	67	-67	175	-66	-70	179	176	179	-118	-168	79
5	7.352	108	-66	-69	178	-71	-59	-178	175	-179	-118	-161	-94
6	7.688	80	66	-176	175	-68	-60	-178	163	179	60	-170	76
7	7.759	66	61	-174	178	-71	76	-178	172	-177	62	-168	70
8	7.831	71	62	-179	82	-89	129	177	-126	-117	60	-161	87
9	8.072	86	179	-176	173	-67	112	-173	118	178	-119	-162	76
10	8.204	68	58	-177	87	-76	-56	-179	149	178	60	-175	83
11	8.474	105	-65	68	65	71	-60	-177	158	179	120	-174	-99
12	8.581	62	66	-66	174	-71	168	178	-170	179	-118	-148	77
13	8.717	104	-66	70	173	-71	-62	-179	165	178	-67	-159	86
14	8.759	90	-178	-178	177	-69	-62	-178	168	179	119	-165	84
15	8.806	89	-179	-67	177	-69	-61	-178	169	179	-118	-165	-93
16	9.193	117	-64	-70	176	-75	169	179	-167	-179	-118	-149	-102
17	9.242	112	-62	66	172	74	58	-177	-153	-177	-114	-167	73
18	9.246	109	-70	-175	178	-32	-60	-178	166	179	-124	-161	-93
19	9.315	163	-169	59	78	-79	-54	178	170	179	118	-170	82
20	9.448	93	178	177	75	-76	-52	177	159	178	-118	-170	86
21	9.516	73	67	-173	67	-175	-63	-175	168	179	62	-171	83
22	9.601	104	-65	70	172	-73	172	179	-173	179	-69	-146	-103

TABLE 6.17



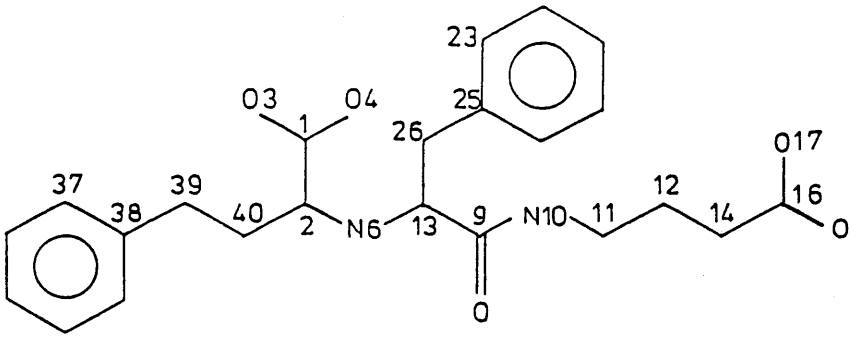
Homo - Phe - [N] - L - Phe - β Ala

TORSION ANGLES

1) C37 C38 C39 C40 2) C38 C39 C40 C2 3) C39 C40 C2 N6
4) C40 C2 N6 C13 5) C2 N6 C13 C9 6) N6 C13 C9 N10
7) C13 C9 N10 C11 8) C9 N10 C11 C12 9) N10 C11 C12 C14
10) C11 C12 C14 O15 11) C40 C2 C1 O3 12) N6 C13 C26 C25

MINIMA	ENERGY kCAL/MOL	TORSION ANGLES 1	2	3	4	5	6	7	8	9	10	11	12	13
1	4.946	64	61	-175	169	-70	114	-177	103	63	113	61	-167	79
2	5.207	68	60	-176	176	-66	112	-173	116	63	61	60	-166	75
3	5.333	69	59	-71	172	-65	109	-172	124	-60	120	62	-164	78
4	5.436	69	58	-176	173	-66	111	-171	91	-179	62	61	-166	75
5	5.495	69	65	-72	176	-61	-56	176	142	-59	116	-119	-161	81
6	5.730	68	57	-176	173	-66	112	-172	116	-177	117	61	-165	75
7	5.738	68	59	-177	173	-65	112	-171	114	-176	64	63	-166	76
8	5.811	69	65	-72	179	-61	-56	179	144	-59	68	61	-162	81
9	5.980	110	-63	-68	172	-71	110	-171	121	62	62	62	-160	76
10	6.021	110	-61	-69	172	-70	110	-172	122	-178	63	62	-159	76
11	6.023	109	-62	-69	171	-71	108	-170	126	59	119	61	-158	75
12	6.175	110	-63	-68	171	-72	111	-171	120	65	117	63	-160	76
13	6.302	69	65	-72	179	-61	-57	179	120	-179	63	61	-165	75
14	6.318	69	60	-176	173	-66	111	-172	119	-178	-179	61	-165	74
15	6.343	69	66	-70	175	-66	108	-173	123	63	61	118	-164	78
16	6.524	108	-165	-69	178	-71	62	176	93	-71	123	62	-163	75
17	6.586	69	65	-72	179	-61	-56	179	143	-178	119	61	-161	81
18	6.708	69	66	-71	173	-68	109	-172	119	179	64	62	-164	79
19	6.709	69	63	-71	173	-65	109	-171	119	-177	118	62	-164	76
20	6.712	109	-63	-69	172	-71	108	-170	126	-177	-179	118	-158	77
21	7.015	108	-65	-70	178	-70	-61	-178	92	178	117	62	-163	86
22	7.025	108	-65	-69	177	-70	-62	-177	92	176	62	62	-162	85
23	7.074	108	-61	-67	179	-71	-58	-176	169	-62	63	63	-163	87
24	7.093	108	-64	-68	178	-71	-59	-177	168	-69	119	61	-160	86
25	7.178	70	65	-72	179	-61	-57	179	144	-178	-178	60	-161	80
26	7.323	109	-61	-69	177	-71	-59	-178	170	177	118	62	-162	87
27	7.404	109	-63	-67	178	-71	-57	-178	169	176	63	61	-161	78
28	7.606	69	67	-70	172	-66	106	172	128	178	-177	60	-163	77
29	8.068	108	-65	-70	178	-71	-60	179	170	177	179	63	-162	87

TABLE 6.18



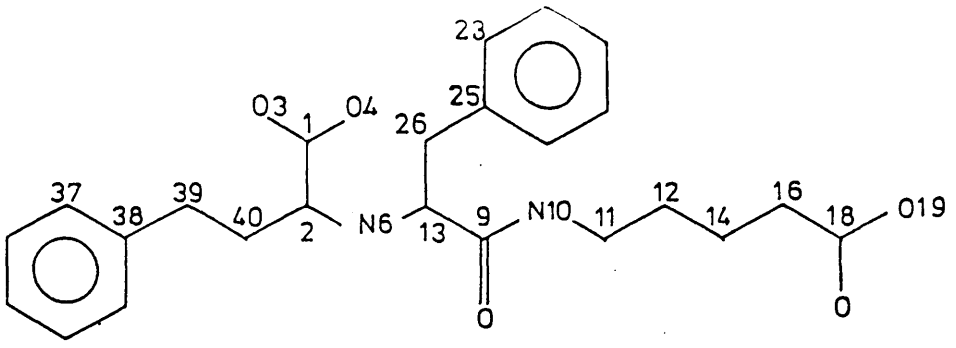
Homo-Phe-[N]-L-Phe-γ Aminobutyric acid

TORSION ANGLES

1) C37 C38 C39 C40 2) C38 C39 C40 C2 3) C39 C40 C2 N6
4) C40 C2 N6 C13 5) C2 N6 C13 C9 6) N6 C13 C9 N10
7) C13 C9 N10 C11 8) C9 N10 C11 C12 9) N10 C11 C12 C14
10) C11 C12 C14 C16 11) C12 C14 C16 O17 12) C40 C2 C1 O3
13) N6 C13 C26 C25 14) C13 C26 C25 C23

MINIMA	ENERGY kCAL/MOL	TORSION ANGLES 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	4.909	67	57	-177	172	-65	115	-174	117	64	-173	118	160	-166	79
2	5.354	69	56	-176	172	-64	115	-175	122	-175	61	61	60	-167	78
3	5.528	68	57	-178	171	-66	1121	-177	73	-176	178	118	60	-168	81
4	5.587	67	65	-70	173	-65	109	-175	104	-78	-66	118	-118	-165	78
5	5.590	66	57	-176	172	-65	116	-174	123	71	-174	68	61	-167	79
6	5.789	110	-63	-69	171	-71	110	-170	120	-176	-179	-119	-118	-160	-104
7	5.814	67	57	-171	172	-65	114	-175	125	-174	-170	62	60	-166	78
8	5.828	110	-63	-68	171	-70	111	-170	119	62	-176	118	-118	-160	-103
9	5.910	67	57	-178	172	-65	116	-175	124	67	62	60	62	-167	80
10	5.948	66	57	-177	172	-66	114	-176	120	-178	-60	117	61	-166	78
11	6.069	67	55	-175	173	-67	114	-175	125	-173	-60	117	61	-166	78
12	6.081	65	56	-177	172	-65	112	-175	121	-178	84	121	63	-164	79
13	6.206	109	-63	-69	171	-71	111	-170	120	62	61	-117	-118	-160	-104
14	6.271	111	-61	-68	171	-71	110	-171	121	-179	178	-179	62	-159	76
15	6.301	67	57	-177	171	-63	116	-173	115	64	78	117	60	-167	79
16	6.492	68	66	-71	172	-65	109	-173	118	-85	-63	119	-117	-164	78
17	6.591	69	65	-71	173	-64	109	-172	105	-178	-61	117	-118	-164	77
18	7.125	90	179	-175	67	-174	175	-175	-174	-179	180	180	61	-149	70
19	7.321	67	66	-71	173	-65	109	171	120	-179	-177	-179	62	-164	78

TABLE 6.19



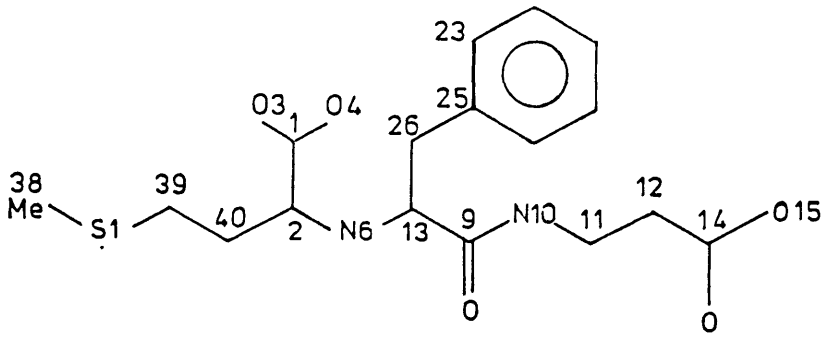
Homo-Phe-[N]-L-Phe-δAminovaleric acid

TORSION ANGLES

1) C37 C38 C39 C40 2) C38 C39 C40 C2 3) C39 C40 C2 N6
4) C40 C2 N6 C13 5) C2 N6 C13 C9 6) N6 C13 C9 N10
7) C13 C9 N10 C11 8) C9 N10 C11 C12 9) N10 C11 C12 C14
10) C11 C12 C14 C16 11) C12 C14 C16 C18 12) C14 C16 C18 O19
13) C40 C2 C1 O3 14) N6 C13 C26 C25 15) C13 C26 C25 C23

MINIMA	ENERGY kCAL/MOL	TORSION ANGLES														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	4.151	110	-63	-67	172	-71	112	-172	114	62	-178	65	61	-120	-162	-103
2	4.745	110	-63	-68	172	-71	112	-172	115	60	-178	68	-178	-118	-161	-103
3	5.164	67	57	-178	171	-65	116	-174	122	63	173	-60	117	60	-167	80
4	5.544	68	65	-72	178	-61	-55	-179	140	-59	-176	-61	119	-118	-163	82
5	5.552	67	57	-178	171	-65	116	-174	122	63	-172	179	119	61	-167	80
6	5.779	68	65	-71	174	-64	113	-172	110	63	-178	69	112	-116	-166	78
7	6.010	68	65	-72	179	-61	-55	179	140	-60	-177	179	117	-118	-163	82
8	6.143	67	57	-178	171	-65	115	-176	126	-175	-168	179	121	60	-167	79
9	6.154	68	65	-71	173	-64	111	-171	115	62	-175	178	117	-116	-165	78
10	6.382	110	-63	-69	171	-70	111	-170	120	-175	179	75	117	-118	-160	-103
11	6.408	68	65	-72	179	-61	-55	179	140	-60	-177	179	-179	-119	-163	82
12	6.464	110	-63	-69	172	-70	110	-170	120	-175	-179	-177	179	-118	-160	-104
13	6.525	68	65	-71	173	-64	110	-172	117	-179	-177	61	-119	-118	-165	78
14	6.586	110	-63	-69	171	-70	110	-170	119	61	-178	-179	179	-118	-160	-103
15	6.592	67	65	-72	178	-62	-55	178	140	-58	101	-177	123	-117	-164	86
16	6.701	109	-65	-69	178	-70	-59	-177	91	65	-177	-178	117	-119	-164	-94
17	6.819	109	-65	-69	178	-70	-60	-178	91	-85	179	178	118	-117	-163	-93
18	7.172	109	-65	-69	177	-70	-60	-177	92	177	-179	-179	118	-118	-162	-92
19	7.207	68	65	-72	179	-61	-56	178	141	-179	-177	-179	-179	-119	-163	82
20	7.248	109	-65	-69	178	-71	-59	-178	170	177	-179	61	-119	-118	-164	-93
21	7.351	68	66	-71	173	-64	110	-172	117	-179	-176	-179	-178	-118	-165	78
22	8.188	109	-65	-69	179	71	-59	178	169	177	-178	179	-179	118	-164	-94
23	8.213	109	-65	-69	178	-72	-57	-178	169	176	-179	60	-120	-117	-169	-95

TABLE 6.20



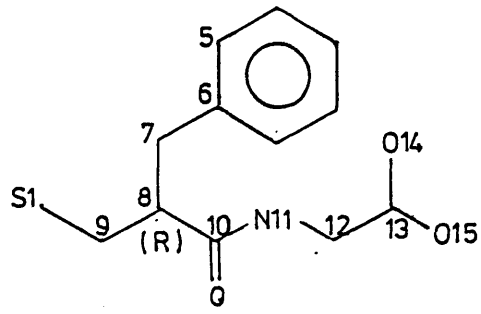
Met - [N] - L - Phe - β Ala

TORSION ANGLES

1) C38 S1 C39 C40 2) S1 C39 C40 C2 3) C39 C40 C2 N6
4) C40 C2 N6 C13 5) C2 N6 C13 C9 6) N6 C13 C9 N10
7) C13 C9 N10 C11 8) C9 N10 C11 C12 9) N10 C11 C12 C14
10) C11 C12 C14 O15 11) C40 C2 C1 O3 12) N6 C13 C26 C25
13) C13 C26 C25 C23

MINIMA	ENERGY kCAL/MOL	TORSION ANGLES 1	2	3	4	5	6	7	8	9	10	11	12	13
1	2.563	80	174	-175	149	-85	116	-176	111	-61	118	-120	-167	-104
2	3.137	75	73	-66	147	-87	115	179	114	65	116	-119	-167	-103
3	3.384	-177	179	-66	147	-87	116	-177	111	61	116	-120	-168	-104
4	3.502	78	175	-174	148	-85	116	-178	110	62	115	-118	-167	-103
5	3.635	74	77	-63	149	-83	115	-177	111	-178	118	-118	-168	-104
6	3.659	80	176	-62	150	-83	115	-176	110	-178	118	-118	-168	-104
7	3.971	79	174	-174	149	-85	114	-177	111	-178	118	-120	-167	-104
8	4.401	75	73	-66	147	-87	67	179	-116	65	116	-119	-167	-103
9	4.580	-178	177	65	145	-87	66	179	-115	65	116	-119	-166	-103
10	4.580	70	60	-178	75	-150	115	-176	109	61	116	-118	-162	-102
11	4.626	73	60	-172	153	-143	115	-177	111	-178	118	-118	-168	-104
12	4.676	72	173	-177	156	-147	116	-179	113	-60	118	-118	-158	-109
13	4.837	76	174	-172	152	-142	115	-177	109	62	116	-118	-164	-104
14	4.858	78	177	-170	144	-87	65	178	-113	64	115	-119	-166	-103
15	4.887	-74	171	-178	76	-149	115	-176	110	60	116	-118	-162	-103
16	4.888	-98	65	-171	78	-153	114	-175	109	61	116	-118	-164	-100
17	4.939	71	171	-179	75	-149	115	-176	110	60	116	-118	-161	-103
18	5.131	76	66	-178	74	-147	59	178	-116	65	116	-118	-160	-105
19	5.182	78	178	-170	143	-86	56	179	-113	-58	115	-118	-165	-104
20	5.306	80	176	-169	151	-84	-115	174	-127	58	116	-118	-171	-109
21	5.346	77	175	-113	149	140	58	175	-112	60	115	-118	-162	-104
22	5.362	79	173	-170	143	-87	59	178	-112	-178	118	-119	-165	-104
23	5.376	73	173	-178	155	-146	178	178	-67	-57	116	-116	-156	-110
24	5.392	72	61	-177	155	-146	178	178	-67	-57	117	-118	172	-110
25	5.416	-74	172	179	73	-147	58	178	-116	65	116	-118	-160	-110
26	5.466	75	172	179	73	-147	58	178	-116	65	116	-118	-160	-105

TABLE 6.21



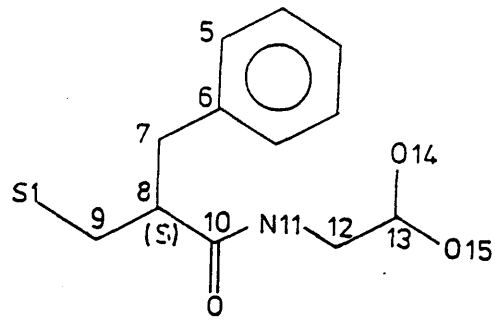
(R) - Thiorphan

TORSION ANGLES

1) S1 C9 C8 C10 2) C9 C8 C10 N11 3) C8 C10 N11 C12
4) C10 N11 C12 C13 5) N11 C12 C13 O14 6) C9 C8 C7 C6
7) C8 C7 C6 C5

MINIMA	ENERGY kCAL/MOL	TORSION ANGLES 1	2	3	4	5	6	7
1	4.259	58	-116	-178	116	119	166	88
2	4.282	57	-114	177	118	57	167	82
3	4.553	62	67	175	-102	65	-168	133
4	4.612	168	-120	-178	134	177	171	102
5	4.798	168	-121	-176	140	179	169	88
6	4.857	61	-119	-179	-105	62	178	-95
7	4.901	60	62	-178	121	119	168	85
8	5.077	167	62	177	-139	179	177	94
9	5.195	61	-116	178	-108	63	73	-99
10	5.260	167	62	176	-141	-60	-165	-101
11	5.395	60	62	-178	-139	178	169	88
12	5.492	167	62	-178	136	178	167	89
13	5.866	167	-114	177	-163	178	162	87
14	6.007	64	-115	178	-167	179	-58	86
15	6.591	58	62	178	110	118	-149	122
16	7.565	168	-151	177	102	179	137	88
17	8.169	-58	-169	-178	174	179	-89	88
18	8.771	172	20	-179	141	178	-65	85

Table 6.22



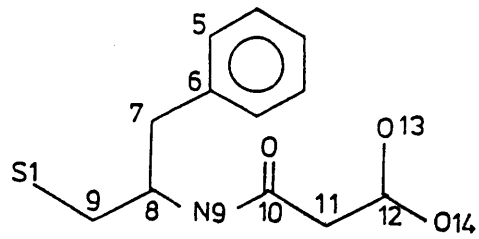
(S) - Thiorphan

TORSION ANGLES

1) S1 C9 C8 C10 2) C9 C8 C10 N11 3) C8 C10 N11 C12
4) C10 N11 C12 C13 5) N11 C12 C13 O14 6) C9 C8 C26 C25
7) C8 C26 C25 C23

MINIMA	ENERGY kCAL/MOL	TORSION ANGLES 1	2	3	4	5	6	7
1	3.773	-63	119	176	-109	62	-174	80
2	3.861	-63	119	176	-125	62	-174	80
3	3.897	-63	119	176	-109	119	-174	80
4	4.212	-61	119	178	-130	61	-172	94
5	4.605	-61	-60	179	119	118	-166	-101
6	4.623	-60	-62	-178	117	60	170	92
7	4.664	-63	117	179	117	118	-167	-100
8	4.671	-167	120	177	-129	-178	-172	94
9	4.695	-61	-62	178	120	112	-170	91
10	4.691	-169	125	176	-128	-179	-164	89
11	4.723	-62	-60	-179	118	59	-165	-101
12	5.008	-60	-63	179	-130	62	-171	94
13	5.247	-167	-62	-179	158	179	-178	81
14	5.366	-167	-63	-176	148	178	-171	93
15	5.508	-167	121	177	154	178	-172	94
16	5.572	-167	-64	-177	-130	-177	-172	95
17	5.803	-167	113	-177	177	-178	-163	91
18	6.133	-155	121	176	-155	-178	-71	106
19	6.239	-171	167	178	-156	-179	76	94
20	6.345	-151	115	179	-130	-177	-72	99
21	6.347	-151	-65	-179	-131	-178	-72	100
22	6.404	-61	-58	176	-124	-42	-170	43
23	6.467	-151	-63	-179	153	179	-72	100
24	6.490	-151	117	-179	154	178	-72	96
25	6.562	-150	-62	-179	149	178	-73	97

TABLE 6.23



Retrothiorphan

TORSION ANGLES

1) S1 C9 C8 N9 2) C9 C8 N9 C10 3) C8 N9 C10 C11
4) N9 C10 C11 C12 5) C10 C11 C12 O13 6) C9 C8 C7 C6
7) C8 C7 C6 C5

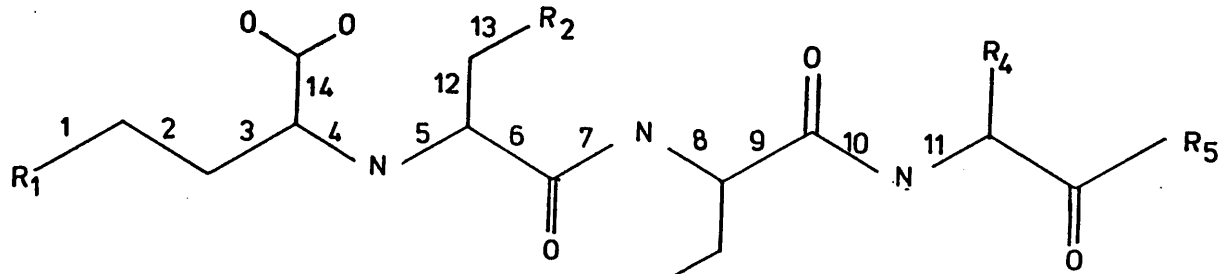
MINIMA	ENERGY kCAL/MOL	TORSION ANGLES 1	2	3	4	5	6	7
1	2.703	74	-129	-178	113	114	178	106
2	2.855	159	-130	-178	114	115	178	106
3	3.431	72	-119	-179	-61	117	-178	104
4	3.436	69	-117	179	-61	117	-179	101
5	3.848	71	38	178	-56	116	-59	81
6	3.885	65	61	179	-60	116	-179	99
7	3.923	155	38	178	-56	116	-59	81
8	3.966	-51	-117	-179	-61	117	-178	104
9	4.070	157	-74	177	-59	117	-53	96
10	4.157	70	36	178	120	117	-60	82
11	4.179	100	-80	178	56	117	-53	98
12	4.303	110	-86	178	-58	116	-60	95
13	4.316	156	36	178	120	116	-59	81
14	4.453	128	-72	179	-61	116	60	94
15	4.536	-33	-71	178	-58	117	-58	97
16	4.546	66	60	179	120	116	-179	99
17	4.566	157	-75	179	0	115	-53	95
18	4.742	-51	-69	177	-61	117	58	96
19	4.800	-56	-67	179	-61	117	-178	103
20	4.835	76	57	178	-60	118	55	97
21	4.865	128	57	178	-60	118	55	95
22	5.107	-82	75	178	-65	116	173	101
23	5.107	64	60	177	2	116	-178	97

TABLE 6.24

6.25. contains a list of the torsion angle values for the lowest energy conformer of each of the Phe-[N]-L-X inhibitors compared with the torsion angles of the binding conformation of CLT. The majority of the lowest energy conformers showed at least some similar torsion angle values when compared with CLT especially for the binding of the phenyl group in the S_1` subsite and it was noticable that Phe-[N]-L-Phe-pAB gave the best agreement. The low energy conformations of each of the inhibitors are discussed in more detail in the next section.

(S) Phe-[N]-L-Phe-pAB

The torsion angle values of the lowest energy conformer of this compound are in reasonable agreement with the torsion angles from the binding conformation of CLT. This compound would therefore be expected to be able to bind well to Thermolysin in a low energy conformation. However it is found to have a K_i with Thermolysin of 10^{-6} as opposed to 10^{-8} in EC 3.4.24.11. The reason for the decrease in activity was proposed to be due to the loss of binding energy, since Thermolysin has no functional group to bind to the terminal carboxylate of this compound whereas EC 3.4.24.11. is predicted to have an arginine at the S_2` subsite. However, when the minimum energy conformer of Phe-[N]-L-Phe-pAB is superimposed upon CLT (see Fig 6.17) it can be seen that the pAB group is oriented towards the hydrophobic area of the $S_2$` subsite and due to the rigid nature of this group it is involved in bad interactions with the sides of this hydrophobic area. If the pAB side chain is to be accommodated in this hydrophobic region there must either be a considerable displacement in position of the rest of the molecule, or there must be major changes in the backbone torsion angles. Both of these rearrangements would probably



Torsion Angles

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
CLT	-81	-177	-90	176	-50	-43	-176	-146	8			-171	60	79
Phe-N-L-Phe- βAla		85	-68	159	-64	-53	178	131	63	114		-168	75	62
S-Phe-N-L-Phe-pAB		94	-67	162	-68	-56	-178	22				174	64	61
R-Phe-N-L-Phe-pAB		91	-48	-177	-91	-20	178	-175				59	86	-63
HomoPhe-N-L-Phe-Gly	-70	-62	-69	172	-70	110	-172	113	179			-162	77	62
HomoPhe-N-L-Phe- βAla	64	61	-175	169	-70	114	-177	103	63	113		-167	79	61
HomoPhe-N-L-Phe- γABA	67	57	-177	172	-65	116	-174	117	64	-173	118	-166	79	160
HomoPhe-N-L-Phe- δAVA	-70	-63	-67	172	71	112	-172	114	62	-178	65	-162	77	60
Met-N-L-Phe- βAla	-100	174	-175	149	-85	116	-176	111	-61	118		-167	76	60

TABLE 6.25

Superimposition of the lowest energy conformation of Phe-N-L-Phe-pAB and CLT from Thermolysin

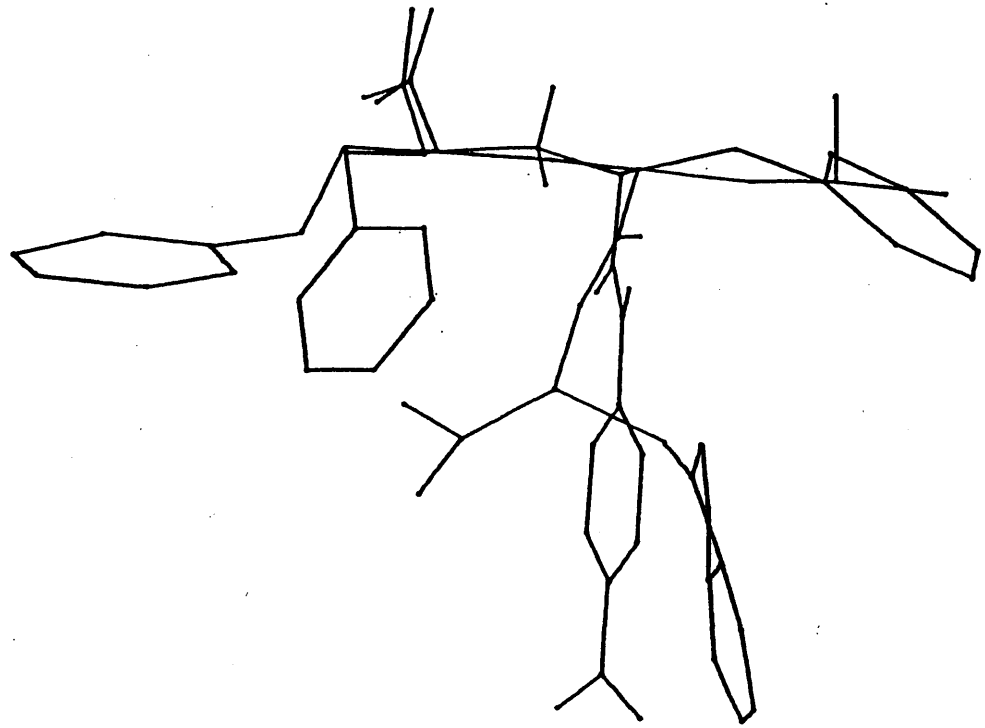


Figure 6.17

lead to a decrease in binding energy and a resulting loss in potency. It is also possible that the terminal carboxylate group is causing bad interactions with the end of the hydrophobic pocket. Therefore the active site of EC 3.4.24.11. must have a hydrophobic pocket at the S_2` which is both large enough to accommodate the 2-napthylamide derivatives and is also in a more favourable orientation to accommodate the p-aminobezoate derivatives. This is the first difference found to exist between the two active sites from examination of the low energy conformers.

Homo-Phe-[N]-L-Phe-Gly, Homo-Phe-[N]-L-Phe-βAla, Homo-Phe-[N]-L-Phe-γABA, Homo-Phe-[N]-L-Phe-δAVA and Phe-[N]-L-Phe-βAla.

A comparison of the low energy conformers of these compounds with the binding conformation of CLT shows that the P_1` phenyl ring is again optimally placed to bind with the hydrophobic pocket in a low energy position. Torsion angle values of roughly 90°, -90°, 180° and -60° for torsion angles 1, 3, 4 and 5 respectively are often found in the low energy conformations. However, values in the region of 180°, -70° and -120° for torsion angles 2, 6 and 8 are associated with high energy conformations. These torsion angles are very important since the value of torsion angle 2 dictates the position of the phenyl group in the P_1 hydrophobic pocket. The value of torsion angle 6 determines the orientation of the P_1` carbonyl group (which in Thermolysin is binding to Arg-203) and the value of torsion angle 8 determines the orientation of the $P_2$` side chain which is occupying the S_2` hydrophobic pocket. Assuming the active sites of EC 3.4.24.11. and Thermolysin are similar it is possible that the above compounds are

lower in potency than expected since they have to adopt relatively high energy conformations in order to maximise binding to the active site.

Met-[N]-L-Phe- βAla

As with other inhibitors the most favourable position for the phenyl of the P_1' Phe residue is where the two torsion angles are equal to -170° and 75° respectively. As was discussed previously this compound may be relatively more potent than the rest of this series of compounds and this could be related to the fact that the values found for torsion angles 1 to 5 of CLT are also found in the low energy conformers of Met-[N]-L-Phe- βAla. Again, assuming that the S_1 Hydrophobic pockets of Thermolysin and EC 3.4.24.11. are in similar orientations, Met-[N]-L-Phe- βAla, although unable to maximise binding to the S_1 pocket, does not need to adopt a high energy conformation to do so and may well be able to bind in a relatively lower energy conformation than the other compounds in the series.

(R)-Phe-[N]-L-Phe-pAB

As previously mentioned the N linkeage of the Phe-[N]-L-X series of compounds inverts the stereochemistry of the P_1 Phe (see Fig. 6.16). Therefore (R)-Phe-[N]-L-Phe-pAB is stereochemically equivalent to L-Phe of a normal substrate.

From Fig. 6.25 we see that the values of torsion angles 4 to 8 of the lowest energy conformer found for (R)-Phe-[N]-L-Phe are in reasonable agreement with those of CLT, but the values associated with the P_1' phenyl side chain are 59° and 86°, these are quite different from the values -170° and 80° which would be associated with binding in

Thermolysin. From Table 6.16 it can be seen that a value of -170° is not found in any of the low energy conformers of this compound.

Torsion angles 2 and 3 of (R)-Phe-[N]-L-Phe-pAB also seem to be in reasonable agreement with those of CLT however this is misleading since as previously mentioned this Phe has inverted stereochemistry to the other inhibitors and in fact if this phenyl group is to occupy the P_1 hydrophobic patch torsion angles 2 and 3 would have to adopt values of 80° and 120° respectively, values which are not found for any of the low energy conformers in Table 6.16. This may indicate that compounds containing R stereochemistry at this centre may have to undergo large conformational rearrangements to bind in EC 3.4.24.11 and this may partly explain why such compounds are generally two orders of magnitude less potent than their corresponding S isomers.

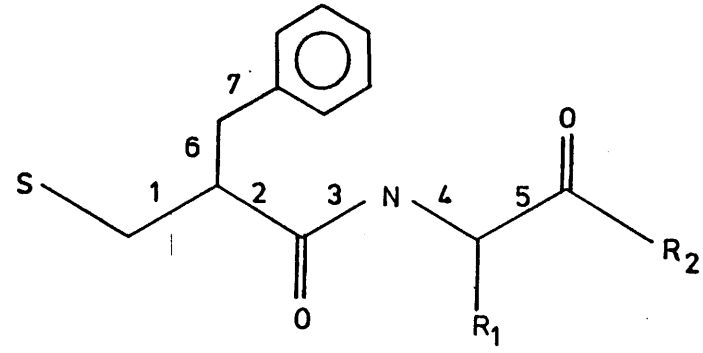
(S)-Thiorphan, (R)-Thiorphan and Retrothiorphan

The R and S isomers of Thiorphan are more potent inhibitors of EC 3.4.24.11. than Thermolysin (see Table 6.10). There may be many explanations for this phenomenon eg. a) Thiorphan binds to both enzymes in a similar manner, but it is better able to utilise the binding groups of EC 3.4.24.11. than of Thermolysin b) Thiorphan may be able to bind to EC 3.4.24.11. in a lower energy conformation than it can bind to Thermolysin c) the terminal carboxylate of Thiorphan may be able to form a salt linkage with the proposed arginine at the S_2' subsite of EC 3.4.24.11. giving stronger binding and greater potency. Retrothiorphan is also a potent inhibitor of EC 3.4.24.11. and several low energy conformers of these three compounds were generated (see Tables 6.22 to 6.24).

The X-ray conformation of a similar inhibitor of Thermolysin, (2-benzyl-3-mercaptopropanyl)-L-Ala-glycinamide, BAG, is known and Fig 6.18 shows a comparison of the torsion angles values for the binding conformation of BAG in Thermolysin with the lowest energy conformer of each of the three inhibitors: (S)-Thiorphan, (R)-Thiorphan and Retrothiorphan.

For (S)-Thiorphan the torsion angles for the Phe side chain are in good agreement with those of the previously mentioned inhibitors (-174°, 80°). However, they are not exactly the same as those found for BAG since in BAG the Phenyl ring torsion angle has a value of 34°. This torsion angle value is necessary if BAG is to optimise binding to both the zinc and the S_1' hydrophobic pocket and it is possible that when Thiorphan binds in EC 3.4.24.11. it will also need to adopt a similar position. Comparing the other torsion angle values we see that the main differences between BAG and the minimum energy conformer of (S)-Thiorphan occur for torsion angles 2 and 4. In BAG these torsion angles have values of -69° and 115° and these are values associated with higher energy conformers of (S)-Thiorphan (see Table 6.23).

(R)-Thiorphan is also a potent inhibitor of EC 3.4.24.11. This is slightly unexpected since this enzyme is known to be highly specific for those substrates possessing the naturally occuring L (equivalent to S stereochemistry) amino acid residue a the P1' subsite. To superimpose (R)- Thiorphan upon BAG requires that the torsion angle values for the phenyl side chain (torsion angles 6 and 7 of Fig. 6.18) have values of 150° and 80°. From Table 6.22 it can be seen that similar values are found for several low energy conformers.



Torsion Angles

	1	2	3	4	5	6	7
BAG	-76	-69	164	-115	-23	-163	35
(S)-Thiorphan	-63	119	176	-109	62	-174	80
(R)-Thiorphan	58	-116	-178	116	119	166	88
Retrothiorphan	74	-129	-178	113	114	178	106

Figure 6.18

Due to inversion of the amide bond in Retrothiorphan the isomer which corresponds to the S stereochemistry of Thiorphan has R stereochemistry in Retrothiorphan. On comparing the torsion angles of Retrothiorphan and BAG it can be seen that once again the phenyl ring is optimally placed to bind in the S_1' subsite. Since the amide bond is inverted the values of torsion angles 2 and 4 would have to be ~ 120° and ~ -60° to give the correct directionality to the hydrogen bonding groups. From Table 6.24 it can be seen that although a value of -60° is found for torsion angle 4 for one of the higher energy conformers, none of the conformers listed possess a torsion angle of 120° for torsion angle 2 and if Retrothiorphan is to bind to EC 3.4.24.11. in a comparable mode to BAG binding in Thermolysin it must bind in a relatively high energy conformation.

Discussion

Comparisons of the low energy conformers of the inhibitors of EC 3.4.24.11. with binding conformations of similar inhibitors of Thermolysin show that there are several possibilities for the structure of the active site of EC 3.4.24.11.

There is a distinct possibility that the active site of the renal endopeptidase shows a high structural similarity to that of Thermolysin having a slightly smaller S_1 hydrophobic pocket, a zinc atom involved in coordination to the carbonyl of the scissile bond, a functional group similar to Ala 113 of Thermolysin which will form a hydrogen bond to the nitrogen of the scissile bond, an S_1' hydrophobic pocket of exactly the same size as that of Thermolysin and an S_2' hydrophobic patch which is larger and slightly displaced from that of Thermolysin. It is not clear

whether there are any residues in the active site of EC 3.4.24.11. capable of binding to the P_1' amide bond since the enzyme is unable to hydrolyse terminal residues - unlike Thermolysin - and inversion of this amide bond, as in Retrothiorphan, seems to make little difference to the potency of the compounds. In addition all these groups would have the same spatial arrangements as those of Thermolysin giving a high homology for the three dimensional geometries of the functional groups of both compounds. Such similarity between a bacterial enzyme and a mammalian enzyme is not unlikely since enzymes which have similar substrate specificities and functions would be expected to present similar catalytic surfaces to the incoming substrates.

Another plausible structure for the active site of EC 3.4.24.11. involves a different mode of inhibitor binding than is found for Thermolysin. Several of the inhibitors studied are much more potent in EC 3.4.24.11. than in Thermolysin and it is possible that these inhibitors are binding to the endopeptidase in lower energy conformations than would be found for Thermolysin. For nearly all the inhibitors studied it was found that the major differences between the low energy conformers and the corresponding Thermolysin inhibitor binding conformation occured for the values of the two backbone torsion angles on either side of the P_1' amide bond. For Thermolysin inhibitors the values for these two torsion angles would have to be roughly -60° and -120°. From Tables 6.14 to 6.24 it can be seen that for the low energy conformers there are two pairs of values for these torsion angles which seem to be highly favourable, namely (120°, 120°) and (-60°,120°). It is important to note that the first value dictates the directionality of the P_1' carbonyl group - the binding of which would appear to be more

important in Thermolysin than EC 3.4.24.11. The second value dictates the orientation of the P_2' side chain.

From a comparison of CLT and (S)-Phe-[N]-L-Phe-pAB (see Fig. 6.17) it would appear that the para-aminobenzoate group is occupying the S_2' hydrophobic patch in both Thermolysin and EC 3.4.24.11. Since the para-aminobenzoate group is totally rigid modeling of the other inhibitors on this group should give the torsion angles necessary for binding to EC 3.4.24.11. with the P_2' side chain occupying the S_2' hydrophobic patch.

(S)-Thiorphan was chosen as a likely candidate for this comparison since it was relatively more potent than expected (see Table 6.10) and was also known to be a more potent inhibitor of EC 3.4.24.11. both factors which may indicate binding of a relatively low energy conformation.

An (S)-Thiorphan conformation was built which had a value of -60° for torsion angle 2 (see Fig. 6.18). On superimposition of this conformation with the para-aminobenzoate group it was found that torsion angle 4 had to adopt a value of 80° or -100° to align a P_2' side chain with the para-aminobenzoate group. A torsion angle 4 value of 120° would not place the side chain in the correct direction and the binding of low energy conformers with the torsion angle pair (-60°,120°) can therefore be discounted as a possibility.

A Thiorphan conformation was then built with a value of 120° for torsion angle 2. when this was superimposed upon the para-aminobenzoate group it was found that torsion angle 4 had to adopt a value of -120° or 60° to position the side chain correctly.

There are therefore four possible pairs of torsion angle values for the torsion angles to either side of the P_1' amide bond which would correctly position a P_2' side chain in the S_2' hydrophobic region, namely (-60°,-100°), (-60°,80°), (120°,-120°) and (120°,60°). The values (-60°,-100°) are similar to those found for the binding conformations of the Thermolysin inhibitors and as previously discussed are associated with high energy conformers as are the values of (-60°,80°) and (120°,60°). Whilst it is possible that the inhibitors could bind with any of these values , if they bind in low energy conformation they may adopt the values (120°,-120°). These values occur in the lowest energy conformer found for (S)-Thiorphan - the lowest energy conformer of Retrothiorphan and the third lowest energy conformer of (R)-Thiorphan are also very similar in structure to this (see Figs. 6.19 and 6.20) and similar torsion angle values are found for low energy conformers of all the other inhibitors.

During the time I spent working at Pfizer's low energy conformers of three other potent inhibitors were also generated. These compounds were not similar in structure to any of those described in this thesis, however it was found that the low energy conformers of these compounds had similar functional groups in similar positions to those conformers discussed above.

The active site prediction which arose from the modeling of low energy conformers was still very similar to that of Thermolysin except that its S_2' hydrophobic patch was in the region which corresponds to the position of a P_2' side chain when the backbone torsion angles on either side of the P_1' amide bond have values of 120° and -120° respectively. The S_1' hydrophobic pocket was identical to that of

A Superimposition of Low Energy Minima of Retrothiorphan and Thiorphan

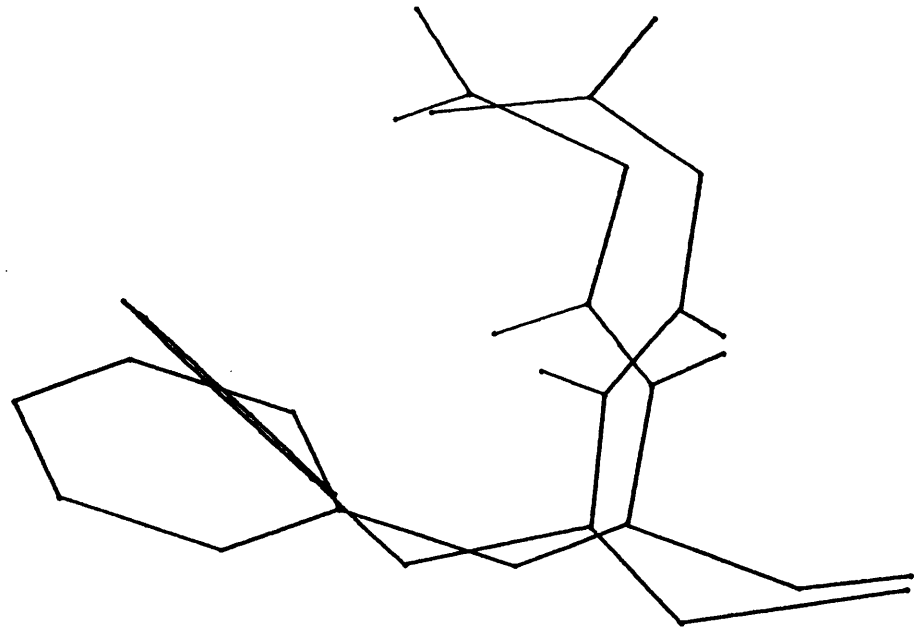


Figure 6.19

A Superimposition of Low Energy Minima of (R)- and (S)- Thiorphan

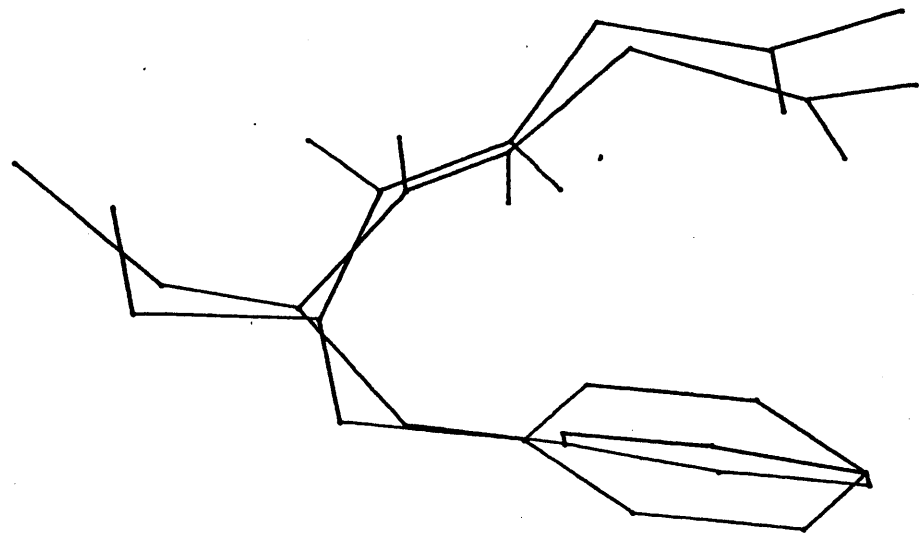


Figure 6.20

Thermolysin and there were still marked similarities in the position of the S_1 hydrophobic patch , the zinc and the functional group which binds to the nitrogen of the scissile bond.

Future Work

There is, of course, no way of knowing if either of the two active site predictions discussed in the previous section bear any resemblence to the active site of EC 3.4.24.11. since the enzyme has so far proved difficult to crystallise and very little is known of its structure.

Assuming that there will be future work on modeling the active site of EC 3.4.24.11. from Thermolysin there are two approaches which could provide valuable information on the active site of the endopetidase.

The first approach would involve a direct comparison of the kinetic parameters of several series of substrates and inhibitors of both Thermolysin and EC 3.4.24.11. This would give much information as to the differences in the subsites of the two enzymes and particular attention could be paid to the preference shown by each enzyme at the S_1' amide binding site and the S_2' hydrophobic binding patch as it would appear that there may be major differences between the two enzymes in these regions.

The second approach could involve the use of novel inhibitors. Such compounds could contain rigid groups which would restrict the allowed conformation of the molecule, carbon carbon double bonds in their backbones to explore the backbone torsion angles required for binding, replacement of the P_1' amide bond by an ethylene group or methylation of the amide nitrogen to investigate whether or not this amide takes part in any hydrogen bonding interactions, bulky side chains could be used

to test the size and direction of the S_2' hyrophobic region and since (R)-Thiorphan is as potent as (S)-Thiorphan, D-amino acids could be substituted for their naturally occuring L counterparts to see if this phenomenon occurs elsewhere.

Chapter Seven

Section 7.1.

Introduction

Much computer programming was done during the course of this project. This included debugging and modifying many existing programs, writing and converting programs to run in the FPS-100 array processor and some graphics programming.

Section 7.2.

Program Development

The hardware and software used for this project has been discussed in detail in chapter 4. The MOL software package was originally used for modeling small molecules and cyclic peptides. When any software is used for a new application bugs tend to surface in the programs and several were found in the MOL system. Most of the problems which arose were caused by the size of the molecules beind studied and tended to fall into two catagories. Either the routines were not dimensioned to cope with such large, flexible molecules and all the associated variables, or the algorithms used were not particularily efficient. Therefore the system had to be extensively modified to cope with the applications described in this thesis.

One of the major changes made to the MOL system was the implementation of a new superimposition program with which several points could be fitted using a least squares fitting procedure in an iterative manner.

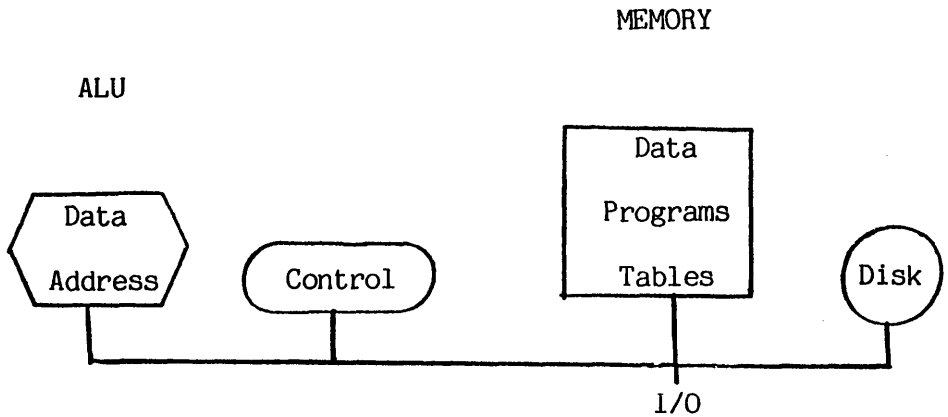
A new force field was also implemented, since existing force fields did not contain the relevant parameters for the molecular systems which were being modelled. The majority of the force field constants were supplied by Dr. M. S. Tute although some had to be derived from experimental data.

Section 7.3.

The Array Processor

As discussed in Chapters One - Three molecular mechanics has proved to be an extremely powerful tool for modeling and understanding molecular structures. However it has one main drawback in that it is highly CPU intensive. Even dedicated mini and supermini computers (eg PDP 11's, VAX 11/780's) can take days to compute the complicated functions involved. However with an array processor the time involved can be shortened quite dramatically since it performs far faster than any supermini computer of similar price. Several studies have shown that array processors have one of the best price/performance ratios to be found.[1,2,3] In fact for the molecular mechanics calculations described in this thesis the array processor performed 35 - 40 times faster than the PDP 11/40 and this increase in speed could have improved even further by programming the machine in assembly language.

The reason why array processors are so much faster than other computers is due to their architecture. The typical elements of a minicomputer are shown in Figure 7.1. The single bus provides a pathway for access to all parts of the system. However the speed



Single bus provides a pathway for access to all parts.

Figure 7.1

of a calculation is restricted, since each device must perform multiple functions and the single bus limits access to all the devices such that only one device can be accessed in one machine cycle. There are three ways of increasing the speed of this kind of system a)by increasing the speed of the logical elements, b)calculations can be spread out amongst parallel elements which have specific functions and c)calculation steps can be spread along a pipeline. The FPS-100 array processor successfully combines all three approaches. It consists of several independent memories and arithmetic units with specialized functions - all of which are interconnected by a network of buses. Figure 7.2. gives an general overview of this type of architecture. This arrangement allows the array processor to be programmed with parallel code which allows all the units to be addressed, or initiated within a single line of code. For example, one line of code may contain instructions to deposit a number from a bus into one of the datapad registers, initiate an addition in the floating adder and initiate a multiplication in the floating point multiplier. This parallel assembly language is highly unusual and difficult to learn and is further complicated by many features of the machine's architecture eg. the result of a multiplication requires to be pushed through the floating multiplier and will not appear on the FM bus until three cycles after the multiplication was initiated. Thus, although array processors are very efficient in terms of computing power versus price the cost of software development is significantly higher than that of standard computers.

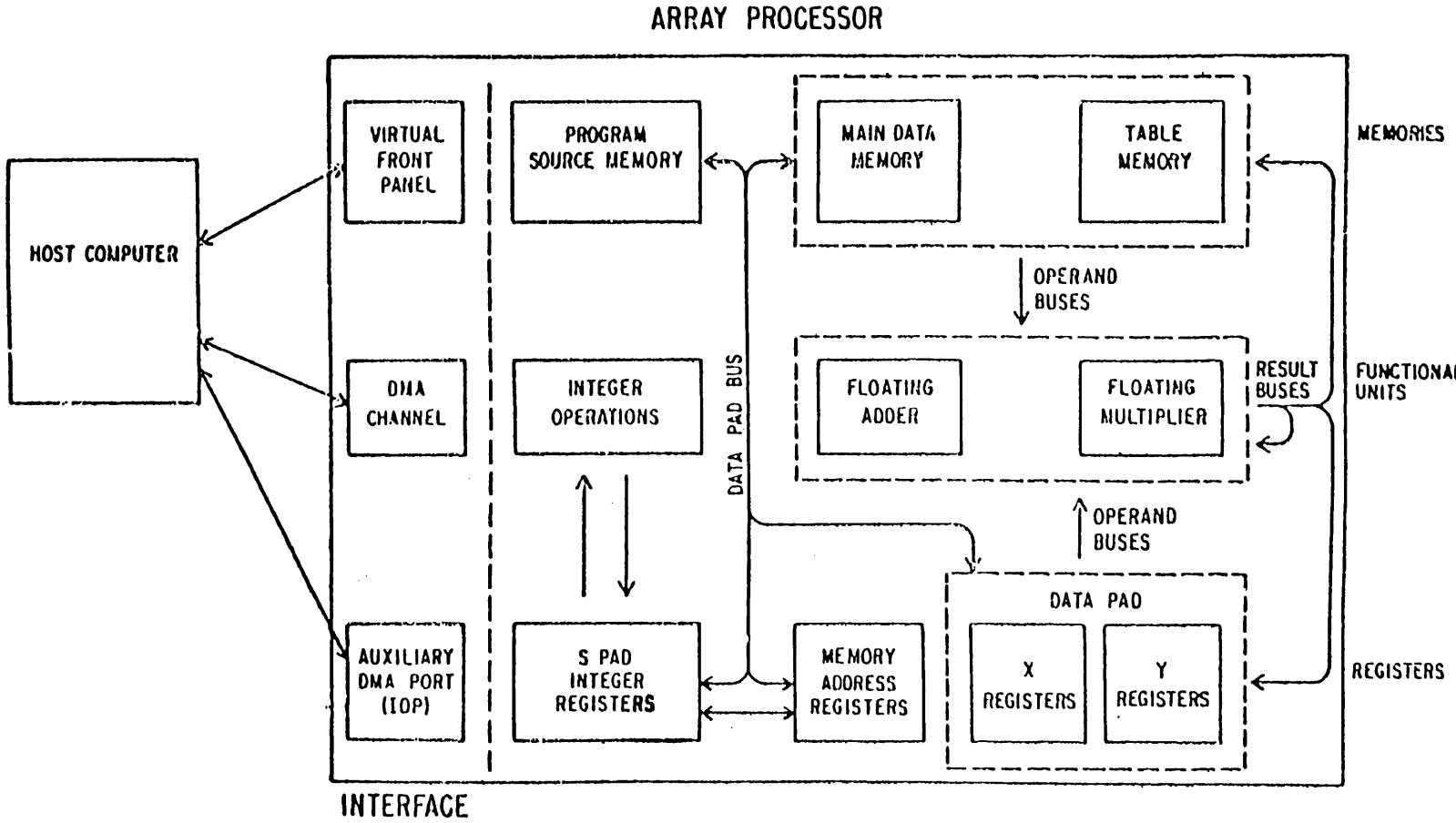

ARRAY PROCESSOR
HOST COMPUTER
VIRTUAL FRONT PANEL
DMA CHANNEL
AUXILIARY DMA PORT (IOP)
PROGRAM SOURCE MEMORY
INTEGER OPERATIONS
S PAD INTEGER REGISTERS
DATA PAD BUS
MAIN DATA MEMORY
TABLE MEMORY
MEMORIES
OPERAND BUSES
FLOATING ADDER
FLOATING MULTIPLIER
RESULT BUSES
FUNCTIONAL UNITS
OPERAND BUSES
MEMORY ADDRESS REGISTERS
DATA PAD
X REGISTERS
Y REGISTERS
REGISTERS
INTERFACE

Figure 7.2

There is a way of overcoming this problem. A company called System Software Factors have produced a software package called the TOAST FORTRAN Development System. This software enables the user to write/convert standard FORTRAN programs to run in the FPS-100 array processor without having to learn how to program in complicated parallel assembly language.

The TOAST FORTRAN system is composed of several components. Of particular interest are the language, the compiler, the linker and the host/AP interface routines.

Creating executable array processor programs with the TOAST system requires two separate source programs - a host program which will run in the host machine (eg a PDP 11/40) and an array processor program which will run in the array processor.

The FPS-100 array processor is incapable of any I/O, therefore the host program is required to fetch and initialise the array processor, find and load the array processor program, transfer any external variables to the array processor, run the array processor program and recover the results from specific array processor addresses. All these operations are carried out using a set of host/array processor interface routines which are stored in a library as a set of FORTRAN callable modules.

The array processor program is written in TOAST FORTRAN. This is a subset of the FORTRAN IV language which contains no I/O statements such as READ, WRITE, OPEN, CLOSE etc. This source file is compiled and linked with the TOAST compiler and TOAST linker respectively to produce the binary file which will run in the array processor. The compiler has been designed to optimise the

linear TOAST FORTRAN source code producing parallel code which makes much more efficient use of the machine's parallel architecture.

In many ways this system can be thought of analogous to subroutine usage - the main (or host) program calls the subroutine (or array processor program), variables are transferred across, the subroutine runs, results are transferred back and the main program continues. This makes writing (and converting existing programs) to run in the array processor much simpler since the user is not required to know anything about parallel programming.

During the course of this project several programs were converted to run in the array processor. These include a rotation program, a least squares fitting routine, energy minimisation programs and a conformer generation program - the GLOMIN program which is given as an example of this type of programming in Appendix A.

Section 7.4.

Graphics Programming

The main graphics terminal on the system was a Megatek 7000S calligraphic terminal. Calligraphic terminals use the same technology as oscilloscopes, in that they draw vectors between specified points. For the purpose of programming the Megatek screen had 4096 x 4096 addressable points on its screen. Programming involved a library of FORTRAN callable routines - the MGS software and the input for these consisted of addresses, character strings and parameters describing character size,

intensity etc. One important feature of the Megatek 7000S is that it has eight planes of memory - each of which can contain a unique picture (display list) and each picture can be switched on or off individually. This feature proved very useful for maintaining menus whilst displaying structures and writing information to the screen and was also used for double buffering to give the effect of dynamic rotation.

Although the MOL system had already been set up before this project began several cosmetic alterations had to be made to the system in particular a windowing/clipping function was written to prevent wrapping of vectors which had exceeded the dimension of the screen. The rotation speed was also increased by packing two sixteen bit integers into a thirtytwo bit integer and sending that over to the Megatek with one call rather than two. Some statistical routines were also written to draw and give hardcopy of graphs and an example of this is given in Appendix B, Figures 1 and 2.

Chapter Eight

Section 8.1

Introduction

This chapter contains a discussion of work carried out in conjunction with Professor A.Y.Meyer while he was on sabbatical leave at Glasgow University.

The work was an attempt to suggest a reasonable way of modifying existing force fields so that their non-bonding parameters will be more transferable. This work shows that the non-bonding potential functions and parameters currently used by different force fields give hugely different estimates as to the size and hardness of the hydrogen atom. However, if an electrostatic term is introduced in the calculation of the non-bonded interactions, the different force fields give more consistant values for the non-bonded parameters. It is therefore proposed that the inclusion of this term would both improve the performance of the individual force field and make the non-bonded parameters more transferable.

This work was published as a paper in the Journal of Computational Chemistry and the paper is given in Appendix B.

Section 8.2

The Force Field Method

The force field method has been described in great detail in Chapters One to Three of this thesis. It originated in vibrational spectroscopy where it was developed in an attempt to predict

vibrational frequencies from a simple mathematical model and was later developed for the calculation of other molecular properties.

In the force field method the molecule is regarded as a group of atoms held together by simple harmonic forces which can be represented by potential functions derived from classical mechanics. The set of functions chosen to reproduce particular molecular properties is called the force field.[1]

Setting up a force field involves selection of potential functions, parameterisation for the functions using reliable data and refinement of the force field functions and parameters such that they compare with observable properties for a group of test compounds.[2]

Section 8.3

The Divergence of Non-Bonded Functions and Parameters

Several force fields have been set up by individual groups which use different sets of potential functions. In general the nature of the individual functions is regarded to be unimportant as long as the set of functions chosen gives good agreement with experimental data. However, certain functions can be used to calculate physical properties of the molecule. Comparison of such properties as calculated using different force fields show wide divergence in values. This is particularly noticable in the area of non-bonded interactions[3,4] where several different functions are in use.

As discussed in Chapter Two non-bonded interactions are perhaps one of the most important forces in stabilising molecular

conformation. The non-bonded interaction arises from two independent forces which are related to interatomic separation.. There is a strong repulsive force at short distances and a weak attractive force over larger interatomic separation. The non-bonded function is a combination of these two forces and the general shape of this function is shown in Figure 8.1. E_m is the minimum energy and is a measure of the depth of the potential well (the hardness of the atom) and R_m is the minimum energy distance i.e. the size of the atom.

Several functions have been formulated to describe non-bonded interactions and the following three are perhaps the most frequently used.

1) The Buckingham Potential[5]

$$E_{nb} = \frac{-P}{r^6} + Ae^{(-Br)} \qquad <1>$$

2) The Lennard Jones Potential[6]

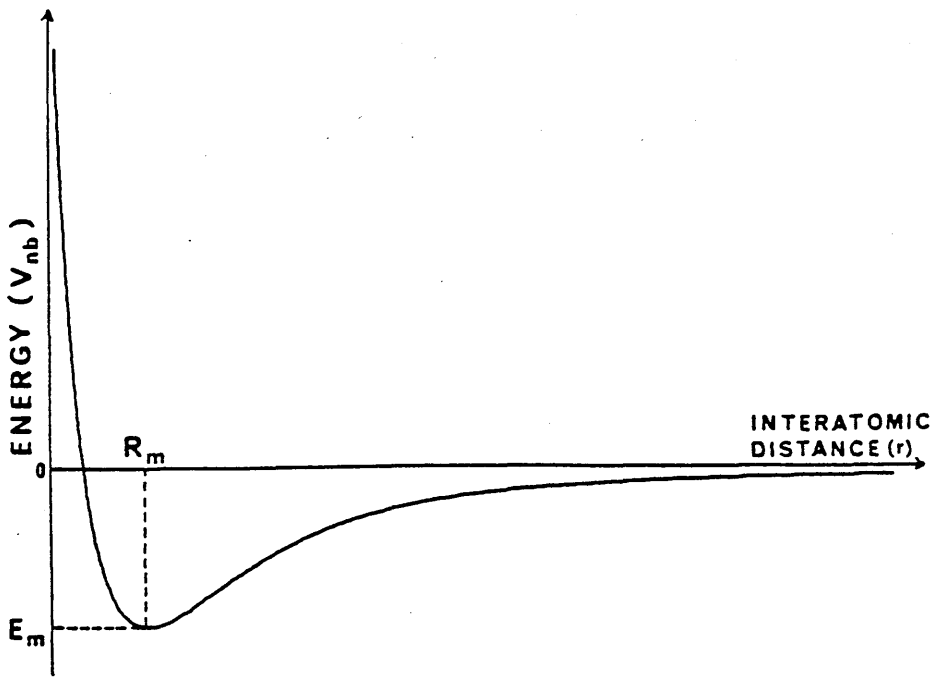
$$E_{nb} = \frac{A}{r^n} - \frac{P}{r^6} \qquad n=9,12 \qquad <2>$$

3) A Modified Version of the Buckingham Potential

$$E_{nb} = E_m ((-2/R_m)^6/r^6) + e^{(12(1-(r/R_m)))} \qquad <3>$$

where r is interatomic separation, B is constant for two interacting atoms and A and P are related to the minimum energy point, E_m, of the function and its associated distance R_m.

All of these different equations can be used to calculate E_m and R_m. This is further complicated by the fact that each force field will be parameterised using experimental data from sources



Van der Waals Interaction Energy

R_m is the minimum interatomic distance at which the attractive and repulsive forces are balanced and E_m is the energy associated with this distance.

Figure 8.1

which are of particular interest to each individual group. It is therefore easy to see why calculated values of physical properties such as E_m and R_m vary so much between force fields.

To illustrate this point the non-bonding parameters for the H···H interaction were calculated using four different force fields. The first force field, FF1, is from the MM2 program by Allinger[7] and is one of the most popular force fields in use today. The second, FF2, was developed at Glasgow University by Bovill and White[8] and was parameterised paying particular attention to compounds with very short H···H non-bonded interactions. The third force field, FF3, is by Jaime and Osawa[9] and modifies the MM2 package so that it can be applied to compounds with short H···H non-bonded interactions and the fourth force field FF4 is by Ermer and Lifson.[10]

FF1 and FF3 employ the Buckingham potential (eqn 1), FF2 uses a modified version of the Buckingham potential (eqn 3) and FF4 uses a Lennard-Jones type potential (eqn 2).

Table 8.1 contains the values of E_m and R_m for a hydrogen atom as calculated by each force field. When compared with FF1 the hydrogen is large and soft in FF2, somewhat small in FF3 and extremely large and hard in FF4.

Therefore it can be seen that the use of different non-bonding functions and the way in which a force field is parameterised can have a major effect upon the values of the non-bonding parameters produced.

Table 8.1

	R_m (A)	E_m (kcal mol^{-1})	Ratio R_m	Ratio E_m
FF1	1.50	0.047	1	1
FF2	1.55	0.016	1.03	0.34
FF3	1.426	0.047	0.95	1
FF4	1.816	0.064	1.21	1.36

Non bonding parameters for H···H interaction as calculated by four different force fields.

Table 8.2

r (A)	E_q (kcal mol^{-1})
1.5	0.354
2.0	0.265
2.5	0.213
3.0	0.177
4.0	0.144
7.0	0.076

Coulmbic interaction between two hydrogen with partial charges of 0.04 charge units.

Thus when parameterising a force field, concentrating on particular structures such as short H···H interatomic distances introduces a bias into the system.

It is therefore obvious that the functions currently used to calculate non-bonded interaction are not adequate. There is, however, another force involved in the attraction/repulsion of non-bonded atoms - namely the electrostatic interaction. None of the above fields contains a function with which this force can be implicitly calculated.

Section 8.4

Introduction of an Electrostatic Term

The electrostatic force between two non-bonded atoms is a measure of the interactions of the partial charges on the atoms and the energy involved in this interaction is given by the equation:-

$$E_q = 332.17\ q_i q_j\ /\ Dr$$

where q_i and q_j are the partial charges for atoms i and j in charge units, r is the interatomic distance in Angstroms, 332.17 is a conversion factor to return a value in kcal mol^{-1} and D is the Dielectric constant which has a value of one in vacuo.

Still considering the H···H interaction, Table 8.2 contains values for the energy associated with two non-bonded hydrogen atoms having partial charges of 0.04 units at different interatomic separations (0.04 is felt to be a reasonable value for the partial charge on a hydrogen in a saturated hydrocarbon.). As

expected the electrostatic force is far more repulsive at short interatomic distances than at larger separations.

Combining the electrostatic interaction E_q with the non-bonded interaction E_{nb} to give a new non-bonded function E_{nb}' such that

$$E_{nb}' = E_{nb} - E_q$$

since both E_{nb} and E_q are known E_{nb}' can be plotted point by point and Figure 8.2 shows the shape of this function. With this new non-bonding pontential E_{nb}' there is a greater repulsion at shorter distances due to the contribution made by E_q. The new values of E_m' and R_m' can be calculated from the new function E_{nb}'

The effect of introducing an electrostatic term to the four force fields previously used can be examined by recalculating the values of the non-bonding parameters. The new values are given in Table 8.3. This shows that when charge interaction is made explicit, all four fields give a more or less identical value for the hardness and size of the atom in FF1, FF2 and FF3 and although the hydrogen atom is still fairly large in FF4 the values are converging.

Since $E_q + E_{nb}'$ is equal to E_{nb} the individual fields should still perform as well as before however it would appear that the new non-bonding parameters E_m' and R_m' as calculated by the different force fields have been made more transferable.

The major problem with implementing this change is that there are many different ways of estimating partial charge. The most basic approach would be to assign a partial charge to each atom by

Table 8.3

	R_m (A)	E_m (kcal mol^{-1})	Ratio R_m	Ratio E_m
FF1'	1.44	0.202	1	1
FF2'	1.40	0.198	0.97	0.98
FF3'	1.37	0.210	0.95	1.04
FF4'	1.76	0.213	1.22	1.05

Non-bonding parameters for H···H interaction on inclusion of electrostatic term.

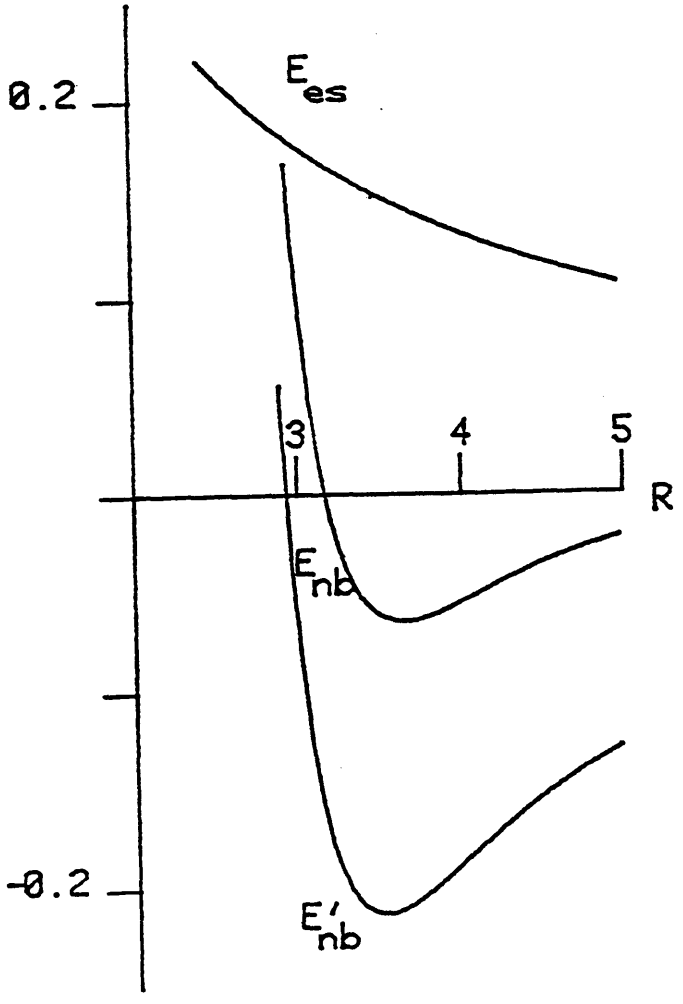


Figure 8.2

type - although this may prove to be somewhat approximate since it is known that the partial charge on an atom varies with its molecular environment.

A better approach might involve calculation of the partial charges for each molecular conformation using say a CNDO type method, but again this would lead to problems as it would impose a limit on the size of the molecules which could be studied.

It is however clear that a force field F can be converted into F' relatively easily without affecting its performance but giving improved calculation of non-bonding parameters.

To summarise, this work shows that the inclusion of electrostatic interactions in force fields is of great importance especially since force fields are now being used for calculations involving heteromolecules, peptides and even proteins in which the partial charge interactions play a fundamental part in stabilising molecular conformations.

REFERENCES

Chapter One

1. Lesk, A.M. and Hardman, K.D., "Supercomputers in Chemistry", American Chemical Society, Symposium 173, p143 (1981).
2. Chabot, A.A., Geddes, A.J., North, A.C.T. and Potterton, E.A., XII International Congress of Crystallography, Abs 18 (1981).
3. Bond, P.J., Comput. Graphics, ACM SIGGRAPH, 6, 13 (1972).
4. Katz, L. and Levinthal, C., Ann. Rev. Biophys. & Bioeng., 1, 465 (1972) .
5. Langridge, R., Federation Proc., 33, 2332 (1974).
6. Fox, J.L., Chem. & Eng. News, 21 July, p27 (1980).
7. White, D.N.J. and Bovill, M.J., J.C.S. Perkin II, 1610 (1977).
8. Niketic, S.R. and Rasmussen, K., "The Consistent Force Field", Springer-Verlag (1977).
9. Engler, E.M., Andose, J.D. and Schleyer, P von R., J. Am. Chem. Soc., 95, 8005 (1973).
10. Ermer, O. and Lifson, S., J. Am. Chem. Soc., 95, 4121 (1973).
11. Allinger, N.L., Tribble, M.T. and Miller, M.A., Tetrahedron, 28, 1173 (1972).
12. Jorgensen, W.L. and Swenson, C.J., J. Am. Chem. Soc., 107, 569 (1985).
13. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Algona, G., Profeta, S. Jr. and Weiner, P., J. Am. Chem. Soc., 106, 765 (1984).
14. Hagler, A.T., Huler, E. and Lifson, S., J. Am. Chem. Soc., 96, 5319 (1974).
15. Altona, C. and Farber, D.H., Top. Curr. Chem., 45, 1 (1974).

16. Allinger, N.L., Adv. Phys. Org. Chem., 13, 7 (1976).

17. Hursthouse, M.B., Moss, G.P. and Sales, K.D., Ann. Rep. Prog. Chem., Sect B, 75, 23 (1978).

18. Momany, F.A., Top. Curr. Phys., 26, 47 (1981).

19. Osawa, E. and Musso, H., Top. Stereochem., 13, 117 (1982).

20. Boyd, D.B. and Lipkowitz, K.B., J. Chem. Ed., 59, 269 (1982).

21. Burkert, U. and Allinger, N.L., "Molecular Mechanics", ACS Monograph 177, American Chemical Society (1982).

22. Barry, C.D., IV International Biophysics Congress Procedings, Acad. Sci. USSR., p366 (1973).

23. Olsen, E.C. and Christofferson, R.E., "Computer-Assisted Drug Design", American Chemical Society, Symposium Series 112 (1979).

24. Gund, P., Andose, J.D., Rhodes, J.B. and Smith, G.M., Science, 208, 1425 (1980).

25. Humblet, C. and Marshall, G.R., Drug Dev. Res., 1, 409 (1981).

26. Bartusiak, M., Discover, August 1981, p47.

27. Blaney, J.M., Jorgensen, E.C., Connolly, M.L., Ferrin, T.E., Langridge, R., Oatley, S.J., Burridge, J.M. and Blake, C.C.F., J. Med. Chem., 25, 785 (1982).

28. Kerr, M.A. and Kenny, A.J., Biochem. J., 137, 477 (1974).

29. Matthews, B.W., Weaver, L.H. and Kester, W.R., J. Biol. Chem., 249, 8030 (1974).

30. Orlowski, M., Michaud, C. and Chu, T.G., Eur. J. Biochem., 135, 81 (1983).

31. Almenoff, J. and Orlowski, M., Biochemistry, 22, 590 (1983).

32. Hudgin, R.L., Charleson, S.E., Zimmerman, M., Mumford, R. and Wood, P.L., Life Sci., 29, 2593 (1981).

33. Roques, B.P., Fournie-Zaluski, M-C., Saroca, E., Lacomte, J.M., Malfroy, B., Llorens, C.,and Schwartz, J.C., Nature, 288, 286 (1980).

34. Pozsgay, M., Michaud, C. and Orlowski, M., Biochem. Soc. Trans., 13, 44 (1985).

35. Kester, W.R., and Matthews, B.W., Biochemistry, 16, 2506 (1977).

36. Weaver, L.H., Kester, W.R. and Matthews, B.W., J. Mol. Biol., 114, 119 (1977).

37. Bolognesi, M.C. and Matthews, B.W., J. Biol. Chem., 254, 634 (1979).

38. Holmes, M.A. and Matthews, B.W., Biochemistry, 20, 6912 (1981).

39. Monzingo, A.F. and Matthews, B.W., Biochemistry, 21, 3390 (1982).

40. Hangauer, D.G., Monzingo, A.F. and Matthews, B.W., Biochemistry, 23, 5730 (1984).

Chapter Two

1. Wertz, D.H. and Allinger, N.L., Tetrahedron, 30, 1579 (1974).

2. Klyne, W. and Prelog, V., Experentia, 16, 521 (1960).

3. Bartell, L.S., J. Am. Chem. Soc., 99, 3279 (1977).

4. Allinger, N.L., Hindman, O. and Honig, H., J. Am. Chem. Soc., 99, 3282 (1977).

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

6. Ermer, O., Tetrahedron, 31, 1849 (1975).

7. Margenau, H. and Kestner, N.R., "Theory of Intermolecular Forces", Pergammon, Oxford (1969).

8. Williams, J.E., Stang, P.J. and Shleyer, P.von R., Ann. Rev. Phys. Chem., 19, 531 (1968).

9. Allinger, N.L. and Sprague, J.T., Tetrahedron, 31, 21 (1975).

10. White, D.N.J., Bovill, M.J., J.C.S. Perkin II, 1610 (1977).

11. Jacob, E.J., Thompson, H.B. and Bartell, L.S., J. Chem. Phys., 47, 531 (1967).

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

13. Niketic, S.R. and Rasmussen, K., "The Consistent Force Field", Springer-Verlag, New York (1977).

Chapter Three

1. Walsh, G.R., "Methods of Optimisation", John Wiley and Sons, p75 (1979).

2. Ramachandran, G.N., Venkatachalam, C.M. and Krumm, S., Biophys. J., 6, 849 (1966).

3. Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V., J. Mol. Biol., 7, 95 (1966).

4. Hooke, R. and Jeeves, T.A., J. Assn. Comp. Mach., 8, 212 (1961).

5. Rosenbrock, H.H., The Computer Journal, 3, 175 (1960).

6. Nelder, J.A. and Mead, R., The Computer Journal, 7, 308 (1965).

7. Spendley, W., Hext, G.R. and Himsworth, F.R., Technometrics, 4, 441 (1962).

8. Smith, C.S., N.C.B. Scientific Dept., Report N° SC846/MR/40 (1962).

9. Powell, M.J.D., The Computer Journal, 7, 155 (1964).

10. Walsh, G.R., "Methods of Optimisation", John Wiley and Sons, p106, (1979).

11. Engler, E.M., Anglose, J.D. and Schleyer, P.von R., J. Am. Chem. Soc., 95, 8005 (1973).

12. Davidon, W.C., A.E.C., R & D Report ANL-5990 (Rev.),Argonna National Laboratory (1959).

13. Fletcher, R. and Powell, M.J.D., The Computer Journal, 6, 163 (1963).

14. Walsh, G.R., "Methods of Optimisation", John Wiley and Sons, p108 (1979).

15. Burkert, U. and Allinger, N.L., "Molecular Mechanics", ACS Monograph 177, American Chemical Society, p67 (1982).

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

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

18. Faber, D.H. and Altona, C., Comput. Chem., 1, 203 (1977).

19. Thomas, M.W. and Emerson, D.J., Mol. Struct., 16, 473 (1973).

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

21. White, D.N.J., Acta Cryst., A33, 1010 (1977).

22. White, D.N.J., Computers and Chemistry, 1, 225 (1977).

Chapter Four

1. Sutherland, I.E., "SKETCHPAD: A Man-Machine Graphical Communication System", Spartan Books, Baltimore, p329 (1963).

2. Ramachandran, G.N., Venkatachalam, C.M. and Krumm, S., Biophys. J., 6, 849 (1966).

3. Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V., J. Mol. Biol., 7, 95 (1966).

4. Johnson, C.K., ORNL Tech. Report 3794 (1965).

5. Wipke, W.T. and Gund, P., J. Am. Chem. Soc., 96, 299 (1974).

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

Chapter Five

1. Hammes, G.G., Nature, 204, 342 (1964).
2. Jenks, W.P., "Catalysis in Chemistry and Enzymology", Chapt. 5, McGraw-Hill, New York (1969).
3. Koshland, D.E., J. Theor. Biol., 2, 75 (1962).
4. Engel, P.C., "Enzyme Kinetics The Steady-State Approach", p45, Chapman and Hall (1981).
5. Kretsinger, R.H., Ann. Rev. Biochem., 45, 239 (1976).
6. Nozaki, M., J. Syn. Org. Chem., 34, 805 (1976).
7. Lehninger, A.L., "Biochemistry", 2nd Ed., Worth, New York (1975).
8. Neurath, H. and Walsh, K.A., Proc. Nat. Acad. Sci. USA, 73, 3825 (1976).
9. de Haan, C., Neurath, H. and Teller, D.C., J. Mol. Biol., 92, 225 (1975).
10. Rubin, C.S., and Rosen, O.M., Ann. Rev. Biochem., 44, 831 (1975).
11. Sutherland, E.W., Science, 177, 401 (1972).
12. Umberger, H.E., Science, 123, 848 (1956).
13. Yates, R.A. and Pardee, A.B., J. Biol. Chem., 221, 757 (1956).
14. Umberger, H.E. and Brawm, B., J. Biol. Chem., 233, 415 (1958).
15. Helmreich, E. and Cori, C.F., Proc. Nat. Acad. Sci., 51, 131 (1964).
16. Lienhard, G.E., Science, 180, 149 (1973).
17. Metzler, D.E., "Biochemistry", Academic Press, New York (1977).
18. Page, M.I. and Jenks, W.P., Proc. Nat. Acad. Sci., 68, 1678 (1971).
19. Bruice, T.C.B, The Enzymes, 2, 217 (1970).

20. Storm, D.R. and Koshland, D.E. Jr., Proc. Nat. Acad. Sci. USA, 66, 445 (1970).

21. Storm, D.R. and Koshland, D.E. Jr., J. Am. Chem. Soc. USA, 94, 5815 (1972).

22. Jenks, W.P., Adv. Enzymol., 43, 220 (1975).

23. Brown, A.J., Trans. Chem. Soc. , 81, 373 (1902).

24. Michaelis, L. and Menten, M.L., Biochem. Z., 49, 333 (1913).

25. Lineweaver, H. and Burk, D., J. Am. Chem. Soc., 56, 658 (1934).

26. Krebs, H.A. and Eggleston, L.V., Biochem. J., 34, 442 (1940).

27. Harris, J.I., Meriwether, B.P. and Park, J.H., Nature, 198, 154 (1963).

28. Dixon, M., Biochem. J., 55, 170 (1953).

29. Cheng, Y.C. and Prusoff, W.H., Biochem. Pharamacol., 22, 3099 (1973).

30. Needleman, S.B. and Wunsch, C.D., J. Mol. Biol., 48, 443 (1970).

31. Gilbert, W., Nature, 271, 501 (1978).

32. Britten, R.J. and Davidson, E.H., Fed. Proc., 35, 2151 (1974).

33. Dickerson, R.E., "Molecular Evolution and Polymorphism", (Ed.) Kimura, M., National Institute of Genetics, Mishima, Japan (1977).

34. Hartley, B.S., Altosaar, I., Dothie, J.M. and Neuberger, M.S., "Structure-Function Relationships in Proteins", (R. Markham and R.W. Horne, Eds.) New York (1976).

35. Blow, D.M., Acc. Chem. Res., 9, 145 (1976).

36. Rossman, M.G. and Argos, P., Mol. Cel. Biochem., 21, 161 (1978).

Chapter Six

1. Martin, W.R., Pharmacol. Rev., 19, 463 (1967).

2. Atweh, S.F. and Kuher, M.J., Br. Med. Bull., 39, 47 (1983).

3. Paterson, S.J., Robson, L.E. and Kosterlitz, H.W., Br. Med. Bull., 39, 31 (1983).

4. West, R.E. and Miller, R.J., Br. Med. Bull., 39, 53 (1983).

5. Henderson, G., Br. Med. Bull., 39, 59 (1983).

6. Morley, J.S., Br. Med. Bull., 39, 5 (1983).

7. Hughes, J., Smith, T.W., Kosterlitz, H.W., Fothergill, L.A., Morgan, B.A. and Morris, H.R., Nature, 258, 577 (1975).

8. Hughes, J., Brain Res., 88, 295 (1975)

9. Meek, J.L., Yang, H-T.T. and Costa, E., Neuropharmac., 16, 151 (1977).

10. Dupont, A., Cusan, L., Caron, M., Alvarado-Urbina, G. and Labrie, F., Life Sci., 21, 907 (1977).

11. Miller, R.J., Chany, K-J., Cuatrecasas, P. and Wilkinson, S., Biochem. Biophys. Res. Comm., 74, 1311 (1977).

12. Craves, F.B., Law, P.Y., Hunt, C.A. and Loh, H.N., J. Pharmac. Exp. Ther., 206, 492 (1978).

13. Hughes, J., Br. Med. Bull., 39, 17 (1983).

14. Swerts, J.P., Perdrisot, R., Malfroy, B. and Schwartz, J.C., Eur. J. Pharmac., 53, 209 (1979).

15. Malfroy, B., Swerts, J.P., Llorens, C. and Schwartz, J.C., Neurosci., Lett., 11, 329 (1979).

16. Benuck, M. and Marks, N., Biochem. Biophys. Res. Comm., 95, 822 (1980).

17. Hudgin, R.L., Charleson, S.E., Zimmerman, M., Mumford, R. and Wood, P.L., Life Sci., 29, 2593 (1981).

18. Checler, F., Vincent, J.P., and Kitabgi, P., J. Neurochem., 41, 375 (1983).

19. Fournie-Zaluski, M-C., Lucas, E., Waksman, G. and Roques, B.P., Eur. J. Biochem., 139, 267 (1984).

20. Hollenberg, N.K., Am. J. Med., 77, 60 (1984).

21. Mastas, R., Kenny, A.J. and Turner, A.J., Biochem. J., 223, 433 (1984).

22. Sullivan, S., Akil, H. and Barchas, J.D., Commun. in Physchopharm., 2, 525 (1978).

23. Roques, B.P., Fournie-Zaluski, M-C., Saroca, E., Lacompte, J.M., Malfroy, B., Llorens, C.,and Schwartz, J.C., Nature, 288, 286 (1980).

24. Murthy, L.R., Glick, S.D., Almenoff, J., Wilk, S. and Orlowski, M., Eur. J. Pharmac., 102, 305 (1984).

25. Scott, W.L., Mendelsohn, L.G., Cohen, M.L., Evans, D.A. and Frederickson, R.C.A., Life Sci., 36, 1307 (1985).

26. Kerr, M.A. and Kenny, A.J., Biochem. J., 137, 477 (1974).

27. Gafford, J.T., Skidgel, R.A., Erdos, E.G. and Hersh, L.B., Biochemistry, 22, 3265 (1983).

28. Almenoff, J. and Orlowski, M., Biochemistry, 22, 590 (1983).

29. Orlowski, M., Michaud, C. and Chu, T.G., Eur. J. Biochem., 135, 81 (1983).

30. Orlowski, M. and Wilk, S., Biochemistry, 20, 4942 (1981).

31. Almenoff, J., Tierstein, A.S., Thornton, J.C. and Orlowski, M., J. Lab. Clin. Med., 103, 420 (1984).

32. Almenoff, J., Wilk, S. and Orlowski, M., Biochem. Biophys. Res. Comm., 102, 206 (1981).

33. Fulcher, I.S., Mastas, R., Turner, A.J. and Kenny, A.J., Biochem. J., 203, 519 (1982).

34. Diss. Abst. Int. B., 44 (10) 3067 (1984).

35. Pozsgay, M., Michaud, C. and Orlowski, M., Biochem. Soc. Trans., 13, 44 (1985).

36. Almenoff, J. and Orlowski, M., J. Neurochem., 42, 151 (1984).

37. Matthews, B.W., Weaver, L.H. and Kester, W.R., J. Biol. Chem., 249, 8030 (1974).

38. Rossman, M.G. and Argos, P., Mol. Cel. Biochem., 21, 161 (1978).

39. Andrews, P.R., Craik, D.J. and Martin, J.L., J. Med. Chem., 27, 1648 (1984).

40. Kester, W.R. and Matthews, B.W., Biochemistry, 16, 2506 (1977).

41. Weaver, L.H., Kester, W.R. and Matthews, B.W., J. Mol. Biol., 114, 119 (1977).

42. Bolognesi, M.C. and Matthews, B.W., J. Biol. Chem., 254, 634 (1979).

43. Monzingo, A.F. and Matthews, B.W., Biochemistry, 21, 3390 (1982).

44. Holmes, M.A. and Matthews, B.W., Biochemistry, 20, 6912 (1981).

45. Monzingo, A.F. and Matthews, B.W., Biochemistry, 23, 5724 (1984).

46. Hangauer, D.G., Monzingo, A.F. and Matthews, B.W., Biochemistry, 23, 5730 (1984).

47. Hangauer, D.G., Gund,P., Andose, J.D., Bush, B.L., Fluder, E.M., McIntyre, E.F. and Smith G.M., Ann. N. Y. Acad. Sci., 439, 124 (1985).

48. Benuck, M., Berg, M.J. and Marks, N., Biochem. Biophys. Res. Comm., 107, 1123 (1982).

49. Malfray, B. and Schwartz, J.C., Biochem. Biophys. Res. Comm., 106, 276 (1982).

50. Mumford, R.A., Zimmerman, M., ten Broeke, J., Taub, D., Joshua, H., Rothrock, J.W., Hirshfield, J.M., Springer, J.P. and Patchett, A.A., Biochem. Biophys. Res. Comm., 109, 1303 (1982).

Chapter Seven

1. Bucy, R.S. and Senne, K.D., Comp. Math. Applications, 6, 317 (1980).

2. Karplus, W.J. and Cohen, D., Computer, 11, (Sept 81).

3. Berens, P.H. and Wilson, K.R., J. Comp. Chem., 4, 313 (1983).

Chapter Eight

1. Burkert, U. and Allinger, N.L., "Molecular Mechanics", ACS Monograph 177, American Chemical Society (1982).

2. Niketic, S.R. and Rasmussen, K., "The Consistent Force Field", Springer-Verlag (1977).

3. Williams, J.E., Stang, P.J. and Schleyer, P. von R., Ann. Rev. Phys. Chem., 19, 531 (1968).

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

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

6. Ermer, O., Tetrahedron, 31, 1849 (1975).

7. Allinger, N.L., J. Am. Chem. Soc., 99, 8127 (1977).

8. White, D.N.J. and Bovill, M.J., J. Chem. Soc. Perkin Trans. 2, 1610 (1977).

9. Jaime, C. and Osawa, E., Tetrahedron, 39, 2769 (1983).

10. Ermer, O. and Lifson, S.J., J. Am. Chem. Soc., 94, 4121 (1973).

APPENDICES

APPENDIX A

The GLOMIN Program

```
      PROGRAM GLOMIN
      COMMON/ROTAT/NATOM(10,4),NR(10),NRATOM(10,50),ISTART(16)
      COMMON/CONST/RAD1,RAD2,RADI,NRING,NT,NRI,NTP,NP,NAV
      COMMON/CHECK/IGEN,EXCLUD(10,15),NEXCLD,J1(15),NBONDS,
     1 DMAX,HOMOP
      COMMON/TCHECK/NREST,ITREST(3,4),TVAL(3),TTOL(3)
      COMMON/VARS/NATOMS,NAMES,XI,ICONN,ITYPES,TITLE,NTITLE
      COMMON/TAUS/TANGLE(9,2),TAU(10),NGEN,ICHEND,JCHEND,TDIST,
     1 TOL
      LOGICAL IGEN,HOMOP
      LOGICAL*1 EXCLUD,TITLE(80)
      DIMENSION NAMES(50,3),XI(50,3),ICONN(50,4),ITYPES(50),
     1 J1S(15)
      DIMENSION DT(8),DELTAU(2),XS1(50),XS2(50),XS3(50),NRA(50)
      DATA DT/5.,4.,3.,2.,2.,2.,1.,1./
C
C  INITIALIZE THE ARRAY PROCESSOR
C
      ICHAN=IGETC()
      CALL SSFLDP(ICHAN,'GLOAP',0,0,ISTAT)
C
C*******GET CURRENT VALUE OF JSW
      IVAL=IPEEK("44)
      CALL SETUP
C*******ENABLE INTERCEPTION OF CTRL/C USING SYSLIB SCCA ROUTINE
      IFLAG=0
      CALL SCCA(IFLAG)
C*******OPEN A FILE WHICH WILL CONTAIN THE OUTPUT COORDINATES.
      TYPE 9000
9000  FORMAT('$OUTPUT FILE NAME ')
      CALL ASSIGN(1,' ',-1,'NEW','CC')
      WRITE(1,9001)
9001  FORMAT('+GLOBAL MINIMISATION - GENERATED CONFORMERS.')
C*******RADIANS TO DEGREES (AND BACK) CONVERSION FACTORS.
      PI=3.1415926536
      RAD1=PI/180.
      RAD2=RAD1*RAD1
      RADI=1./RAD1
      ICNF=ISTART(16)
C*******SAVE ORIGINAL COORDINATES FOR REUSE AS STARTING COORDS.
      DO 2 I=1,NATOMS
      WRITE(1,9002) ITYPES(I),(NAMES(I,J),J=1,3)
      XS1(I)=XI(I,1)
      XS2(I)=XI(I,2)
      XS3(I)=XI(I,3)
 2    CONTINUE
9002  FORMAT(X,I5,3A2)
C*******ALL DONE LETS GET STARTED
      J1S(NTP)=0
      DO 10 I1=ISTART(1),NGEN
      J1(1)=I1
      DO 10 I2=ISTART(2),NGEN
      J1(2)=I2
      DO 10 I3=ISTART(3),NGEN
      J1(3)=I3
      DO 10 I4=ISTART(4),NGEN
      J1(4)=I4
```

```
        DO 10 I5=ISTART(5),NGEN
        J1(5)=I5
        DO 10 I6=ISTART(6),NGEN
        J1(6)=I6
        DO 10 I7=ISTART(7),NGEN
        J1(7)=I7
        DO 10 I8=ISTART(8),NGEN
        J1(8)=I8
        DO 10 I9=ISTART(9),NGEN
        J1(9)=I9
        DO 10 I10=ISTART(10),NGEN
        J1(10)=I10
        DO 10 I11=ISTART(11),NGEN
        J1(11)=I11
        DO 10 I12=ISTART(12),NGEN
        J1(12)=I12
        DO 10 I13=ISTART(13),NGEN
        J1(13)=I13
        DO 10 I14=ISTART(14),NGEN
        J1(14)=I14
        DO 10 I15=ISTART(15),NGEN
        J1(15)=I15
        IF(J1(NTP).EQ.J1S(NTP)) GO TO 10
        J1S(NTP)=J1(NTP)
C*******MODIFY JSW TO SET CONSOLE I/O TO SPECIAL MODE
C*******AND TO ALLOW ITTINR TO RETURN A NEGATIVE
C*******ARGUMENT IF NO CHAR AVAILABLE (SEE DEC
C*******ADVANCED PROGRAMMERS GUIDE).
        CALL IPOKE("44,"10100.OR.IVAL)
C*******IS THERE ANYBODY THERE?
        I=ITTINR()
C*******RESET JSW
        CALL IPOKE("44,IVAL)
C*******WAS A CTRL/C TYPED? IF SO GO TO ABORT ROUTINE
        IF(I.EQ.3) CALL CNTRLC(ICNF,NTP,J1)
        IF(NEXCLD.EQ.0) GO TO 7
        CALL IMPOSS
        IF(.NOT.IGEN) GO TO 10
7       IF(.NOT.HOMOP) GO TO 71
        CALL INVERS
        IF(.NOT.IGEN) GO TO 10
C*******GET ORIGINAL COORDINATES BACK.
71      CONTINUE
C
C   BEGIN AP LOOP
C INITIALISE AP PROGRAMME
C
        CALL SSFSGL('')
C
C  TRANSFER DATA
C
        IADR=ISFADR('TRANS')
C
        CALL SSFFTO(TANGLE   ,IADR  ,18  )
        CALL SSFITO(NATOM(1,1) ,IADR+18  ,NT )
        CALL SSFITO(NATOM(1,2) ,IADR+28  ,NT )
```

```
      CALL SSFITO(NATOM(1,3) ,IADR+38   ,NT )
      CALL SSFITO(NATOM(1,4) ,IADR+48   ,NT )
      CALL SSFITO(NT     ,IADR+58 ,1    )
      CALL SSFFTO(XS1    ,IADR+59 ,NATOMS )
      CALL SSFFTO(XS2    ,IADR+109 ,NATOMS )
      CALL SSFFTO(XS3    ,IADR+159 ,NATOMS )
C
      IADR=ISFADR('ROTA')
C
      CALL SSFFTO(RAD1   ,IADR     ,1    )
      CALL SSFITO(NR     ,IADR+1   ,NT   )
C
      IADR=ISFADR('COOR')
C
      CALL SSFITO(NRATOM,IADR+150 ,500)
      CALL SSFITO(NATOMS,IADR+650,1    )
      CALL SSFFTO(RADI   ,IADR+651,1    )
C
      IADR=ISFADR('LSAR')
C
      CALL SSFITO(ICHEND ,IADR     ,2    )
C
      IADR=ISFADR('GLO')
C
      CALL SSFFTO(TDIST  ,IADR      ,1    )
      CALL SSFFTO(TOL    ,IADR+1    ,1    )
      CALL SSFFTO(DT     ,IADR+2    ,8    )
      CALL SSFITO(NTP    ,IADR+10   ,3    )
      CALL SSFITO(J1     ,IADR+13   ,15   )
C
      CALL APWD
C
C RUN GLOAP
C
      CALL SSFRUN
      CALL APWR
C
C GET NEW COORDS BACK FROM AP
C
      IADR=ISFADR('COOR')
C
      CALL SSFFFR(IADR     ,XI    ,150     )
C
      IADR=ISFADR('GLO')
      CALL SSFFFR(IADR+28 ,A    ,1      )
C
      CALL APWD
C
      IF(A.EQ.1.) GOTO 105
      GO TO 10
105   CALL BONDEX
      IF(.NOT.IGEN) GO TO 10
      IF(NREST.LE.0) GO TO 106
      CALL CHECKT
      IF(.NOT.IGEN) GO TO 10
```

```
106       ICNF=ICNF+1
          WRITE(7,9009)ICNF,(J1(IN),IN=1,NTP)
          WRITE(1,9010) (J1(IN),IN=1,NTP)
          DO 11 IA=1,NATOMS
11        WRITE(1,9011) (NAMES(IA,IN),IN=1,3),(XI(IA,IN),IN=1,3)
10        CONTINUE
9009      FORMAT(I10,5X,20I1)
9010      FORMAT(X,20I1)
9011      FORMAT(X,3A2,10X,3F15.5)
          STOP
          END

          SUBROUTINE CHECKT
          COMMON/ROTAT/NATOM(10,4),NR(10),NRATOM(10,50),ISTART(16)
          COMMON/CONST/RAD1,RAD2,RADI,NRING,NT,NRI,NTP,NP,NAV
          COMMON/CHECK/IGEN,EXCLUD(10,15),NEXCLD,J1(15),NBONDS,
         1 DMAX,HOMOP
          COMMON/TCHECK/NREST,ITREST(3,4),TVAL(3),TTOL(3)
          COMMON/VARS/NATOMS,NAMES,XI,ICONN,ITYPES,TITLE,NTITLE
          COMMON/TAUS/TANGLE(9,2),TAU(10),NGEN,ICHEND,JCHEND,TDIST,
         1 TOL
          LOGICAL IGEN,HOMOP
          LOGICAL*1 EXCLUD,TITLE(80)
          DIMENSION NAMES(50,3),XI(50,3),ICONN(50,4),ITYPES(50),
         1 J1S(15)
          IGEN=.FALSE.
          DO 10 I=1,NREST
          CALL TCAL(ITREST(I,1),ITREST(I,2),ITREST(I,3),ITREST(I,4),
         1 TA)
          IF(ABS(ABS(TA)-TVAL(I)).LE.TTOL(I)) GO TO 10
          RETURN
10        CONTINUE
          IGEN=.TRUE.
          RETURN
          END

          SUBROUTINE CNTRLC(ICNF,NTP,J1)
          DIMENSION J1(15)
          TYPE 1
1         FORMAT('1***ABORT REQUEST RECEIVED.')
          TYPE 2,ICNF+1,(J1(I),I=1,NTP)
2         FORMAT(' RESTART INFORMATION:'/
         +' NEXT CONFORMER TO BE GENERATED WOULD HAVE BEEN NO.',I6/
         +' CURRENT GENERATOR SEQUENCE IS ',15I1//)
          CALL EXIT
          END

          SUBROUTINE INVERS
          COMMON/ROTAT/NATOM(10,4),NR(10),NRATOM(10,50)
          COMMON/CONST/RAD1,RAD2,RADI,NRING,NT,NRI,NTP,NP,NAV
          COMMON/CHECK/IGEN,EXCLUD(10,15),NEXCLD,J1(15),NBONDS,
         1 DMAX,HOMOP
          COMMON/VARS/NATOMS,NAMES,XI,ICONN,ITYPES,TITLE,NTITLE
          COMMON/TAUS/TANGLE(9,2),TAU(10),NGEN,ICHEND,JCHEND,TDIST,
         1 TOL
          LOGICAL IGEN
          LOGICAL*1 EXCLUD,TITLE(80)
          DIMENSION NAMES(50,3),XI(50,3),ICONN(50,4),ITYPES(50)
          IREM=MOD(NTP,2)
```

```
      NTP=NTP/2+IREM
      DO 10 I=1,NTP2
      JTEM1=J1(I)
      JTEM2=J1((NTP+1)-I)
      IF(JTEM1.GT.JTEM2) GO TO 11
10    CONTINUE
      IGEN=.TRUE.
      RETURN
11    IGEN=.FALSE.
      RETURN
      END

      SUBROUTINE IMPOSS
C THIS SUBROUTINE REMOVES CODE NUMBERS WHICH CONTAIN PARTIAL
C GENERATOR SEQUENCES REPRESENTING TORSION ANGLE SEQUENCES
C WHICH CANNOT POSSIBLY BE INCORPORATED IN THE CLOSED RING.
C THE PARTIAL SEQUENCES MAY CONTAIN UP TO FOURTEEN USER
C SPECIFIED DIGITS.
      LOGICAL IGEN
      LOGICAL*1 IBUF(16),IEXCLD(15),EXCLUD
      COMMON/CHECK/IGEN,EXCLUD(10,15),NEXCLD,J1(15),NBONDS,
     1  DMAX
      COMMON/ROTAT/NATOM(10,4),NR(10),NRATOM(10,50)
      COMMON/CONST/RAD1,RAD2,RADI,NRING,NT,NRI,NTP,NP,NAV
C ENCODE THE INTEGERS COMPRISING THE CODE NUMBER INTO A
C CHARACTER STRING.
      ENCODE(NTP,1,IBUF) (J1(I),I=1,NTP)
1     FORMAT(15I1)
C TERMINATE CHARACTER STRING WITH A ZERO BYTE
      NTP1=NTP+1
      IBUF(NTP1)=0
C REMOVE ONE PARTIAL SEQUENCE AT A TIME FROM THE EXCLUSION
C ARRAY AND USE THE LIBRARY ROUTINE INDEX TO CHECK IF THE
C PARTIAL SEQUENCE APPEARS IN THE CODE NUMBER.
      DO 10 I=1,NEXCLD
      DO 11 J=1,15
11    IEXCLD(J)=EXCLUD(I,J)
      CALL INDEX(IBUF,IEXCLD,,M)
      IF(M.NE.0) GO TO 12
C IF M.NE.0 GO TO 12 SO THAT WE CAN REMOVE CODE NUMBER
      IC=0
C CLOCKWISE CHECKING ALL DONE REVERSE PARTIAL SEQUENCE
C TO SEARCH IN AN ANTICLOCKWISE DIRECTION.
      DO 20 J=1,15
      JREV=16-J
      IF(EXCLUD(I,JREV).EQ.0) GO TO 20
      IC=IC+1
      IEXCLD(IC)=EXCLUD(I,JREV)
20    CONTINUE
C CHECK FOR ANTICLOCKWISE PARTIAL SEQUENCE
      DO 21 J=IC+1,15
21    IEXCLD(J)=0
      CALL INDEX(IBUF,IEXCLD,,M)
      IF(M.NE.0) GO TO 12
10    CONTINUE
C PARTIAL SEQUENCE DOESN'T APPEAR ,PASS CODE NUMBER
```

```
        IGEN=.TRUE.
        GO TO 13
12      IGEN=.FALSE.
13      RETURN
        END

      SUBROUTINE TCAL(JA,JB,JC,JD,WA)
      DIMENSION AA(3,3),V1(3),V2(3)
        COMMON/ROTAT/NATOM(10,4),NR(10),NRATOM(10,50)
        COMMON/CONST/RAD1,RAD2,RADI,NRING,NT,NRI,NTP,NP,NAV
        COMMON/CHECK/IGEN,EXCLUD(10,15),NEXCLD,J1(15),NBONDS,
     1  DMAX
        COMMON/VARS/NATOMS,NAMES,XI,ICONN,ITYPES,TITLE,NTITLE
        COMMON/TAUS/TANGLE(9,2),TAU(10),NGEN,ICHEND,JCHEND,
     1  TDIST,TOL
        LOGICAL IGEN
        LOGICAL*1 EXCLUD,TITLE(80)
        DIMENSION NAMES(50,3),XI(50,3),ICONN(50,4),ITYPES(50),
     1  J1S(15)
      R1=0.0
      R2=0.0
      COSW=0.0
      DO 401 I=1,3
      AA(I,1)=XI(JA,I)-XI(JB,I)
      AA(I,2)=XI(JC,I)-XI(JB,I)
  401 AA(I,3)=XI(JC,I)-XI(JD,I)
      DO 402 J=1,3
        JTEM=J
      K=MOD(JTEM,3)+1
      L=MOD(K,3)+1
      V1(J)=AA(K,1)*AA(L,2)-AA(L,1)*AA(K,2)
      V2(J)=AA(K,2)*AA(L,3)-AA(L,2)*AA(K,3)
      R1=R1+V1(J)**2
  402 R2=R2+V2(J)**2
      R1=SQRT(R1)
      R2=SQRT(R2)
      DO 403 K=1,3
  403 COSW=COSW+(V1(K)/R1)*(V2(K)/R2)
      COSW=ARGCHK(COSW)
      WA=ACOS(COSW)*RADI
      SIGN1=AA(1,1)*(AA(2,2)*AA(3,3)-AA(2,3)*AA(3,2))-AA(2,1)
     1*(AA(1,2)*AA(3,3)-AA(1,3)*AA(3,2))+AA(3,1)*(AA(1,2)
     2*AA(2,3)-AA(1,3)*AA(2,2))
      WA=SIGN(WA,SIGN1)
      RETURN
      END

      FUNCTION ARGCHK(X)
      ARGCHK=SIGN(AMIN1(ABS(X),1.0),X)
      RETURN
      END
```

```
      FUNCTION ACOS(X)
      IF (ABS(X)-1.0000001) 5,5,6
 6    WRITE (7,10) X
 10   FORMAT('0ARGUMENT =',E15.6,'  > 1 IN ARC COSINE')
      CALL EXIT
 5    IF (ABS(X).GT.1.) X=SIGN(1.,X)
      IF(X) 3,4,3
 4    ACOS=1.5707964
      RETURN
 3    ACOS=ATAN(SQRT(1.-X*X)/X)
      IF(X)1,2,2
 1    ACOS=ACOS+3.1415927
 2    RETURN
      END

      SUBROUTINE BONDEX !CHECKS A CONF. FOR THE NO OF BONDS
        COMMON/ROTAT/NATOM(10,4),NR(10),NRATOM(10,50)
        COMMON/CONST/RAD1,RAD2,RADI,NRING,NT,NRI,NTP,NP,NAV
        COMMON/CHECK/IGEN,EXCLUD(10,15),NEXCLD,J1(15),NBONDS,
     1  DMAX
        COMMON/VARS/NATOMS,NAMES,XI,ICONN,ITYPES,TITLE,NTITLE
        COMMON/TAUS/TANGLE(9,2),TAU(10),NGEN,ICHEND,JCHEND,
     1  TDIST,TOL
        LOGICAL IGEN
        LOGICAL*1 EXCLUD,TITLE(80)
        DIMENSION NAMES(50,3),XI(50,3),ICONN(50,4),ITYPES(50),
     1  J1S(15)
        NBOND=0
        DO 10 I=1,NATOMS
        DO 10 J=I,NATOMS
        IF(I.EQ.J) GO TO 10
        DTEST=SQRT((XI(I,1)-XI(J,1))**2+(XI(I,2)-XI(J,2))**2+
     1(XI(I,3)-XI(J,3))**2)
        IF(DTEST.LE.DMAX) NBOND=NBOND+1
 10     CONTINUE
        IGEN=.TRUE.
        IF(NBOND.GT.NBONDS) IGEN=.FALSE.
        RETURN
        END

        SUBROUTINE SETUP
        COMMON/ROTAT/NATOM(10,4),NR(10),NRATOM(10,50),ISTART(16)
        COMMON/CHECK/IGEN,EXCLUD(10,15),NEXCLD,J1(15),NBONDS,
     1  DMAX,HOMOP
        COMMON/TCHECK/NREST,ITREST(3,4),TVAL(3),TTOL(3)
        COMMON/CONST/RAD1,RAD2,RADI,NRING,NT,NRI,NTP,NP,NAV
        COMMON/VARS/NATOMS,NAMES,XI,ICONN,ITYPES,TITLE,NTITLE
        COMMON/INPUT/INPLUN,LINNO,INBUF,INPNTR,NCHARS,ERROR,
     1  EMPTY,ENDSEC
        COMMON/TAUS/TANGLE(9,2),TAU(10),NGEN,ICHEND,JCHEND,
     1  TDIST,TOL
        DIMENSION NAMES(50,3),XI(50,3),ICONN(50,4),ITYPES(50)
        DIMENSION INDEXS(8)
        LOGICAL*1 EXCLUD,TITLE(80),INBUF(80),ANSWER,YES
        LOGICAL ERROR,EMPTY,ENDSEC,IGEN,HOMOP
        DATA YES /'Y'/
C READ IN AND PRINT MESSAGE CONTAINED IN FILE 'GMINFO.DAT'
C WHICH SHOULD RESIDE ON THE SYSTEM DISK.THERE IS NO LIMIT
C TO THE SIZE OF THIS FILE.
```

```
        CALL ASSIGN(1,'GMINFO.DAT')
        INPLUN=1
        LINENO=0
        INPNTR=0
1       CALL NXLINE
        IF(ENDSEC) GO TO 9
        WRITE(7,3)(INBUF(I),I=1,NCHARS)
        GO TO 1
3       FORMAT(X,72A1)
9       CALL CLOSE(1)
        INPLUN=7
C ALL DONE WITH THE START-UP MESSAGE. LETS GET ON WITH IT.
        TYPE 9300
9300    FORMAT('$DOES THE FOLLOWING DATA REFER TO A
     1  HOMOPOLYMER [Y/N] ')
        READ(7,9402) ANSWER
        HOMOP=.FALSE.
        IF(ANSWER.EQ.YES) HOMOP=.TRUE.
C GET THE NUMBER OF TORSION ANGLES
403     TYPE 9403
9403    FORMAT('$NUMBER OF VARIABLE TORSION ANGLES [I] ? ')
        CALL NXLINE
        CALL NUMBER(Q)
        IF(ERROR) GO TO 403
        NT=IFIX(Q)
C FIND OUT WHETHER IT'S PAIRS OR SINGLES HE 'S AFTER.
        TYPE 9401
9401    FORMAT('$TORSION ANGLES BY PAIRS [Y/N] ? ')
        READ(7,9402) ANSWER
9402    FORMAT(A1)
C YES CHOOSES PAIRWISE TORSION ANGLES.
        IF(ANSWER.EQ.YES) GO TO 410
        NP=1
C MAKE SURE HE KNOWS WHAT HE HAS DONE.
        TYPE 9410
9410    FORMAT('INDIVIDUAL TORSION ANGLES SELECTED .')
        GO TO 411
410     NP=2
411     CONTINUE
C NOW FIND OUT IF THE BOND ANGLES ARE TO BE ALLOWED TO CHANGE
C TO EFFECT RING CLOSURE (AS WELL AS TORSION ANGLES)
        TYPE 9411
9411    FORMAT('$ARE THE BOND ANGLES TO REMAIN FIXED [Y/N]')
        READ(7,9402) ANSWER
        IF(ANSWER.EQ.YES) GO TO 420
        TYPE 9420
9420    FORMAT('BOND ANGLES WILL VARY DURING RING CLOSURE.')
        NAV=2
        GO TO 421
420     NAV=1
421     CONTINUE
C READ IN NAMES AND COORDINATES OF THE RING ATOMS FROM FILE
        CALL GETFIL
C READ IN TORSION ANGLE SPECIFIERS
```

```
        NP4=NP*4
        NTP=NT/NP
        DO 432 J=1,NTP
430     TYPE 9430,NP
9430    FORMAT('$ENTER',I2,'TORSION ANGLE[S](ATOM LABELS)')
        CALL NXLINE
        CALL ATNAMS(NP4,INDEXS,NATOMS,NAMES)
        DO 431 I=1,NP4
        IF(INDEXS(I).LE.0) GO TO 8430
        IF(NP.EQ.2) GO TO 433
        NATOM(J,I)=INDEXS(I)
        GO TO 431
433     J2=J*2
        J21=J2-1
        IF(I.LE.4) NATOM(J21,I)=INDEXS(I)
        IF(I.GT.4) NATOM(J2,I-4)=INDEXS(I)
431     CONTINUE
        GO TO 432
8430    TYPE 9843
9843    FORMAT('ILLEGAL NAME IN TORSION SPECIFIER.TRY AGAIN.')
        GO TO 430
432     CONTINUE
C SO NOW WE KNOW WHICH TORSION ANGLES ARE VARYING. WE NOW NEED
C TO KNOW WHICH ATOMS TO ROTATE WHEN CHANGING THE TORSION
C ANGLES.THIS CAN BE DECODED FROM THE CONNECTIVITY (IN ICONN)
C I.E. WE NEED TO KNOW WHO IS CONNECTED TO (BY WHATEVER ROUTE)
C TO ATOM 3 ON THE SIDE AWAY FROM ATOM 2.(OBVIOUSLY 3 & 4 ARE
THE FIRST TWO). (WHERE 1,2,3 & 4 REFER TO THE T.A. SPECIFIERS)
        DO 440 I=1,NT
        NR(I)=2
        NRATOM(I,1)=NATOM(I,3)
        NRATOM(I,2)=NATOM(I,4)
        DO 441 INR=3,50
        IC1=NRATOM(I,INR-2)
        IC2=NRATOM(I,INR-1)
        NRI=NR(I)
        DO 442 J=1,4
        IC3=ICONN(IC2,J)
        IF(IC3.EQ.0) GO TO 442
        IF(IC3.EQ.IC1) GO TO 442
        NRATOM(I,INR)=IC3
        NR(I)=NR(I)+1
442     CONTINUE
        IF(NR(I).EQ.NRI) GO TO 440
441     CONTINUE
440     CONTINUE
C GET NUMBER OF GENERATORS AND THEN THE GENERATORS THEMSELVES.
        INPLUN=7
500     TYPE 9500
9500    FORMAT('$ENTER NUMBER OF GENERATORS [I] ')
        CALL NXLINE
        CALL NUMBER(Q)
        IF (ERROR) GO TO 500
        NGEN=IFIX(Q)
        GO TO (501,502) NP
501     TYPE 9501
```

```
9501     FORMAT(' ENTER A GENERATOR (IN DEGREES) [F] ')
         GO TO 503
502      TYPE 9502
9502     FORMAT('ENTER A GENERATOR(BY PAIRS,IN DEGREES)[F]')
503      DO 504 I =1,NGEN
5032     TYPE 9503
9503     FORMAT('$>')
         CALL NXLINE
         CALL NUMBER(TANGLE(I,1))
         IF(ERROR) GO TO 5031
         IF(NP.EQ.1) GO TO 504
         CALL NUMBER(TANGLE(I,2))
         IF(ERROR) GO TO 5031
         GO TO 504
5031     TYPE 9531
9531     FORMAT(' ***FORMAT ERROR. RETYPE GENERATOR.')
         GO TO 5032
504      CONTINUE
C GET NUMBER OF EXCLUSION STRINGS I.E. THOSE PARTIAL
C GENERATOR COMBINATIONS WHICH RESULT IN IMMEDIATE
CONFORMER DISQUALIFICATION AND THEN THE STRINGS THEMSELVES.
600      TYPE 9600
9600     FORMAT('$ENTER NUMBER OF EXCLUSION STRINGS(MAX10)[I]')
         CALL NXLINE
         CALL NUMBER(Q)
         IF(ERROR) GO TO 600
         NEXCLD=IFIX(Q)
         IF(NEXCLD.EQ.0) GO TO 700
         IF(NEXCLD.GT.10) NEXCLD=10
C ZERO EXCLUSION ARRAY.
         DO 601 I=1,10
         DO 601 J=1,15
601      EXCLUD(I,J)=0
         TYPE 9601
9601     FORMAT(' ENTER EXCLUSION STRINGS (MAX 14 INTEGERS)')
         DO 602 I=1,NEXCLD
         TYPE 9503
         READ(10,9602) J,(EXCLUD(I,K),K=1,J)
9602     FORMAT(Q,14A1)
602      CONTINUE
C READ IN TARGET DISTANCE AND DECODE CHAIN END ATOMS.
700      TYPE 9700
9700     FORMAT('$ENTER TARGET END-TO-END DIST AND MAX
     +   DEVIATION THEREFROM [F] ')
         CALL NXLINE
         CALL NUMBER(TDIST)
         IF(ERROR) GO TO 700
         CALL NUMBER(TOL)
         IF(ERROR) GO TO 700
         IC1=0
         DO 701 I=1,NATOMS
         IC=0
         DO 702 J=1,4
         IF(ICONN(I,J).GT.0) IC=IC+1
702      CONTINUE
         IF(IC.GT.2) GO TO 999
```

```
          IF(IC.NE.1) GO TO 701
          IC1=IC1+1
          IF(IC1.EQ.2) GO TO 703
          ICHEND=I
          GO TO 701
703       JCHEND=I
          GO TO 704
701       CONTINUE
704       CONTINUE
800       TYPE 9800
9800      FORMAT('$ENTER SHORTEST TOLERABLE NON-BONDED
     +    DISTANCE [F] ')
          CALL NXLINE
          CALL NUMBER(DMAX)
          IF(ERROR) GO TO 800
801       TYPE 9801
9801      FORMAT('$ENTER MAXIMUM PERMISSABLE NUMBER OF
     +    SUCH CONTACTS [I] ')
          CALL NXLINE
          CALL NUMBER(Q)
          IF(ERROR) GO TO 801
          NBONDS=IFIX(Q)+NATOMS
C FIND OUT IF HE WANTS ANY RESTRICTIONS ON THE GEOMETRY
C OF THE RING JUNCTION. (IN TERMS OF T.A.'S ONLY).
850       TYPE 9850
9850      FORMAT('$DO YOU WANT TO RESTRICT THE DEPENDENT
     +    TORSION ANGLES [Y/N] ')
          READ(7,9402) ANSWER
          NREST=0
          IF(ANSWER.NE.YES) GO TO 900
C*******OK HE WANTS SOME RESTRICTIONS - SO....
851       TYPE 9851
9851      FORMAT('$ENTER THE NUMBER OF DEPENDENT T.A.''S
     +    TO BE RESTRICTED [I] ')
          CALL NXLINE
          CALL NUMBER(Q)
          IF(ERROR) GO TO 851
          NREST=IFIX(Q)
C******THERE CAN ONLY BE THREE SO JUST MAKE SURE.
          IF(NREST.LE.3) GO TO 855
          TYPE 9853
9853      FORMAT(' YOU HAVE SPECIFIED TOO MANY RESTRICTED
     +    T.A.''S.')
          GO TO 851
855       DO 858 ITR=1,NREST
8551      TYPE 9855
9855      FORMAT('$ENTER DEPENDENT T.A., DESIRED VALUE &
     +    TOLERANCE ')
          CALL NXLINE
          CALL ATNAMS(4,INDEXS,NATOMS,NAMES)
          DO 856 I=1,4
          IF(INDEXS(I).LE.0) GO TO 857
856       ITREST(ITR,I)=INDEXS(I)
          CALL NUMBER(TVAL(ITR))
          IF(ERROR) GO TO 857
```

```
        CALL NUMBER(TTOL(ITR))
        IF(.NOT.ERROR) GO TO 858
857     TYPE 9857
9857    FORMAT(' ***FORMAT ERROR. TRY AGAIN. ')
        GO TO 8551
858     CONTINUE
C NOW THAT WE KNOW EXACTLY WHAT HE WANTS TO DO FIND OUT
C WHERE TO START IT ALL FROM.
900     TYPE 9900
9900    FORMAT('$ENTER STARTING CONFORMER NUMBER [I] ')
        CALL NXLINE
        CALL NUMBER(Q)
        IF(ERROR) GO TO 900
        ISTART(16)=IFIX(Q)-1
        TYPE 9901
9901    FORMAT('$ENTER STARTING GENERATOR STRING [I] ')
        READ(7,9902)(ISTART(I),I=1,NTP)
9902    FORMAT(15I1)
        DO 9001 I=(NTP+1),15
9001    ISTART(I)=NGEN
        RETURN
999     TYPE 9999
9999    FORMAT('YOU HAVE INPUT A BRANCHED CHAIN!')
        CALL EXIT
        END
C
C*****************************************
C
C
        SUBROUTINE GETFIL
C
C OPENS A CRYSTAL FILE, RETURNS THE ATOM NAMES, ORTHO COORDS,
C CONNECTIVITY, AND TYPES. THEN CLOSES THE FILE.
C
C NOTE: THIS MUST BE LINKED WITH SYSLIB BECAUSE OF THE CALL CLOSE
C
        COMMON/VARS/NATOMS,NAMES,X,Y,Z,ICONN,ITYPES,TITLE,NTITLE
        COMMON/INPUT/INPLUN,LINNO,INPBUF,INPNTR,NCHARS,ERROR,EMPTY
     1  ENDSEC
        DIMENSION NAMES(50,3),X(50),Y(50),Z(50),ICONN(50,4),
     +  ITYPES(50)
        LOGICAL*1 TITLE(80),INPBUF(80)
        LOGICAL ERROR,EMPTY,ENDSEC
        LOGICAL EXISTS,GOTITL
        DATA LUNTTY/7/
C
C CHOOSE A CHANNEL FOR INPUT
        LUN=11
C
C PROMPT FOR NAME OF FILE
        INPLUN=LUNTTY
 1      TYPE 9000
 9000   FORMAT('$NAME OF INPUT FILE :')
C
C READ REPLY TO INPBUF, SEE IF FILE EXISTS
23      CALL GFILN(EXISTS)
        IF(EMPTY) RETURN        !BLANK LINE MEANS EXIT
        IF(EXISTS) GOTO 7
```

```
C
C MESSAGE IF FILE DOSN'EXIST
        TYPE 9007
 9007   FORMAT('+*** FILE DOSN''T EXIST')
        GOTO 1                      !GO PROMPT AGAIN
C
C OPEN THE FILE, SWITCH INPUT FROM TTY TO FILE
 7      CALL ASSIGN(LUN,INPBUF,NCHARS)
        INPLUN=LUN
        LINENO=0
        GOTITL=.FALSE.
        TITLE(1)=0
C
C PRINT TITLE  IN COLUMN 1
        CALL POSN('TITL')           !FIND LINE BEGINNING WITH TITLE
        IF(ERROR) GOTO 100          !NO SUCH LINE IN FILE
 2      TYPE 9010, (INPBUF(I),I=1,NCHARS)
 9010   FORMAT(X,80A1)
        CALL NXLINE                 !PRINT ALL LINES UNTIL      1
        IF(ENDSEC) GOTO 3           !FOUND ONE BEGINNING WITH  **
        IF(GOTITL) GOTO 2
        DO 6 I=1,NCHARS
 6      TITLE(I)=INPBUF(I)          !SAVE FIRST LINE OF TITLE
        NTITLE=NCHARS
        IF(NTITLE.GT.72) NTITLE=72
        TITLE(NTITLE+1)=0
        GOTITL=.TRUE.
        GOTO 2
C
C PRINT COMMENTS
 3      CALL POSN('COMM')           !FIND LINE BEGINNING WITH COMM
        IF(ERROR) GOTO 10           !NO SUCH LINE IN FILE
 5      TYPE 9010, (INPBUF(I),I=1,NCHARS)
        CALL NXLINE                 !PRINT ALL LINES UNTIL
        IF(.NOT.ENDSEC) GOTO 5      !FOUND ONE BEGINNING WITH **
C
10      CONTINUE
C
C GET ATOM NAMES AND ORTHOGONAL COORDINATES
        CALL COORDS(NATOMS,NAMES,X,Y,Z)
        IF(ERROR) GOTO 130          !NO LINE BEGINNING WITH COOR
C
C GET CONNECTIVITY
 30     CALL CONECT(NATOMS,NAMES,ICONN)
        IF(ERROR) TYPE 9030         !NO LINE BEGINNING WITH CONN
 9030   FORMAT(' *** NO CONNECTIVITY SECTION IN FILE')
C
C CHECK CONNECTIVITY
        TYPE 9040
 9040   FORMAT(' NOW I''M CHECKING FOR IRREGULARITIES')
        CALL OKCONN(NATOMS,NAMES,X,Y,Z,ICONN,ITYPES)
C
C GET TYPES
        CALL GTYPES(NATOMS,NAMES,ITYPES)
        IF(ERROR) TYPE 9060         !NO LINE BEGINNING WITH ATOM
 9060   FORMAT(' *** NO ATOM TYPES SECTION IN FILE')
```

```
C CLOSE THE FILE AND RELEASE THE UNIT
 99     CALL CLOSE(INPLUN)
C
C THATS ALL FOLKS
        EMPTY=.FALSE.
        RETURN
C
C NO TITLE: PRESUME ,NOT A CRYSTAL FILE. ASK FOR ANOTHER
 100    TYPE 9100
 9100   FORMAT('NO TITLE. NOT AN ACCEPTABLE FILE')
999     CALL CLOSE(INPLUN)
        GOTO 1                    !GO TRY AGAIN
C
C NO COORDINATES
 130    TYPE 9130
 9130   FORMAT(' *** NO COORDINATES SECTION IN FILE')
        GOTO 999
        END
C
C
C*****************************************
C
C
        SUBROUTINE COORDS(N,NAMES,X,Y,Z)
C
C READS IN THE ATOM NAMES AND FRACTIONAL COORDINATES.
C THE COORDS ARE THEN ORTHONORMALISED (USING TMAT)
C AND RETURNED IN X,Y,Z
C N              RETURNS THE NUMBER OF ATOMS READ IN
C NAMES(I,J)     J=1,3 RETURNS THE NAME OF ATOM I IN 3A2
C                FORMAT PADDED WITH NULLS
C X(I),Y(I),Z(I) RETURN THE ORTHONORMAL COORDS OF ATOM I
C
C MAXN    MAXIMUM NUMBER OF ATOMS ALLWED IN
C         SET THIS TO THE DIMENSION OF X,Y,Z,NAMES ARRAYS
C
        LOGICAL*1 INPBUF(80)
        LOGICAL*1 OPBRAK
        LOGICAL AGAIN,NEWPIC,HYDRO,BALLS,INTENS,LABEL,INVTED
        LOGICAL ERROR,EMPTY,ENDSEC,CHARG
        DIMENSION NAMES(50,3),X(50),Y(50),Z(50),C(3)
        COMMON/INPUT/INPLUN,LINENO,INPBUF,INPNTR,NCHARS,ERROR,
     +  EMPTY,ENDSEC
        DATA OPBRAK/'('/
        DATA MAXN/50/
        DATA ITAG/20/
        DATA AGAIN,NEWPIC,HYDRO,BALLS,INTENS,LABEL,
     +  INVTED/F,T,T,F,T,T,F/
        DATA ISTAGE,CHARG /1,F/
C
C POSTION FILE AT COORDINATES SECTION
        CALL POSN('COOR')
        IF(ERROR) RETURN
C
C GET EACH ATOM NAME AND COORDS
        N=0
```

```
C       --READ IN NEXT LINE--
 10     CALL NXLINE
        IF(ENDSEC) GOTO 30       !END OF LIST
        IF(N.GE.MAXN) GOTO 80    !TOO MANY ATOMS
        N=N+1
C       --GET NAME--
        CALL GETNAM( NAMES(N,1),NAMES(N,2),NAMES(N,3) )
        IF(ERROR) GOTO 90
C       -- REPORT IF ALREADY GOT THAT NAME --
        CALL NAMNUM(N-1,NAMES,NAMES(N,1),NAMES(N,2),NAMES(N,3),
     +  INDEX)
        IF(INDEX.GT.0) TYPE 9010, (NAMES(N,I),I=1,3)
 9010   FORMAT(' ATOM NAME ',3A2,' IS USED MORE THAN ONCE')
C       --READ FRACTIONAL COORDINATES INTO C(1),C(2),C(3)--
        DO 20 I=1,3
         CALL NUMBER(C(I))
         IF(ERROR) GOTO 90
         IF(INPBUF(INPNTR).EQ.OPBRAK) CALL NUMBER(SIGMA)
         IF(ERROR) GOTO 90
 20     CONTINUE
        X(N)=C(1)
        Y(N)=C(2)
        Z(N)=C(3)
        GOTO 10          !LOOP FOR NEXT ATOM
 30     IF(N.EQ.0) ERROR=.TRUE. !ERROR IF NO ATOMS
        RETURN
C
C TOO MANY ATOMS
 80     TYPE 9080, MAXN
 9080   FORMAT(' ****TOO MANY ATOMS, ONLY FIRST',I4,
     1          ' HAVE BEEN ACCEPTED')
        RETURN
C
C ERROR
 90     TYPE 9000, (INPBUF(I),I=1,NCHARS)
 9000   FORMAT('FORMAT ERROR IN COORDINATE SECTION'/X,80A1)
        GOTO 10
        END              !OF SUBROUTINE COORDS
C
C
C*****************************************
C
C
        SUBROUTINE CONECT(NATOMS,NAMES,ICONN)
C
C READS IN THE CONNECTIVITY FILE AND SETS UP THE
C CONNECTIVITY MATRIX
C
C NATOMS         NUMBER OF ENTRIES IN NAMES
C NAMES(I,J)     J=1,3 GIVES THE NAME OF ATOM I IN 3A2
C                FORMAT PADDED WITH NULLS
C ICONN(I,J)     J=1,4 GIVES INDICIES OF ATOMS BONDED TO
C                ATOM I (LIMIT=4)
C ERROR SET .TRUE. IF KEYWORD CONN (ECTIVITY) IS MISSING
C FROM INPUT FILE
C FORMAT ERRORS IN THE FILE DO NOT SET ERROR BUT THEY DO
C CAUSE A MESSAGE TO BE PRINTED ON THE TTY.
```

```
C
C
        LOGICAL*1 INPBUF(80)
        LOGICAL ERROR,EMPTY,ENDSEC
        DIMENSION NAMES(50,3), ICONN(50,4), INDEXS(6)
        COMMON/INPUT/INPLUN,LINENO,INPBUF,INPNTR,NCHARS,ERROR,
     1  EMPTY,ENDSEC
C
C ZERO THE CONNECTIVITY ARRAY
        DO 1 I=1,NATOMS
        DO 1 J=1,4
 1      ICONN(I,J)=0
C
C POSITION INPUT FILE AT CONNECTIVIY
        CALL POSN('CONN')
        IF(ERROR) RETURN
C
C READ NEXT NON-BLANK LINE INTO INPUT BUFFER
 10     CALL NXLINE
        IF(ENDSEC) GOTO 30       !END OF CONNECTIVITY LIST
        IF(NCHARS.EQ.0) GOTO 10 !IGNORE BLANK LINES
C
C GET INDICIES FOR THE ATOM NAMES.
        CALL ATNAMS(5,INDEXS,NATOMS,NAMES)
        IF(ERROR) GOTO 90
        I=INDEXS(1)
C
C FILL IN CONNECTIVITY MATRIX
        DO 20 J=2,5
 20     ICONN(I,J-1)=INDEXS(J)
        GOTO 10                              !LOOP FOR NEXT LINE
C
C FOUND END OF SECTION
 30     ERROR=.FALSE.
        RETURN
C
C FOUND AN ERROR
 90     TYPE 9000, (INPBUF(I),I=1,NCHARS)
 9000   FORMAT('FORMAT ERROR IN CONNECTIVITY SECTION'/X,80A1)
        GOTO 10                              !LOOP FOR NEXT LINE
C
        END              !OF SUBROUTINE CONECT
C
C
C*************************************************
C
C
        SUBROUTINE GTYPES(NATOMS,NAMES,ITYPES)
C
C RETURNS THE ATOM TYPES
C
C NATOMS         NUMBER OF ENTRIES IN NAMES
C NAMES(I,J)     J=1,3 GIVES THE NAME OF ATOM I IN 3A2
C                FORMAT PADDED WITH NULLS
C ITYPES(I)      RETURNS THE TYPE CODE OF ATOM I
```

```
C
C ERROR SET .TRUE. IF KEYWORD "TYPE" IS MISSING
C
C FORMAT ERRORS IN THE FILE DO NOT SET ERROR BUT THEY
C DO CAUSE A MESSAGE TO BE PRINTED ON THE TTY.
C
C
      LOGICAL*1 INPBUF(80)
      LOGICAL ERROR,EMPTY,ENDSEC
      DIMENSION NAMES(50,3),ITYPES(50)
      COMMON/INPUT/INPLUN,LINENO,INPBUF,INPNTR,NCHARS,ERROR,
     1  EMPTY,ENDSEC
C
C ZERO THE ITYPES ARRAY
      DO 1 I=1,NATOMS
 1    ITYPES(I)=0
C
C POSITION THE INPUT FILE AT THE KEY WORD 'ATOM'
      CALL POSN('ATOM')
      IF(ERROR) RETURN          !NOT THERE: ERROR RETURN
C
C READ NEXT NON-BLANK LINE INTO THE INPUT BUFFER
 10   ERROR=.FALSE.
      ENDSEC=.TRUE.
      CALL NXLINE
      IF(ENDSEC) RETURN         !END OF LIST: NORMAL RETURN
      IF(NCHARS.EQ.0) GOTO 10 !IGNORE BLANK LINES
C
C GET TYPE
      CALL NUMBER(TYPE)
      IT=IFIX(TYPE)
      IF(ERROR) GOTO 90
C
C GET LIST OF NAMES, FILL IN ITYPES ARRAY
C     -- GET NEXT NAME --
 20   CALL GETNAM(NAM1,NAM2,NAM3)
      IF(EMPTY.AND.INPNTR.GE.NCHARS) GOTO 10
      IF(ERROR) GOTO 90
C     -- GET INDEX ASSOCIATED WITH NAME --
      CALL NAMNUM(NATOMS,NAMES,NAM1,NAM2,NAM3,INDEX)
      IF(INDEX.LE.0) GOTO 90
C     -- FILL IN ITYPES ARRAY --
      ITYPES(INDEX)=IT
      GOTO 20                   !LOOP FOR NEXT NAME
C FOUND A FORMAT ERROR
 90   TYPE 9000, (INPBUF(I),I=1,NCHARS)
 9000 FORMAT('FORMAT ERROR IN TYPES SECTION'/X,80A1)
      GOTO 10                   !LOOP FOR NEXT LINE
C
      END              !OF SUBROUTINE GTYPES
C
C*****************************************
C
      SUBROUTINE OKCONN(NATOMS,NAMES,X,Y,Z,ICONN,ITYPES)
C
C CHECKS THE CONNECTIVITY MATRIX (ICONN) FOR THE FOLLOWING:
C IF ATOM I IS CONNECTED TO J,  IS J CONNECTED TO I. REPORT
C IF NOT IF ATOM I IS CONNECTED TO J,  IS DISTANCE I,J < 1.8
C REPORT IF NOT IF DISTANCE I,J < 1.8,  IS I CONNECTED TO J
C REPORT IF NOT D.S.RICHARDSON  CHEMISTRY DEPT., UNIVERSITY
C COLLEGE LONDON  12/75
```

```
C NATOMS          NUMBER OF ATOMS
C NAMES(I,J)      J=1,3 NAME OF ATOM I IN 3A2 FORMAT
C X(I),Y(I),Z(I) ORTHONORMAL COORDINATES OF ATOM I
C ICONN(I,J)      J=1,4 UP TO 4 ATOMS CONNECTED TO ATOM I
        DIMENSION X(50),Y(50),Z(50),NAMES(50,3),ICONN(50,4)
        DIMENSION ITYPES(50)
        DATA DISMAX,HHMIN/1.8,1.3/
        DO 30 I=1,NATOMS
        DO 10 K=1,4
C FIND CONNECTED ATOMS I,J
        J=ICONN(I,K)
        IF(J.EQ.0) GOTO 10
C SEE IF TOO FAR APART
        DISTIJ=SQRT((X(I)-X(J))**2+(Y(I)-Y(J))**2+(Z(I)-
     +  Z(J))**2)
        IF(DISTIJ.GT.DISMAX)   TYPE 9005,
     1  (NAMES(I,L),L=1,3),(NAMES(J,L),L=1,3),DISTIJ
 9005     FORMAT(' BONDED ATOMS ',3A2,',',3A2,' SEPARATED BY',
     1  F7.2)
C SEE IF J IS CONNECTED TO I
        DO 5 L=1,4
        M=ICONN(J,L)
        IF(M.EQ.0) GOTO 9        !NOT CONNECTED
        IF(M.EQ.I) GOTO 10       !ARE CONNECTED
 5      CONTINUE
 9      TYPE 9009,  (NAMES(I,L),L=1,3),(NAMES(J,L),L=1,3)
 9009   FORMAT(' INCONSISTENT CONNECTIVITY FOR ',3A2,',',3A2)
10      CONTINUE
C CHECK DISTANCE BETWEEN I AND ALL OTHER ATOMS
        IF(I.EQ.NATOMS) RETURN
        IP1=I+1
        DO 30 J=IP1,NATOMS
        DISTIJ=SQRT((X(I)-X(J))**2+(Y(I)-Y(J))**2+(Z(I)-
     1  Z(J))**2)
        IF(DISTIJ.GT.DISMAX) GOTO 30
        DO 25 K=1,4
        IF(ICONN(I,K).EQ.J) GOTO 30
 25     CONTINUE
C REPORT IF LESS THAN DISMAX AND YET NOT CONNECTED
        IF(ITYPES(I).NE.15)GOTO 27
        IF(ITYPES(J).NE.15)GOTO 27
C UNLESS I AND J ARE BOTH HYDROGENS CONNECTED TO THE SAME
C ATOM THEN ONLY REPORT IF LESS THAN HHMIN
        IF(ICONN(I,1).EQ.ICONN(J,1).AND.DISTIJ.GT.HHMIN) GOTO 30
 27     TYPE 9025, (NAMES(I,L),L=1,3),(NAMES(J,L),L=1,3),DISTIJ
 9025   FORMAT(' NB ATOMS ',3A2,',',3A2,' SEPARATED BY',F7.2)
 30     CONTINUE
        RETURN
        END
```

```
      SUBROUTINE GFILN(EXISTS)
C
C USES NXLINE TO INPUT A FILE NAME.  IF THE NAME LOOKS
C LEGAL THEN THIS ROUTINE CHECKS TO SEE IF SUCH A FILE
C ALREADY EXISTS AND SETS EXISTS ACCORDINGLY. IF THE
C NAME LOOKS WRONG THE USER IS ASKED TO TRY AGAIN.
C
C EXISTS        SET .TRUE. IF THE FILE  EXISTS
C               SET .FALSE. IF THE FILE DOSN'T  EXIST
C
C (PASSED THROUGH COMMON)
C INPBUF  TEXT STRING CONTAINING THE FILE NAME
C NCHARS  NUMBER OF CHARACTERS IN INPBUF
C EMPTY   SET .TRUE. IF THE USER INPUTS A BLANK LINE
C ERROR   SET .TRUE. IF NO CHANNELS ARE AVAILABLE, OR
C         IF A BLANK LINE IS INPUT
C
C
      LOGICAL*1 INPBUF(80)
      LOGICAL ERROR,EMPTY,ENDSEC
      LOGICAL EXISTS
      COMMON/INPUT/INPLUN,LINENO,INPBUF,INPNTR,NCHARS,ERROR,
     1  EMPTY,ENDSEC
      LOGICAL ERR
      INTEGER DEVBLK(4)
C
C READ LINE INTO INPUT BUFFER
 10   CALL NXLINE
      IF(NCHARS.EQ.0) GOTO 40          !BLANK LINE
11    EMPTY=.FALSE.
C
C MARK END OF LINE WITH ZERO AS REQUIRED BY ROUTINE FILR50
      INPBUF(NCHARS+1)=0
C
C CONVERT FILE NAME INTO STANDARD RAD 50 FILE DESCRIPTOR BLOCK
      CALL FILR50(INPBUF,DEVBLK,ERR)
      IF(ERR) GOTO 20
C
C GET A FREE RT11 CHANNEL, MARK IT IN USE TO FORTRAN
      LUN=IGETC()
      IF(LUN.LT.0) GOTO 50
C
C BE SURE THE SPECIFIED DEVICE HANDLER IS .FETCH'D
      IF(IFETCH(DEVBLK(1)).NE.0) GOTO 20
      IF(DEVBLK(2).EQ.0) GOTO 15
C
C TRY TO OPEN THE FILE WITH A .LOOKUP
      IF(LOOKUP(LUN,DEVBLK).LE.0) GOTO 30
C
C .LOOKUP WAS SUCCESFUL.  CLOSE THE FILE, RELEASE THE CHANNEL
      CALL CLOSEC(LUN)
      CALL IFREEC(LUN)
      EXISTS=.TRUE.
```

```
 15     ERROR=.FALSE.
        RETURN
C FORMAT ERROR
 20     TYPE 9020
 9020   FORMAT('$TRY AGAIN PLEASE:')
        GOTO 10
C FILE NOT FOUND
 30     EXISTS=.FALSE.
        GOTO 15
C
C BLANK LINE
 40     EMPTY=.TRUE.
        GOTO 55
C
C NO FREE CHANNELS
 50     TYPE 9050
 9050   FORMAT('+*** NO FREE CHANNELS')
 55     ERROR=.TRUE.
        EXISTS=.FALSE.
        RETURN
        END
C
C*************************************************
C
        SUBROUTINE FILR50(FILNAM,DEVBLK,ERROR)
C
C CONVERTS AN RT11 FILE NAME (TEXT STRING)NAME INTO
C A STANDARD FORMAT 4 WORD RAD 50 FILE DESCRIPTOR BLOCK
C FILNAM        TEXT STRING CONTAINING THE FILE NAME
C               TERMINATED WITH A NULL (ZERO) BYTE.
C DEVBLK        4 WORD ARRAY WHICH RETURNS THE REQUIRED
C               RAD 50 FILE DESCRIPTOR BLOCK
C ERROR         SET .TRUE. IF THE TEXT STRING CONTAINS A
C               FORMAT ERROR
C DEFAULT DEVICE IS GIVEN IN DEFDEV
C DEFAULT EXTENSION IS GIVEN IN DEFEXT
C DEFAULT NAME IS GIVEN IN DEFNAM
C
C
        LOGICAL*1 FILNAM(81), COLN, BLANK, DOT
        LOGICAL LEAD, ERROR
        INTEGER DEVBLK(4), DEFDEV, DEFEXT, DEFNAM(2)
        DATA BLANK,COLN,DOT/' ',':','.'/
        DATA DEFDEV/3RDK /
        DATA DEFEXT/3RDAT/
        DATA DEFNAM/2*0/
C
C ZERO THE FILE DESCRIPTOR BLOCK
        DO 5 I=1,4
 5      DEVBLK(I)=0
C
C FIND POSITION OF FIRST NON-BLANK CHARACTER, POSN OF
C COLN, POSN OF DOT, AND NUMBER OF CHARACTERS
        LEAD=.TRUE.
        ICOLN=0
        IDOT=0
        N=0
```

```
C       -- GET NEXT CHARACTER --
 9      IBLANK=N
 10     IF(N.EQ.80) GOTO 20
        IF(FILNAM(N+1).EQ.0) GOTO 20
        N=N+1
        IF(LEAD.AND.FILNAM(N).EQ.BLANK) GOTO 9
        LEAD=.FALSE.
        IF(FILNAM(N).EQ.COLN) ICOLN=N           !FOUND :
        IF(FILNAM(N).EQ.DOT) IDOT=N             !FOUND .
        GOTO 10
C DEVICE SECTION
 20     IF(LEAD) GOTO 90                !BLANK LINE
        IF(ICOLN.EQ.0) GOTO 25          !USE DEFAULT
        NDEV=ICOLN-IBLANK-1
        IF(NDEV.LE.0.OR.NDEV.GT.3) GOTO 90
C       -- CONVERT TO RAD 50, PUT INTO DEVBLK(1),
     1  GOTO 90 IF ERROR --
        IF(NDEV.NE.IRAD50(NDEV,FILNAM(IBLANK+1),
     1  DEVBLK(1))) GOTO 90
        GOTO 30
C USE DEFAULT DEVICE
 25     DEVBLK(1)=DEFDEV
        ICOLN=IBLANK
C EXTENSION SECTION
 30     IF(IDOT.EQ.0) GOTO 35
        NEXT=N-IDOT
        IF(NEXT.LE.0) GOTO 90
        IF(NEXT.GT.3) NEXT=3
C       -- CONVERT TO RAD 50, PUT INTO DEVBLK(4),
     1  GOTO 90 IF ERROR --
        IF(NEXT.NE.IRAD50(NEXT,FILNAM(IDOT+1),DEVBLK(4)))
     1  GOTO 90
        GOTO 40
C USE DEFAULT EXTENSION
 35     DEVBLK(4)=DEFEXT
        IDOT=N+1
C FILE NAME
 40     NUM=IDOT-ICOLN-1
        IF(NUM.EQ.0) GOTO 45
        IF(NUM.LT.0.OR.NUM.GT.6) GOTO 90
C       -- CONVERT TO RAD 50, PUT INTO DEVBLK(2&3),
     1  GOTO 90 IF ERROR --
        IF(NUM.NE.IRAD50(NUM,FILNAM(ICOLN+1),DEVBLK(2)))
     1  GOTO 90
 44     ERROR=.FALSE.
        RETURN
C USE DEFAULT NAME
 45     DEVBLK(2)=DEFNAM(1)
        DEVBLK(3)=DEFNAM(2)
        GOTO 44
C FOUND AN ERROR
 90     ERROR=.TRUE.
        DO 95 I=1,4
 95     DEVBLK(I)=0
        RETURN
        END
```

```
      PROGRAM GLOAP
      COMMON /TRANS/ TANGLE(9,2),NATOM(10,4),NT,XS(50,3)
      COMMON /ROTA/ RAD1,NR(10)
      COMMON /COOR/ XI(50,3),NRATOM(10,50),NATOMS,RADI
      COMMON /LSAR/ ICHEND,JCHEND
      COMMON /GLO/ TDIST,TOL,DT(8),NTP,NP,NAV,J1(15),A
      DIMENSION DELTAU(2),TAU(10)
      DO 8 I=1,NATOMS
      DO 8 J=1,3
      XI(I,J)=XS(I,J)
 8    CONTINUE
      DO 9 I=1,NTP
      DO 9 J=1,NP
      IJ=I
      IF(NP.EQ.2) IJ=2*(I-1)+J
      TAU(IJ)=TANGLE(J1(I),J)
 9    CONTINUE
C
      A=0.
      DO 100 I=1,NT
      ITEM=I
      CALL CFRAME(ITEM)
      CALL TCAL(NATOM(I,1),NATOM(I,2),NATOM(I,3),NATOM(I,4),WA)
      CALL ROT(ITEM,-WA,0)
      CALL ROT(ITEM,TAU(I),0)
 100  CONTINUE
C
      CALL LSQCAL(ICHEND,JCHEND,DIST2)
      DIST=SQRT(DIST2)
      IF(ABS(TDIST-DIST).LE.TOL) GOTO 105
      DO 101 I=1,8
      DELTAU(1)=DT(I)
      DELTAU(2)=-DT(I)
      DO 101 L=1,NAV
      DO 102 J=1,NT
      JTEM=J
      IF(NP.EQ.2.AND.L.EQ.2.AND.MOD(JTEM,2).EQ.0) GOTO 102
      CALL CFRAME(JTEM)
      DO 103 K=1,2
      CALL ROT(JTEM,DELTAU(K),L-1)
      CALL LSQCAL(ICHEND,JCHEND,DIST2)
      DISTN=SQRT(DIST2)
      DELDIS=ABS(TDIST-DISTN)
      IF(DELDIS.GE.ABS(TDIST-DIST)) GOTO 104
      DIST=DISTN
      IF(DELDIS.LE.TOL) GOTO 105
      GOTO 102
 104  CALL ROT(JTEM,-DELTAU(K),L-1)
 103  CONTINUE
 102  CONTINUE
 101  CONTINUE
 105  A=1.
      RETURN
      END
```

```
      SUBROUTINE CFRAME(I)
      COMMON /TRANS/ TANGLE(9,2),NATOM(10,4),NT,XS(50,3)
      COMMON /COOR/ XI(50,3),NRATOM(10,50),NATOMS,RADI
    DIMENSION DC(3,3),XX(3),XOA(3),XA(3)
    R12SQ=0.0
    R23SQ=0.0
    RSQ=0.0
    COSANG=0.0
      N1=NATOM(I,1)
      N2=NATOM(I,2)
      N3=NATOM(I,3)
    DO 9 J=1,3
  9 XOA(J)=-XI(N2,J)
    DO 10 J=1,3
    DC(3,J)=XI(N3,J)-XI(N2,J)
    DC(2,J)=XI(N1,J)-XI(N2,J)
    IF(ABS(DC(3,J)).LT.1.0E-10) DC(3,J)=0.0
    IF(ABS(DC(2,J)).LT.1.0E-10) DC(2,J)=0.0
    R23SQ=R23SQ+DC(3,J)*DC(3,J)
    R12SQ=R12SQ+DC(2,J)*DC(2,J)
 10 CONTINUE
    R23=SQRT(R23SQ)
    R12=SQRT(R12SQ)
    DO 11 K=1,3
    DC(3,K)=DC(3,K)/R23
    DC(2,K)=DC(2,K)/R12
    COSANG=COSANG+DC(3,K)*DC(2,K)
 11 CONTINUE
    RX=-R12*COSANG
    DO 12 L=1,3
    XX(L)=XI(N3,L)-(R23+RX)*DC(3,L)
    DC(2,L)=XI(N1,L)-XX(L)
    IF (ABS(DC(2,L)).LT.1.0E-10) DC(2,L)=0.0
    RSQ=RSQ+DC(2,L)*DC(2,L)
 12 CONTINUE
    R=SQRT(RSQ)
    DO 13 M=1,3
    DC(2,M)=DC(2,M)/R
 13 CONTINUE
    DO 14 N=1,3
      NTEM=N
    K=MOD(NTEM,3)+1
    L=MOD(K,3)+1
    DC(1,N)=DC(2,K)*DC(3,L)-DC(2,L)*DC(3,K)
 14 CONTINUE
    DO 16 J=1,NATOMS
    DO 15 K=1,3
     XA(K)=0.0
    DO 15 L=1,3
 15 XA(K)=XA(K)+DC(K,L)*(XI(J,L)+XOA(L))
    DO 16 M=1,3
 16 XI(J,M)=XA(M)
    RETURN
    END
```

```
      SUBROUTINE TCAL(JA,JB,JC,JD,WA)
        COMMON/COOR/XI(50,3),NRATOM(10,50),NATOMS,RADI
        DIMENSION AA(3,3),V1(3),V2(3)
      R1=0.0
      R2=0.0
      COSW=0.0
      DO 401 I=1,3
      AA(I,1)=XI(JA,I)-XI(JB,I)
      AA(I,2)=XI(JC,I)-XI(JB,I)
  401 AA(I,3)=XI(JC,I)-XI(JD,I)
      DO 402 J=1,3
        JTEM=J
      K=MOD(JTEM,3)+1
      L=MOD(K,3)+1
      V1(J)=AA(K,1)*AA(L,2)-AA(L,1)*AA(K,2)
      V2(J)=AA(K,2)*AA(L,3)-AA(L,2)*AA(K,3)
      R1=R1+V1(J)**2
  402 R2=R2+V2(J)**2
      R1=SQRT(R1)
      R2=SQRT(R2)
      DO 403 K=1,3
  403 COSW=COSW+(V1(K)/R1)*(V2(K)/R2)
        CALL ARGCHK(COSW)
        CALL ABCOS(COSW,WA)
      WA=WA*RADI
      SIGN1=AA(1,1)*(AA(2,2)*AA(3,3)-AA(2,3)*AA(3,2))-
     1AA(2,1)*(AA(1,2)*AA(3,3)-AA(1,3)*AA(3,2))+AA(3,1)
     1*(AA(1,2)*AA(2,3)-AA(1,3)*AA(2,2))
      WA=SIGN(WA,SIGN1)
      RETURN
      END
C
        SUBROUTINE ARGCHK(X)
        X=SIGN(AMIN1(ABS(X),1.0),X)
        RETURN
        END
C
C
        SUBROUTINE ABCOS(X,WA)
        IF(ABS(X)-1.0000001) 5,5,6
 6      CONTINUE
        RETURN
 5      IF(ABS(X).GT.1.) X=SIGN(1.,X)
        IF(X) 3,4,3
 4      WA=1.5707964
        RETURN
 3      WA=ATAN(SQRT(1.-X*X)/X)
        IF(X) 1,2,2
 1      WA=WA+3.1415927
 2      RETURN
        END
```

```
      SUBROUTINE  ROT(I,ANG,IJ)
        COMMON/ROTA/RAD1,NR(10)
        COMMON/COOR/XI(50,3),NRATOM(10,50),NATOMS,RADI
C
        COSANG=COS(ANG*RAD1)
        SINANG=SIN(ANG*RAD1)
        IF(IJ.EQ.1)GOTO 20
        N=NR(I)
      DO 10 J=2,N
      X1=XI(NRATOM(I,J),1)*COSANG-XI(NRATOM(I,J),2)*SINANG
      X2=XI(NRATOM(I,J),1)*SINANG+XI(NRATOM(I,J),2)*COSANG
      XI(NRATOM(I,J),1)=X1
      XI(NRATOM(I,J),2)=X2
   10 CONTINUE
      RETURN
C
   20 DO 21 J=1,N
      X1=XI(NRATOM(I,J),2)*COSANG-XI(NRATOM(I,J),3)*SINANG
      X2=XI(NRATOM(I,J),2)*SINANG+XI(NRATOM(I,J),3)*COSANG
      XI(NRATOM(I,J),2)=X1
      XI(NRATOM(I,J),3)=X2
   21 CONTINUE
      RETURN
      END
C
      SUBROUTINE LSQCAL(I,J,DIST2)
        COMMON/COOR/XI(50,3),NRATOM(10,50),NATOMS,RADI
      DIMENSION DIR(3),DIR2(3)
      DO 703 K=1,3
      DIR(K)=XI(I,K)-XI(J,K)
  703 DIR2(K)=DIR(K)*DIR(K)
      DIST2=DIR2(1)+DIR2(2)+DIR2(3)
      RETURN
      END
```

```
C
C ---------------
C DESCRIPTION OF THE INPUT BUFFER AND ITS ASSOCIATED
C VARIABLES AND FLAGS. THESE ARE USED BY THE FOLLOWING TEXT
C PROCESSING ROUTINES NXLINE, NUMBER, GETNAM, POSN
C
C ALL ARGUMENTS ARE HELD IN LABELLED COMMON /INPUT/
C
C INPLUN  LOGICAL UNIT NUMBER OF INPUT DEVICE
C LINENO  CURRENT LINE NUMBER IN FILE
C INPBUF  INPUT BUFFER: HOLDS 80 CHARACTERS IN A1 FORMAT
C         NOTE THIS IS A BYTE ARRAY
C INPNTR  BUFFER POSITION POINTER.  SET TO ZERO WHEN A NEW LINE
C         IS PUT INTO THE INPUT BUFFER.  POINTS TO THE LAST
C         CHARACTER PROCESSED.
C NCHARS  NUMBER OF CHARS IN THE INPUT BUFFER.  RESET WHEN NEW
C         LINE IS PUT INTO THE BUFFER.
C ERROR   LOGICAL FLAG SET .TRUE. IF THE LOOKED-FOR ITEM HAS A
C         FORMAT ERROR OR IS 'EMPTY' OR IF A SEARCH OPERATION
C         FAILS.
C EMPTY   LOGICAL FLAG SET .TRUE. IF THE LOOKED-FOR ITEM IS 'EMPT
C         (USEFUL WHEN SUPPLYING DEFAULT VALUES)
C         EG.  1, ,3,4, ,6  THE SECOND AND FIFTH ITEMS ARE 'EMPTY
C                    C ENDSEC  SET .TRUE. IF LINE STARTS WITH **
C         (** USED TO MARK END OF SECTION IN CRYSTAL DATA FILES)
C
C
C*****************************************
C
C
        SUBROUTINE NXLINE
C
C READS THE NEXT LINE FROM THE CURRENT INPUT DEVICE (UNIT
C NUMBER=INPLUN) INTO THE BYTE ARRAY INPBUF (THE INPUT BUFFER).
C  LINENO IS INCREMENTED. INPNTR IS SET TO ZERO AND NCHARS=NUMBER
OF CHARACTERS READ IN. IF THE FIRST TWO CHARACTERS ARE ** OR
IF END-OF-FILE IS ENCOUNTERED THEN ENDSEC IS SET .TRUE.,
C OTHERWISE IT IS SET .FALSE.
C D.S.RICHARDSON   12/75  CHEMISTRY DEPT., UCL LONDON
C
        LOGICAL*1 INPBUF(80)
        LOGICAL ERROR,EMPTY,ENDSEC
        LOGICAL*1 ISTAR
        COMMON/INPUT/INPLUN,LINENO,INPBUF,INPNTR,NCHARS,ERROR,
      1 EMPTY,ENDSEC
        DATA ISTAR/'*'/
C
 1      READ(INPLUN,9000,END=10) NCHARS,(INPBUF(I),I=1,NCHARS)
 9000   FORMAT(Q,80A1)  !RETURNS NUMBER OF CHARACTERS READ IN
        INPNTR=0
        LINENO=LINENO+1
        IF(INPBUF(1).EQ.ISTAR.AND.INPBUF(2).EQ.ISTAR) GOTO 10
        ENDSEC=.FALSE.
        RETURN
C
C FOUND END OF SECTION OR END OF FILE
 10     ENDSEC=.TRUE.
```

```
        RETURN
        END                !OF SUBROUTINE NXLINE
C
C
C*****************************************
C
C
        SUBROUTINE NUMBER(X)
C
C X RETURNS THE VALUE OF THE NUMBER REPRESENTED BY THE
C CHARACTER STRING STARTING AT THE CURRENT POSITION OF INPBUF.
C !!! X MUST BE REAL. IF YOU WANT TO INPUT AN INTEGER
C   THEN DO
C        CALL NUMBER(X)
C        INTEGER=IFIX(X)  OR  INTEGER=INT(X)
C
C ON ENTRY INPNTR = POSN-1 OF FIRST CHARACTER OF No IN INPBUF
C ON EXIT INPNTR = POSN OF TERMINATOR (NCHARS IF LINE END FOUND)
C
C LEADING SPACES ARE IGNORED AND THE NUMBER IS TERMINATED BY
C ONE OF:- END OF LINE
C  A SECOND DECIMAL POINT
C  A NON LEADING + OR - OR SPACE
C  ANY OTHER NON-NUMERIC CHARACTER
C EMPTY IS SET .TRUE. IF THE FIRST CHARACTER ENCOUNTERED IS
C NOT NUMERIC. (IN THIS CASE THE VALUE OF X IS UNCHANGED -
C ALLOWS X = DEFAULT VALUE) ERROR IS SET IN THE SAME WAY AS
C EMPTY NOTE THIS ROUTINE DOES NOT FLAG ILLEGAL CHARACTERS -
C THE CALLING ROUTINE SHOULD INSPECT THE TERMINATOR TO SEE IF
C IT IS LEGAL.
C
        LOGICAL*1 INPBUF(80)
        LOGICAL ERROR,EMPTY,ENDSEC
        LOGICAL*1 IBLANK,IPOINT,IPLUS,IMINUS,I0,I9
        COMMON/INPUT/INPLUN,LINENO,INPBUF,INPNTR,NCHARS,ERROR,
     1  EMPTY,ENDSEC
        DATA IBLANK,IPOINT,IPLUS,IMINUS,I0,
     1  I9/' ','.','+','-','0','9'/
C
        EMPTY=.TRUE.
        SIGN=+1.0
        VAL=0.0
        C=0.0
C
C GET NEXT CHARACTER
 5      IF(INPNTR.GE.NCHARS) GOTO 10              !END OF BUFFER
        INPNTR=INPNTR+1
        ICHAR=INPBUF(INPNTR)
        IF(EMPTY.AND.ICHAR.EQ.IBLANK) GOTO 5
        IF(ICHAR.EQ.IMINUS) GOTO 30               !ITS A MINUS
        IF(ICHAR.EQ.IPLUS) GOTO 40                !ITS A PLUS
        IF(ICHAR.EQ.IPOINT) GOTO 50               !ITS A DECIMAL
        IF(ICHAR.GE.I0.AND.ICHAR.LE.I9) GOTO 20 !ITS NUMERIC
```

```
C IT'S A TERMINATOR
C (ANY NON NUMERIC CHARACTER OR END OF LINE
C  OR NON LEADING SPACE, SIGN OR POINT)
 10     ERROR=EMPTY
        INPNTR=INPNTR-1
        IF(EMPTY) RETURN
        X=VAL*SIGN
        RETURN
C
C ITS NUMERIC
 20     EMPTY=.FALSE.
        IF (C.GT.0.) GOTO 25
        VAL=10.*VAL+ICHAR-I0
        GOTO 5
C
C FRACTIONAL PART
 25     C=0.1*C
        VAL=(ICHAR-I0)*C + VAL
        GOTO 5
C
C IT'S A MINUS SIGN
 30     IF (.NOT.EMPTY) GOTO 10
        SIGN=-1.0
C
C IT'S A PLUS SIGN
 40     IF (.NOT.EMPTY) GOTO 10
        EMPTY=.FALSE.
        GOTO 5
C
C IT'S A DECIMAL POINT
 50     IF(C.NE.0) GOTO 10
        C=1.0
        EMPTY=.FALSE.
        GO TO 5
C
        END             !OF SUBROUTINE NUMBER
C
C
C*****************************************
C
C
        SUBROUTINE GETNAM(N1AND2,N3AND4,N5AND6)
C
C RETURNS THE NEXT NAME FROM THE INPUT BUFFER AND UPDATES
C THE POSITION POINTER
C
C LEADING SPACES ARE IGNORED. THE NAME SHOULD START WITH A
C LETTER AND THEREAFTER BE EITHER LETTERS OR NUMBERS.  IT
C IS TERMINATED BY A NON-LEADING SPACE OR ANY NON-ALPHANUMERIC
C CHARACTER OR END-OF-LINE. THE NAME IS TRUNCATED TO 5
C CHARACTERS - ALL CHARACTERS AFTER THE FIFTH ARE IGNORED.
C
C ON ENTRY INPNTR = POSITION-1 OF FIRST CHARACTER OF THE
C  NAME IN INPBUF
```

```
C ON EXIT
C INPNTR POSITION OF TERMINATOR (NCHARS IF LINE END FOUND)
C N1AND2 RETURNS THE 1ST 2 CHARACTERS OF NAME IN A2 FORMAT
C N3AND4 RETURNS THE THIRD AND FOURTH CHARACTERS
C N5AND6 RETURNS THE 5TH CHAR.+ A SPACE IN THE 6TH POSITION
C
C IF THE NAME IS LESS THAN 5 CHARACTERS THE UNUSED CHARACTER
C POSITIONS IN N1AND2,N3AND4,N5AND6 ARE SET TO SPACES
C
C IF THE NAME DOSN'T START WITH A LETTER THEN ON EXIT
C ERROR=EMPTY=.TRUE.
C N1AND2,N3AND4,N5AND6 ARE UNCHANGED
C
C
      LOGICAL*1 INPBUF(80)
      LOGICAL ERROR,EMPTY,ENDSEC
      LOGICAL*1 ICHAR,IBLANK,I0,I9,IA,IZ,NAMTEM(6)
      COMMON/INPUT/INPLUN,LINENO,INPBUF,INPNTR,NCHARS,ERROR,
     +  EMPTY,ENDSEC
      EQUIVALENCE (N12,NAMTEM(1)),(N34,NAMTEM(3)),
     +  (N56,NAMTEM(5))
      DATA IBLANK,IA,IZ/' ','A','Z'/
C
      EMPTY=.TRUE.
      ICOUNT=1          !NUMBER OF CHARACTERS IN THE NAME + 1
C
C GET NEXT CHARACTER
 5    IF(INPNTR.GE.NCHARS) GOTO 10            !END OF  BUFFER
      INPNTR=INPNTR+1
      ICHAR=INPBUF(INPNTR)                    !GOT THE CHAR.
      IF(EMPTY.AND.ICHAR.EQ.IBLANK) GOTO 5    !IGNORE BLANKS
      IF(ICHAR.GE.IA.AND.ICHAR.LE.IZ) GOTO 30 !ITS ALPHABETIC
      IF(ICHAR.GE."41.AND.ICHAR.LE."100)GOTO 20 !ITS PRINTABLE
C
C IT'S A TERMINATOR
 10   ERROR=EMPTY
      IF(EMPTY) RETURN
      DO 11 I=ICOUNT,6
 11   NAMTEM(I)=IBLANK          !PAD REST OF NAME WITH SPACES
      N1AND2=N12
      N3AND4=N34                !NOTE EQUIVALENCE WITH NAMTEM
      N5AND6=N56                !(CAN'T EQIVALENCE DUMMYS VARS)
      RETURN
C
C IT'S NUMERIC
 20   IF(EMPTY) GOTO 10         !TERMINATOR IF FIRST CHARACTER
C
C IT'S O.K.  STORE CHARACTER IN NAMTEM.
 30   EMPTY=.FALSE.
      IF(ICOUNT.GE.6) GOTO 5          !IGNORE CHARS AFTER 5TH
      NAMTEM(ICOUNT)=ICHAR
      ICOUNT=ICOUNT+1
      GOTO 5
      END               !OF SUBROUTINE GETNAM
```

```
      SUBROUTINE POSN(KEYTEX)
C
C POSITION INPUT FILE AT LINE BEGINNING WITH THE
C CHARACTER STRING GIVEN IN KEYTEX (ONLY THE NUMBER
C OF CHARACTERS GIVEN ARE MATCHED, THE REMAINDER OF
C THE LINE IS IGNORED).
C
C INPLUN = LOGICAL UNIT NUMBER FOR THE INPUT FILE
C
C ON ENTRY LINENO=CURRENT LINE NUMBER IN FILE (0 MEANS
C NO LINES YET READ) ON EXIT LINENO = LINE NUMBER IN FILE
C WHERE MATCH FOUND IF NO MATCH IS FOUND THEN ERROR IS SET
C .TRUE. AND LINENO RETAINS IT'S INITIAL VALUE
C
C KEYTEX MUST END WITH A NULL (ZERO) BYTE.  (RT11 FORTRAN
C STRING CONSTANTS SATISFY THIS REQUIREMENT)
C
C EG: CALL POSN('UNITCELL') POSITIONS THE INPUT FILE AT THE
C NEXT LINE BEGINNING WITH THE CHARACTERS 'UNITCELL'
C
C THE SEARCH STARTS AT THE CURRENT POSITION AND GOES FORWARD
C LINE BY LINE.  IF THE END OF THE FILE IS REACHED THEN THE
C FILE IS REWOUND AND THE SEARCH CONTINUES UNTIL THE WHOLE FILE
C IS SEARCHED. USE WITH CARE IF THE DEVICE IS NON-REWINDABLE.
C
      LOGICAL*1 INPBUF(80)
      LOGICAL ERROR,EMPTY,ENDSEC
      LOGICAL*1 KEYTEX(80),INULL
      COMMON/INPUT/INPLUN,LINENO,INPBUF,INPNTR,NCHARS,ERROR,
     1  EMPTY,ENDSEC
      DATA INULL/0/
C START SEARCH AT BEGINNING OF FILE
      REWIND INPLUN
      LINENO=0
      ERROR=.FALSE.
C
C READ NEXT LINE
 1    READ(INPLUN,9001,END=900) NCHARS,(INPBUF(I),I=1,NCHARS)
 9001 FORMAT(Q,80A1)
      LINENO=LINENO+1
      INPNTR=0
C
C SEE IF CHARACTERS MATCH
      DO 2 I=1,NCHARS
      IF(KEYTEX(I).EQ.INULL) RETURN          !FOUND A MATCH
      IF(KEYTEX(I).NE.INPBUF(I)) GOTO 1      !NO MATCH
 2    CONTINUE
      IF(KEYTEX(NCHARS+1).EQ.INULL) RETURN   !FOUND A MATCH
      GOTO 1                                 !TOO FEW CHARS IN LINE
C
C END OF FILE, IE KEYTEX NOT IN FILE. SET ERROR FLAG
 900  ERROR=.TRUE.
      RETURN
      END              !OF SUBROUTINE POSN
```

```
      SUBROUTINE ATNAMS(N,INDEXS,NATOMS,NAMES)
C
C RETURNS ASSOCIATED INDICIES FOR THE NEXT N ATOMS NAMED IN
C THE INPUT BUFFER
C
C N             NUMBER OF NAMES WE'RE LOOKING FOR (LIMIT=20)
C INDEXS(K)     RETURNS THE INDEX OF THE KTH NAME FROM THE
C               CURRENT POSITION IN THE INPUT BUFFER.
C               SET TO 0 IF THE NAME IS EMPTY( TERMINATOR)
C               SET TO -1 IF THE NAME IS NOT IN NAMES ARRAY
C                OR IF A FORMAT ERROR IS FOUND
C               SET TO -2 IF THE NAME IS '*'  'WILD CARD'
C               THE FIRST FORMAT ERROR ABORTS THE ROUTINE, SO
C               ONLY ONE -1 WILL BE FOUND IN THE INDEXES ARRAY
C NATOMS        NUMBER OF ENTRIES IN THE NAMES ARRAY No ATOMS
C NAMES(I,J)    J=1,3 GIVES NAME OF ATOM I IN 3A2 FORMAT
C               PADDED WITH SPACES
C
C ERROR         SET .TRUE. IF A FORMAT ERROR IS FOUND, OR IF
C              ANY NAME IS NOT IN THE NAMES ARRAY, OR IF A
C              'WILD CARD' IS FOUND. ELSE SET .FALSE.
C
C REFER TO SUBROUTINE NXLINE FOR A DESCRIPTION OF THE INPUT BUFFEF
C
      LOGICAL*1 INPBUF(80)
      LOGICAL*1 STAR,FILL
      LOGICAL ERROR,EMPTY,ENDSEC
      LOGICAL ERRTEM
      DIMENSION NAMES(50,3)
      DIMENSION INDEXS(20)
      COMMON/INPUT/INPLUN,LINENO,INPBUF,INPNTR,NCHARS,ERROR,
     1 EMPTY,ENDSEC
      DATA STAR/'*'/
C
      ERROR=.FALSE.
      ERRTEM=.FALSE.  !SET .TRUE. IF NAME NOT IN NAMES ARRAY
C
      DO 10 I=1,N
C      --GET NEXT NAME INTO NAM1,NAM2,NAM3--
       CALL GETNAM(NAM1,NAM2,NAM3)
       IF(.NOT.ERROR) GOTO 9
C        ERROR CAUSES CORRESPONDING INDEX TO BE SET TO -1
 8       INDEXS(I)=-1
C        AND THE ROUTINE IS ABORTED WITH ERROR SET .TRUE.
         IF(.NOT.EMPTY) RETURN
C        UNLESS IT WAS EMPTY, THEN INDEX IS SET TO 0
         INDEXS(I)=0
         ERROR=.FALSE.
C        --WILD CARD (*) IS SPECIAL CASE --
         IF(INPBUF(INPNTR).NE.STAR)   GOTO 10
         CALL GETNAM(NAM1,NAM2,NAM3)          !MOVE TO NEXT ONE
         IF(.NOT.EMPTY) GOTO 8
         INDEXS(I)=-2                         ! SET INDEX TO -2
         GOTO 10
```

```
          GOTO 10
C         --LOOK UP NAME AND RETURN CORRESPONDING INDEX--
 9        CALL NAMNUM(NATOMS,NAMES,NAM1,NAM2,NAM3,INDEXS(I))
          IF(INDEXS(I).LT.0) ERRTEM=.TRUE.
 10       CONTINUE
          ERROR=ERRTEM
          RETURN
          END                !OF SUBROUTINE ATNAMS
C
C
C*****************************************
C
C
          SUBROUTINE NAMNUM(NATOMS,NAMES,N1AND2,N3AND4,N5AND6,INDEX)
C
C SEARCHES THE NAMES ARRAY UNTIL IT FINDS A MATCH WITH THE
C NAME HELD IN N1AND2,N3AND4,N5AND6.  THE CORRESPONDING NAMES
C ARRAY INDEXIS RETURNED IN INDEX.  IF NO MATCH IS FOUND
C INDEX IS SET TO -1.
C
C NATOMS          NUMBER OF ENTRIES IN THE NAMES ARRAY
C NAMES(I,J)      J=1,3 GIVES NAME OF ATOM I IN 3A2 FORMAT
C                 PADDED WITH SPACES
C N1AND2          HOLDS THE FIRST TWO CHARS OF NAME IN A2 FORMAT
C N3AND4          HOLDS THE THIRD AND FOURTH CHARACTERS
C N5AND6          HOLDS THE 5TH CHARACTER + A SPACE IN THE 6TH POSN
C INDEX           RETURNS THE INDEX OF THE CORRESPONDING ATOM.
C                 SET TO -1 IF THERE IS NO ATOM OF THAT NAME
C
C IF THE NAME IS LESS THAN 5 CHARACTERS THE UNUSED CHARACTER
C  POSITIONS IN N1AND2,N3AND4,N5AND6 MUST BE SPACES
C
C
          LOGICAL*1 INPBUF(80)
          LOGICAL ERROR,EMPTY,ENDSEC
          DIMENSION NAMES(50,3)
          COMMON/INPUT/INPLUN,LINENO,INPBUF,INPNTR,NCHARS,ERROR,
     1    EMPTY,ENDSEC
C
          INDEX=0
 1        INDEX=INDEX+1
          IF(INDEX.GT.NATOMS) GOTO 90
          IF(    NAMES(INDEX,1).NE.N1AND2
     1       .OR.NAMES(INDEX,2).NE.N3AND4
     2       .OR.NAMES(INDEX,3).NE.N5AND6 )GOTO 1 !LOOP IF NO MATCH
          RETURN                                  !EXIT WHEN FOUND
C
C NO MATCH IN WHOLE ARRAY
 90       INDEX=-1
          RETURN
          END                !OF SUBROUTINE NAMNUM
```