A COMPUTATIONAL STUDY OF TRISHOMOCUBANE AMINO ACID DIPEPTIDE

BY

POOMANI PENNY GOVENDER

A dissertation submitted in partial fulfilment of the requirements for the degree of Master of Technology in the Department of Chemistry, Durban Institute of Technology, Durban

DECLARATION

I hereby declare that the work presented in this thesis is my own work and has never before been submitted for any degree at this or any other Technikon.

PP GOVENDER

(Student Number: 29414)

15TH day of FEBRUARY 2004

ACKNOWLEDGEMENTS

I would like to extend my thanks and gratitude to the following people: -

Drs K. Bisetty and H.G. Kruger for their assistance, guidance, patience and encouragement throughout this research project.

The support from the Centre de Supercomputació de Catalunya (CESCA), Barcelona Spain, for the generous allocation of computer time.

Financial assistance from the Durban Institute of Technology, the University of Kwa-Zulu Natal, Durban and the National Research Foundation.

My friends and colleagues from the past and present, wherever they may be.

Finally, a special thank you to my parents for always being there for me, for their support and encouragement is gratefully acknowledged.

"Om Shri Sai Ram"

ABSTRACT

4-amino-(D_3)-trishomocubane-4-carboxylic acid (tris-amino acid) is a constrained α-amino acid residue that exhibits peculiar conformational characteristics. The aim of the present study is to provide a deeper understanding of these features, which can be used as a guide when choosing Trishomocubane as suitable building blocks for peptide design. The C_α carbon of Trishomocubane forms part of the cyclic structure, and consequently a peptidic environment was simulated with an acetyl group on its N-terminus and a methylamide group on its C-terminus. This study involved a complete exploration of the conformational profile of Trishomocubane using computational techniques.

The parm94 parametization of the AMBER[♣] force-field was used to explore the conformational space of the peptide, Trishomocubane. The Ramachandran maps computed at the molecular mechanics level[♦] with the parm94 force-field parameters compared reasonably with the corresponding maps computed at the Hartree Fock (HF) level, using the 6-31G* basis set. The results of this study revealed that the conformational profile of the Trishomocubane peptide can be characterized by four low energy regions, viz., C_{7ax}, C_{7eq}, 3_{10} and α_L helical structures.

♣ Computational time spent on each job was approximately 15 minutes. A total of 450 jobs were run.

♦ Computational time spent on each job run from a desktop computer was 4-5 days, on a super computer 3-4 hours was required. A total of 450 jobs were run. The jobs run, exclude all trials and mistrials done when initially setting up.

CONTENTS

SUPPLEMENTARY MATERIAL:

A CD accompanying this Thesis includes the following:

- The cartesian coordinates for all 3D structures presented
- The amber preparation input file
- Text for all chapters including references used

LIST OF FIGURES

PAGE NO./S

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Peptides are highly flexible molecules that could potentially occupy a variety of conformational space.[5] These molecules interact with their receptors in very precise conformations. The bioactive conformation is not usually known, unless analytical studies using X-ray diffraction or NMR spectroscopy conducted on the ligand-receptor complex are available.[1]

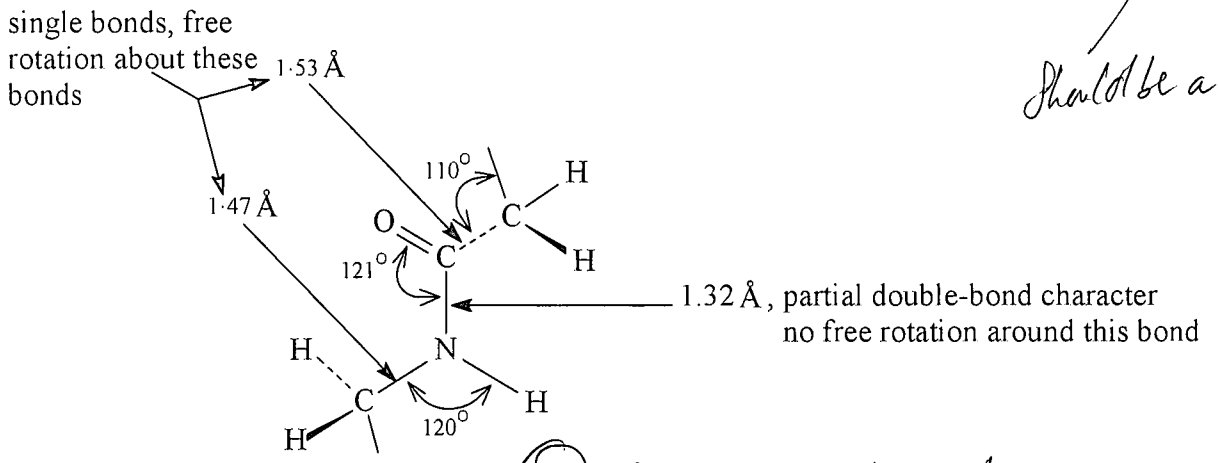
Amino acids play a large role in synthesizing pharmaceutical drugs or biologically active compounds since they comprise the main ingredient of these drugs. Conformationally constrained amino acids are also found in many naturally occurring biologically active compounds eg. alamethicin, a peptide antibiotic.[2] The introduction of constrained amino acids, or larger structural units such as bicyclic dipeptides, into peptides is an important technique for the study of the conformation and their relation to bioactivity.[3] In the past, information has been gathered on conformational preferences of some peptide analogues and a number of naturally occurring amino acid modifications have been proposed in literature.[1] There are several strategies to constrain amino acid residues to reduce their energy-accessible conformational space. Among them, substitution of the α-carbon hydrogen to restrict both backbone and the side-chain conformational freedom, has been widely investigated.[4,5] Methylation of the C_α carbon is one of the simplest examples, to produce aminoisobutyric (Aib) acid. The residue (Aib) exhibits a strong tendency to stabilize 3_{10} and α-helical conformations,[1] this has been established by analysis of a large number of synthetic peptides using crystallographic and theoretical studies.[1] Another possibility is to generate unnatural amino acid residues, which include the C_α carbon in the cyclic structure. A similar residue inducing conformational rigidity which has been extensively studied, is the 1-aminocyclopropanecarboxylic acid (Ac3c). X-ray structures

structures of some peptide analogues containing cyclopropane amino acids show that they adopt a preferred conformation with ϕ near $\pm 80^0$ and ψ near 0^0 or, 180^0 in some cases.[1] This uniqueness confers the amino acid the property to induce β-turn secondary structures in peptides.[1]

Peptides consisting of a number of amino acids assume characteristic conformations. The order of amino acids in the backbone of a polypeptide is known as the primary structure of the peptide. The conformation of the peptide chain gives it a secondary structure and further bending and folding of peptide chains create the tertiary structure of proteins.[6] These local conformations have been classified in terms of regular folding patterns called helices, pleated sheets and turns.[7] All the amino acids found in naturally occurring peptides and proteins have the same configuration, and this imparts a stereochemical regularity (having typical bond lengths and bond angles for the peptide group-see below)[6] to peptide chains. They assume different conformations depending mainly on the types of amino acids that make up the chain.[6]

1

Should be a figure.

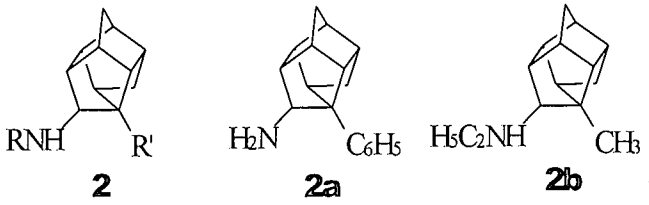


1 Caption and label properly.

The conformations arise from twisting of the backbone because of large or bulky substituent groups positioned on the polypeptide chain. These bulky groups introduce

repulsive interactions with each other therefore in order to reduce these unfavourable steric effects these groups rotate away from each other. This deviation from the totally planar arrangement results in various folding patterns.[7]

The incorporation of the polycyclic cage peptides into conformations of bioactive molecules led to a steady increase of biochemical systems and processes.[8,9] The field of peptide design and protein engineering has expanded due to the synthesis of unnatural amino acids, which in turn resulted in the design of molecules as potentially useful medicinal agents.[8,9,10] Specifically, the incorporation of polycyclic cage compounds into medicinal drugs has shown positive effects on the properties of drugs.[11] The inherent steric bulk of rigid Trishomocubane cage structures, such as compound **2**[12], prove to be advantageous in the slowing down of drug degradation thus making drug administration to patients less frequent.[11]



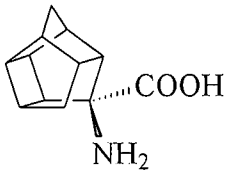
2 **2a** **2b**

In order to understand the function of peptides and proteins in biological processes, knowledge of their molecular structure is an essential prerequisite. Therefore, structural investigations on these novel compounds are most important. The conformational profile of different unnatural amino acids has been extensively reported in literature.[13,14,15,16]

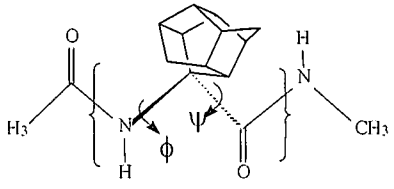
Presently, our research group at the University of Kwa-Zulu Natal, Durban is actively involved in asymmetric synthesis and solid phase peptide synthesis of polycyclic cage derivatives. The polycyclic cage derivatives currently under investigation by the group is (R)-8-amino-pentacyclo[5.4.0.02,6.03,10.05,9] undecane-8-carboxylic (PCU) acid[17],

compound **3a** and 4-amino-(D_3)-trishomocubane-4-carboxylic acid (tris-amino acid), compound **4a**.[18,19]

Indeed computational methods can provide useful insights to this field. As part of a wider research project, the focus of this study is aimed at the understanding of the conformational features induced by the incorporation of conformational constrained trishomocubane amino acid[18,20] into peptides. This work follows a similar study performed by Bisetty et al.,[21,22] who performed a conformational study on the (R)-8-amino-pentacyclo[5.4.0.02,6.03,10.05,9] undecane-8-carboxylic acid dipeptide[17] (Ac-PCU-NHMe), compound **3b**, using computational techniques. The conformational profile of the PCU cage residue was assessed, by studying the cage dipeptide **3b** as model compound, with an acetyl group on its N-terminus and with a methylamide group on its C-terminus to simulate a peptidic environment.



3a



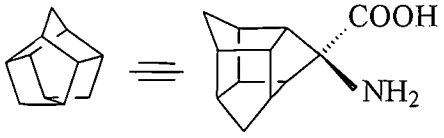
Ac Cage NHMe

3b

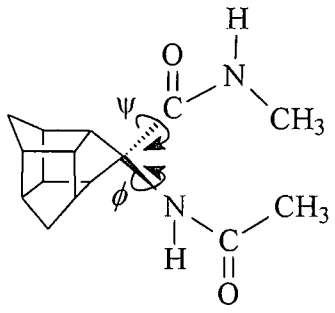
Caption and label.

AMBER

Bisetty et al.,[21] carried out calculations at both the ab-initio level and the molecular mechanics level, using the amber forcefield. They concluded that the PCU cage residue exhibits preferences for helical conformations as well as inducing promising β-turn[21,22,23] characteristics. Consequently, their work inspired us to undertake a similar investigation on the Trishomocubane peptide (Ac-Trish-NHMe) **(4b)** and to compare the results with those obtained for the PCU cage peptide **3b**.



4a



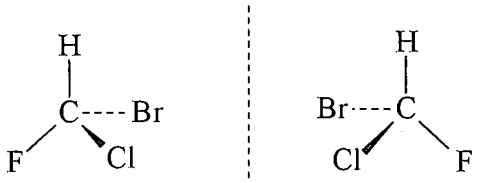
4b

Caption and label figures.

It is expected that a computational study on the conformational characteristics of an additional cage amino acid (**4a**) should enhance our understanding of the manipulation of the three-dimensional structure of peptides. In pursuing the main goal of this work in peptide conformational analysis of the cage dipeptide, new atomic charges (parameters) for the tris-amino acid cage residue need to be generated by fitting the molecular electrostatic potential (ESP) charges, computed at the Hartree-Fock (HF) level using the 6-31G* basis set. The development of new parameters involves the application of the restrained electrostatic potential (RESP) fit model consistent with the parm94 set of parameters for the AMBER force-field.[24] Once these new parameters have been computed they can then be incorporated into the database of the AMBER force-field using the AMBER 5.0 molecular modelling computer program. One reputable method for the critical evaluation of such force-fields has been their ability to reproduce the structures and energetics of a small group of molecules that may be regarded, as models for larger peptides, for which there are comparable experimental or theoretical data.[25]

Not worded correctly.

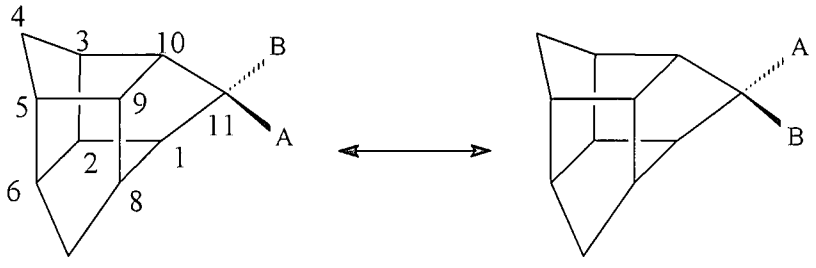
The properties of a protein depend primarily on its three dimensional structure.[26] When a tetrahedral carbon is unsymmetrically substituted the phenomenon of chirality arises. Chirality is a term that describes the "handedness" or mirror image of a molecule. With reference to 5[27], chirality can also be described as a molecule that contains a carbon atom with four different substituents.



5 Caption not laid properly.

Such carbon atoms are termed chiral centres, asymmetric centres or stereogenic centres.[27] Chiral molecules are directly linked to the optical activity of that molecule i.e enantiomers*(R or S configuration) and diastereomers℘.

Trishomocubane is a chiral molecule, shown in 6 below, with only two enantiomers and not any of the expected four diastereomers. This can be explained by the axis of symmetry through each of the methylene groups and the centre of the opposite C-C bond, as well as an axis of symmetry (D_3) passing through C_2 and C_9.



enantiomer A (1) **enantiomer A(2)**

6

---

* Enantiomers are mirror images of each other and have the same physical and chemical properties. Only physical, in an achiral, environment difference is that they are able to rotate plane polarized light in the opposite directions.

℘ Stereoisomers that are not enantiomers are diastereomers. Compounds with 2 or more chiral centers can have many different stereoisomers. Diastereomers possess different chemical and physical properties. A molecule can have only one enantiomer, but it may have many diastereomers.[27]

The model peptide molecule, compound **4b** (derived from **4a**) was chosen on the basis that the residue lies in an environment similar to that on a polypeptide chain. The Ramachandran map,[28] a potential energy map, can be used to investigate the conformational space of an amino acid residue. These maps are generated from systematic rotation of the backbone torsion angles, ϕ and ψ defined by the N-C_α and C_α-C bonds, respectively. Most of the naturally occurring amino acid residues can access about 30% of the (ψ,ϕ) space. For this purpose the Ramachandran plot of the Trishomocubane cage peptide was computed at the molecular mechanics level with the parm94 force field parameter[29,30] of the AMBER 5 computer program.[31] In addition the same maps were computed using an *ab-initio* method at the Hartree-Fock level using the 6-31G* basis set.[32] The *ab-initio* results were used as a benchmark and gives a qualitative indication of the molecular mechanics results.

CHAPTER 2

THEORETICAL TOOLS

Computational chemistry encompasses a number of different types of calculations, based on a wide range of theoretical approaches, ranging from simple computer graphics-based analysis to quantum mechanical calculations of molecular energies and electronic structures. Theoretical calculations fall into two broad categories *viz*, quantum mechanical studies, which try to directly calculate molecular structures and energies from the Schrödinger equation, and molecular mechanics calculations which uses empirical and semi-empirical energy calculations.

2.1 ELECTRONIC STRUCTURE METHOD

Electronic structure methods use the laws of quantum mechanics as the basis for their computations. Quantum mechanics states that the energy and other related properties of a molecule may be obtained by solving the Schrödinger equation, and explains how entities like electrons have both particle-like and wave-like characteristics. The Schrödinger equation, below, describes the wave function of a particle shown below;[33]

$$\left[-\frac{h^2}{8\pi^2}\sum_i \frac{1}{m_j}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+V(r)\right].\Psi(r)=E.\Psi(r) \qquad (2.1)$$

where, Ψ is the wave function, *m* is the mass of the particle, *h* is Planck's constant, and V is the potential field in which the particle is moving. This equation can also be written in its abbreviated form shown in equation 2.2 below.

$$H\Psi = E\Psi \qquad (2.2)$$

H is the Hamiltonian, an operator to derive the kinetic energy (KE) and potential energy (PE) of a system of electrons and nuclei.

$$H = KE + PE \tag{2.3}$$

where

$$KE = \left[\frac{-h^2}{8\pi^2} \sum_i \frac{1}{m_j} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \right] \tag{2.4}$$

and

$$PE = \sum_i \sum_{(j} \left(\frac{e_i e_j}{r_{ij}} \right) \tag{2.5}$$

In equations 2.4 and 2.5 the sums are over all particles, i with masses m_j and over all distinct pairs of particles i and j with electric charges e_i and e_j separated by a distance r_{ij}, respectively.

2.1.1 Molecular Orbital Theory[33]

A complete description of an orbital must be a function, not only of the spatial coordinates, but also of the spin coordinates ξ. If one defines two spin functions $\alpha(\xi)$ and $\beta(\xi)$ such that

$$\alpha(+\frac{1}{2}) = 1 \qquad \beta(+\frac{1}{2}) = 0 \tag{2.6}$$

i.e., for an electron with 'spin up' ($\uparrow$), $\xi = +\frac{1}{2}$, the value of $\alpha(\xi)$ is 1 and $\beta(\xi)$ is 0, and for an electron with 'spin down' ($\downarrow$), $\xi = -\frac{1}{2}$, the values of the two

functions are transposed (0 and 1, respectively). Then, the product of the spatial molecular orbital function $\psi(r)$ with $\alpha(\xi)$ or $\beta(\xi)$ will include the electron spin in the overall electronic wave function. The resulting functions $\chi = \psi(r)\alpha(\xi)$ or $\psi(r)\beta(\xi)$ are called the spin orbital. Thus the molecular orbital theory is an approach to molecular quantum mechanics which uses one-electron functions or orbitals to approximate the full wave function.

2.1.2 Basis Sets[33]

Obtaining quality *ab-initio* calculations depends mainly on the basis set and computational method being employed. In deciding upon which basis set to use, factors such as the molecule being studied and objective of the calculation must be primarily considered.[34] A basis set can be defined as a mathematical description of the orbitals within a system used to perform the theoretical calculation. In the equation 2.7 the wave function, Ψ, can be expanded in terms of a set of atomic orbitals, χ_μ, in the linear combination of atomic orbitals (LCAO) method, to give[35]

$$\Psi = \sum_{\mu} c_\mu \chi_\mu \qquad (2.7)$$

where, c_μ = molecular orbital expansion coefficient, and χ_μ = basis function of atomic orbital.

Using a variation of parameters to describe an individual orbital; results in the lowering of the energy. At some point, a situation can be reached when the energy is no longer decreased and the number of variational parameters is increased and then the best single determinant wave function is obtained. Once this occurs, altering the wave function, Ψ, by an insignificant amount will not alter the energy. The quality of the molecular orbital, Ψ, is determined by the number of atomic orbitals, χ_μ. The number of atomic integrals required is directly related to the

amount of electrons in a molecule regardless of the size of the molecules. Conclusively, a fast computer that has large storage capacity becomes essential.

The two types of atomic basis functions are Slater-type atomic orbitals (STOs) and Gaussian-type atomic orbitals (GTOs). The former is not well suited to numerical work, and their use in practical molecular orbital calculations has been limited. Almost all modern *ab-initio* calculations employ GTO basis sets. These basis sets, in which each orbital is made up of a number of Gaussian probability functions, has considerable advantages over STOs. The Gaussian series of programs deals, as the name implies, almost exclusively with Gaussian-type orbitals.[36] The basis sets, which are most commonly used in *ab-initio* molecular orbital calculations, are described below:

2.1.2.1 The 4-31G Basis Set[33]

The inner (1s) shell of the first row elements is represented by a single function, which is expanded in terms of four primitive Gaussian functions. The valence shell (2s, 2p) is split into an 'inner' or tightly held part, which is expanded in terms of three primitive Gaussian functions, and an 'outer' more diffuse part which is written in terms of one primitive Gaussian function. The hydrogen atom is represented by a split 1s shell, which comprises a three Gaussian inner and a one Gaussian function.

2.1.2.2 Polarisation Basis Sets[32]

Polarisation functions are added to the chosen sp-basis. These basis sets can be described as one that incorporates functions of higher angular quantum than are needed by the atom in its electronic ground-state. It provides for displacement of electronic charge away from the nuclear centres, i.e. charge polarisation.

On the other hand, functions with smaller exponents than the standard functions are called diffuse functions. Diffuse functions are useful in accurately describing anionic systems and describing the delocalisation of π-electrons. A diffuse function is introduced into a basis set by a +, *viz* 6-31+G* indicates diffuse functions on heavy atoms and 6-31++G* indicates diffuse functions on heavy atoms and hydrogens. The addition of diffuse functions seldom makes a tremendous difference in accuracy as shown by Bisetty et.al, in their investigation.[21]

(i) The 4-31G* Basis Set[35]

This basis set is similar to that described in **2.1.2.1** (above), but in addition it has a single set of Gaussian d-functions on each heavy atom, i.e., all atoms excluding hydrogen and helium. This allows for small displacements of the centre of electronic charge away from the nuclear position, thus improving the description of highly polar molecules or small strained ring systems.

(ii) The 6-31G* Basis Set[35]

In this basis set, the inner shell functions are expanded in terms of six, rather than four primitive Gaussian functions. A set of d-functions is included in the outer valence shells for the description of heavy atoms.

(iii) The 6-31G** Basis Set[35]

The basis set described above does not allow for any polarization of the s orbitals of either hydrogen or helium atoms. The 6-31G** basis set is similar to the one described above, except for the inclusion of a set of Gaussian type p-functions in the representation of each hydrogen or helium atom. There is no general rule for choosing an adequate basis set. The level of calculation depends on the desired accuracy and the molecular properties of interest.[37] If the geometry of the

molecule is influenced by polarisation effects, electron delocalisation or hyperconjugative effects, a 6-31G* or higher basis set is necessary to include the d orbitals. Using the hydrogen atom as an example, its electron cloud appears symmetrical around a lone hydrogen atom but when the hydrogen atom forms part of a molecule the electrons are attracted towards the other nuclei.[37]

This distortion leads to the mixing of p-type orbitals with the 1s of the hydrogen atom thus leading to an sp hybrid. Similarly, unoccupied d orbitals introduce the problem of asymmetry into p orbitals which can then be counteracted by introducing polarisation functions into the basis set. Polarisation functions have a higher angular quantum number and therefore correspond to p orbitals for hydrogen and d orbitals for the first and second row elements. Past reviews[38,39,40,41,42] have shown a great success rate for the use of medium sized basis sets at the HF level of calculation for describing the conformational features of amino acid residues. Thus as a result my calculations to plot the ab-initio calculated Ramachandran maps were carried out at the HF/6-31G* level. This will also enable direct comparison with the results obtained on the PCU amino acid (compound **3a**) which was performed with the same basis set.[21]

2.2 MOLECULAR MECHANICS[43]

Chemists routinely use this computational method either by itself or in combination with experimental or other computational techniques.[44] Molecular mechanics can be defined as a mathematical procedure designed to give accurate structural energies of molecules of varying structures. In contrast to quantum mechanical approaches, this method does not explicitly include electrons in the calculations. This can be attributed to the Born-Oppenheimer approximation, which states that the electronic and nuclear motions can be uncoupled from one another and considered separately.[32,35] Molecular mechanics makes the assumption that the electrons in a system find their maximum distribution from the

standpoint of the nuclear structure. This method is most applicable for understanding macromolecular structures and drug design because of the modest time[44] and inexpensive computations[33] required. However, with any theory or technique there are certain limitations that should be understood.

The atoms in molecules, for the molecular mechanics method, are treated as rubber balls of different sizes (atom types) joined by springs of varying length (bonds). In the process of a calculation the total energy is minimized with respect to atomic coordinates where:

$$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec} + \ldots \qquad 2.8$$

E_{tot} is the total energy of the molecule, E_{str} is the bond-stretching energy term, E_{bend} is the angle-bending energy term, E_{tors} is the torsional energy term, E_{vdw} is the van der Waals energy term, and E_{elec} is the electrostatic energy term. The various contributions are represented in the figure below.[43]

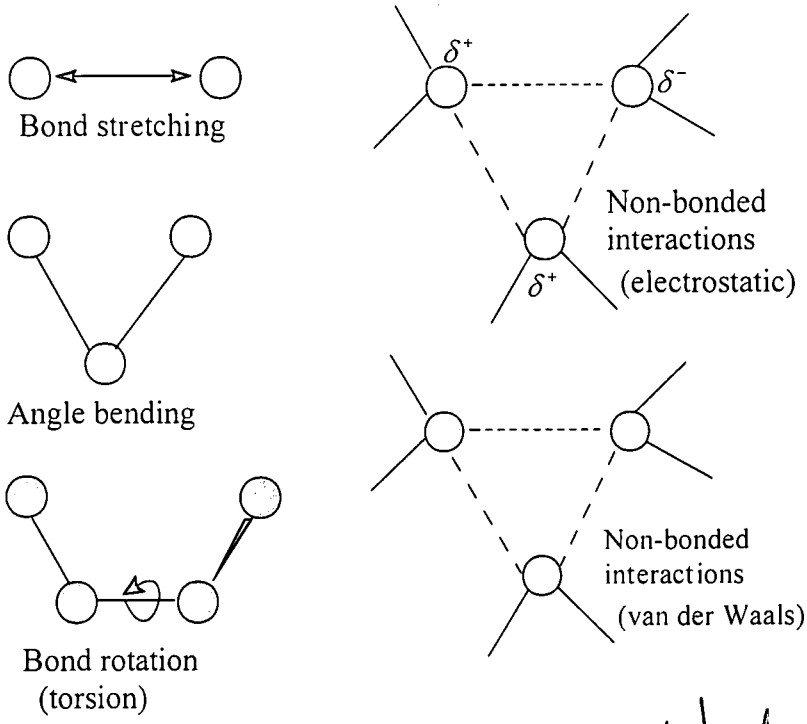


Figure not captioned numbered.

Since there are many different molecular mechanics methods and each one is characterized by its particular force-field, each force-field achieves only good results for those molecules for which it was parameterized. A force-field describes in detail the characteristics of an element within a specified context together within the environment it is found, thus different molecules of interest need different force-fields.

Since quantum mechanical methods are very time consuming, it is computationally expensive to be employed to large or flexible molecules. In contrast, molecular mechanics methods are far less time-consuming, and are indeed used in most molecular modelling programs to perform conformational searches. On the other hand, *ab-initio* methods are far more accurate than the molecular mechanics methods, and in the absence of experimental data, it is therefore necessary to use *ab-initio* methods as benchmark calculations.

CHAPTER 3

CONFORMATIONAL ANALYSIS

The conformations of a given molecular configuration are defined as a set of arrangements of their atoms in space, which can be interconverted by rotation about single bonds. Any changes in the conformation of a compound can have an influence on their physical and chemical behaviour. Most molecules of interest to organic, bio-organic, and medicinal chemists can have more than one conformation.[44] The essence of a conformational analysis is to identify the preferred conformations of a molecule. This requires the location of the conformations that are at minimum points on the potential energy surface, as shown in Figure 3.1. Energy minimization methods thus play a crucial role in conformational analysis. Conformational energies can be calculated using quantum mechanics or molecular mechanics methods. Molecular mechanical methods normally include a mathematical search algorithm to do a systematic or random analysis of the conformational possibilities.

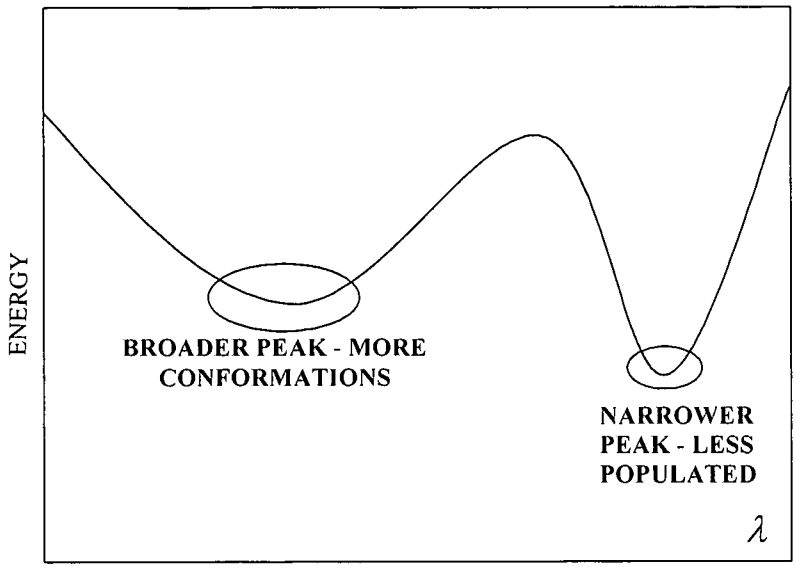


Fig. 3.1 An energy diagram showing potential energy (E) varying with some conformational parameter (λ)[44]

The identification of low-energy conformations is an important part of understanding the relationship between the structure and the biological activity of a molecule since the biological activity of a drug molecule is supposed to depend on the one single unique conformation among all the low-energy conformations.[45] Large energy of stabilization occurs due to the hydrogen bond interactions between the guest ligand and host receptor. Based on the information of the active conformation one may be able to design new agents for a particular receptor system especially when the bioactive conformation of the receptor site is known.[45]

3.1 SYSTEMATIC METHODS OF CONFORMATIONAL ANALYSIS[37,46]

This technique explores the conformational space of a molecule by making regular and predictable changes to their conformations. This involves the identification of all rotatable bonds on the backbone of a protein, which consists of a repeated sequence of three atoms, belonging to one amino acid residue, generally represented

as N_i, C_i^{α}, and $C_i^{'}$, where i indicates the number of the residue starting from the N-terminus of the chain, as shown in Figure 3.2 below.

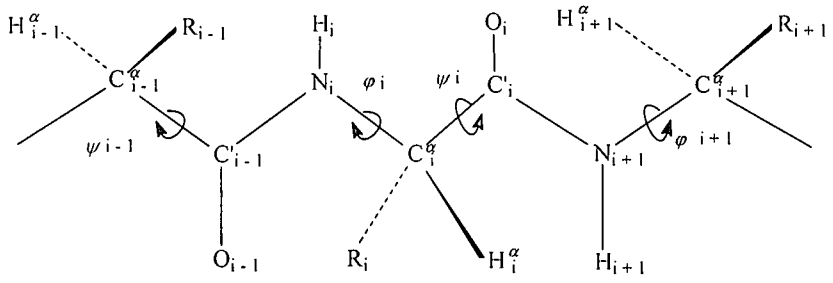


Fig. 3.2 Polypeptide sequence showing all dihedral angles

In proteins the backbone torsion angles are named ϕ (phi), ψ (psi) and ω (omega). The rotation about the N-C_α bond is described by ϕ, rotation about the C_α-C' bond is described by ψ, and the rotation about the peptide bond C'- N is described by ω.

Note that the amide bonds (Ni – C'i-1 and C'i – Ni+1) are planar with the amide proton in a *trans* position with respect to the carbonyl oxygen and is in principle subject to restricted rotation. This is known from X-ray structures and is a direct result of delocalization of the lone pair on nitrogen as well as the π-electrons of the carbonyl bond. The *trans* configuration is energetically more favoured than the *cis* configuration (ω=0°). The other bonds are systematically rotated through 360 degrees in fixed increments, while the bond angles (ψi-1, ϕi, ψi, ϕi+1) remain fixed throughout the calculations. The rest of the molecule is free to find the lowest energy subject to the constraints imposed with respect to the four dihedral angles.

This is done manually for the *ab-initio* (Gaussian) conformational search. The molecular mechanical software has an automated algorithm to create the starting conformations which are subjected to energy minimisation in order to derive the

Bad paragraph structure

associated minimum energy conformation. The search is completed once all possible conformations of torsion angles have been generated and minimised.

The permitted values of ϕ and ψ were first analysed and determined by Ramachandran et.al,[46] and their results gave a systematic method to determine the stable conformations of small peptides. Each conformation represented by a particular (ϕ, ψ) combination was examined for close contacts between atoms, only values of ϕ and ψ for which there are no close contacts between atoms are permitted, and usually they are presented on a 2D contour diagram depicted as the Ramachandran plot shown in Figure 3.3.

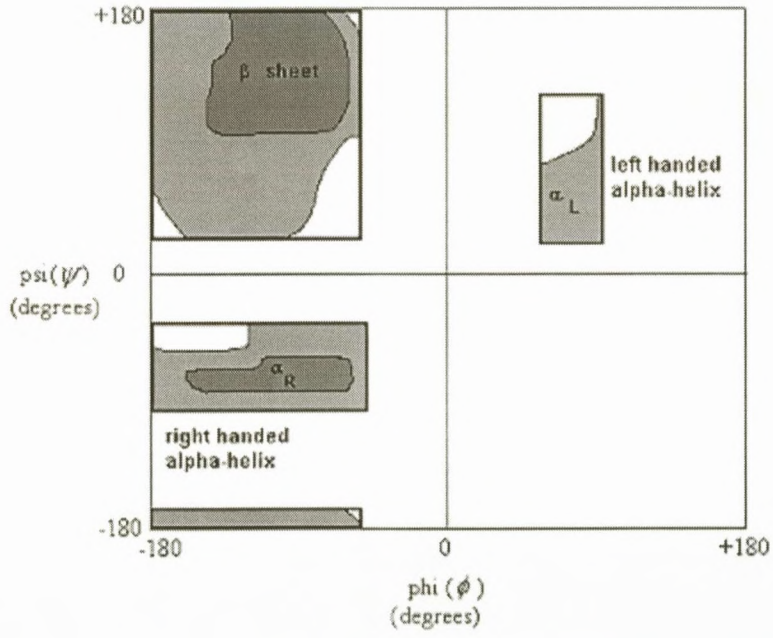


Fig. 3.3 Simplified diagram for the Ramachandran plot[47]

Due to the various structures characterized by the 2D Ramachandran torsion angles ϕ and ψ (Figure 3.2), Ramachandran maps are an important and relative easy-to-analyse test for the nature of a 3D protein structure. Accessible areas on the contour maps calculated by Ramachandran correspond to those conformations that

are observed in X-ray structures of proteins. In Figure 3.3 the areas within the darkly shaded regions are considerably favoured since it is this region where conformations exhibit no steric clashes. The area directly outside (lightly shaded region) corresponds to those conformations whereby slight alterations of bond angles are accepted. All the regions that remain clear include those conformational structures where the atoms in the peptide residue are located in distances closer than the sum of their van der Waals radii.

With the exception of glycine, these clear areas are sterically disallowed for all amino acids. The darkly shaded region represents the sub-regions of the phi, psi space. They are classified as α_L, β-sheet, α_R) the secondary structures that result when corresponding phi, psi angles occur on a repetitive basis.

3.1.1 The α-helix and the β-sheet[37]

The right-handed alpha helix (α_R in Figure 3.3) occupies the area in the lower left near -60° and -40°, while the β-sheet resides in the upper left around -120° and 140°. The slightly unfavoured α_L helix conformations tend to be found in the upper right hand coordinates of 60° and 40°. The α_R helix is the most easily recognised secondary structural element in proteins. The right handed alpha-helical conformer is mainly formed and stabilised by repeated hydrogen bonds between the carbonyl function of one residue and the NH of the neighbouring residue. On the other hand it has been reported[37] that left handed alpha-helix are found only for individual residues and not for a series of neighbouring residues. The second most common and identifiable secondary structural motif is the β-sheet. Beta sheets form from a collection of beta strands. These beta strands develop from linear extended conformations of the backbone of polypeptides. β-sheets are energetically less favoured due to the intramolecular bonding of the beta strands while the alpha helix conformation shows preference due to the much more stable intermolecular bonding.

The following two Figures illustrate the formation of secondary structures when hydrogen bonding between the amino acids causes the polypeptide to form an alpha helix or a pleated beta sheet.

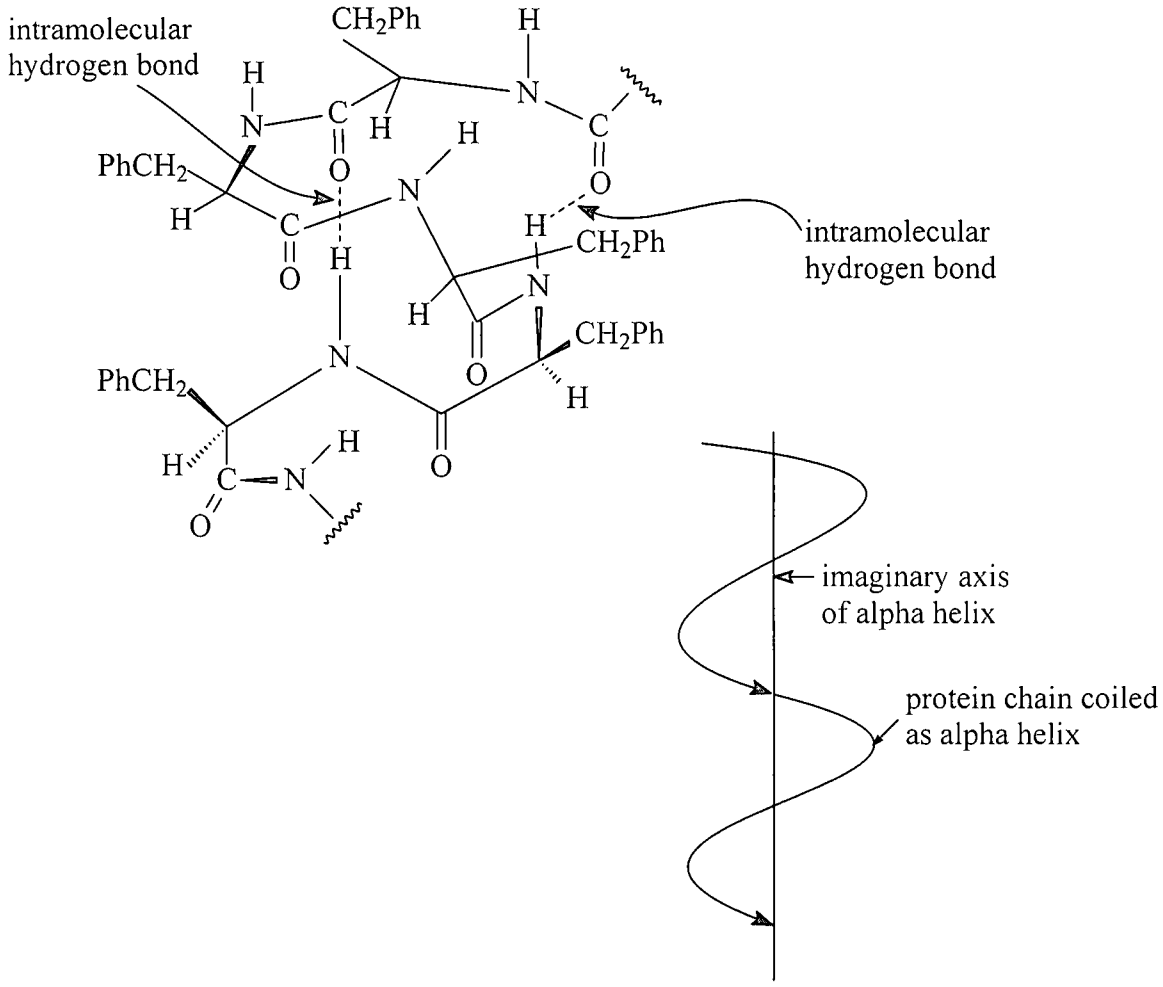


Fig. 3.4 Partial structure of a phenylalanine protein residue α-helix[6,27]

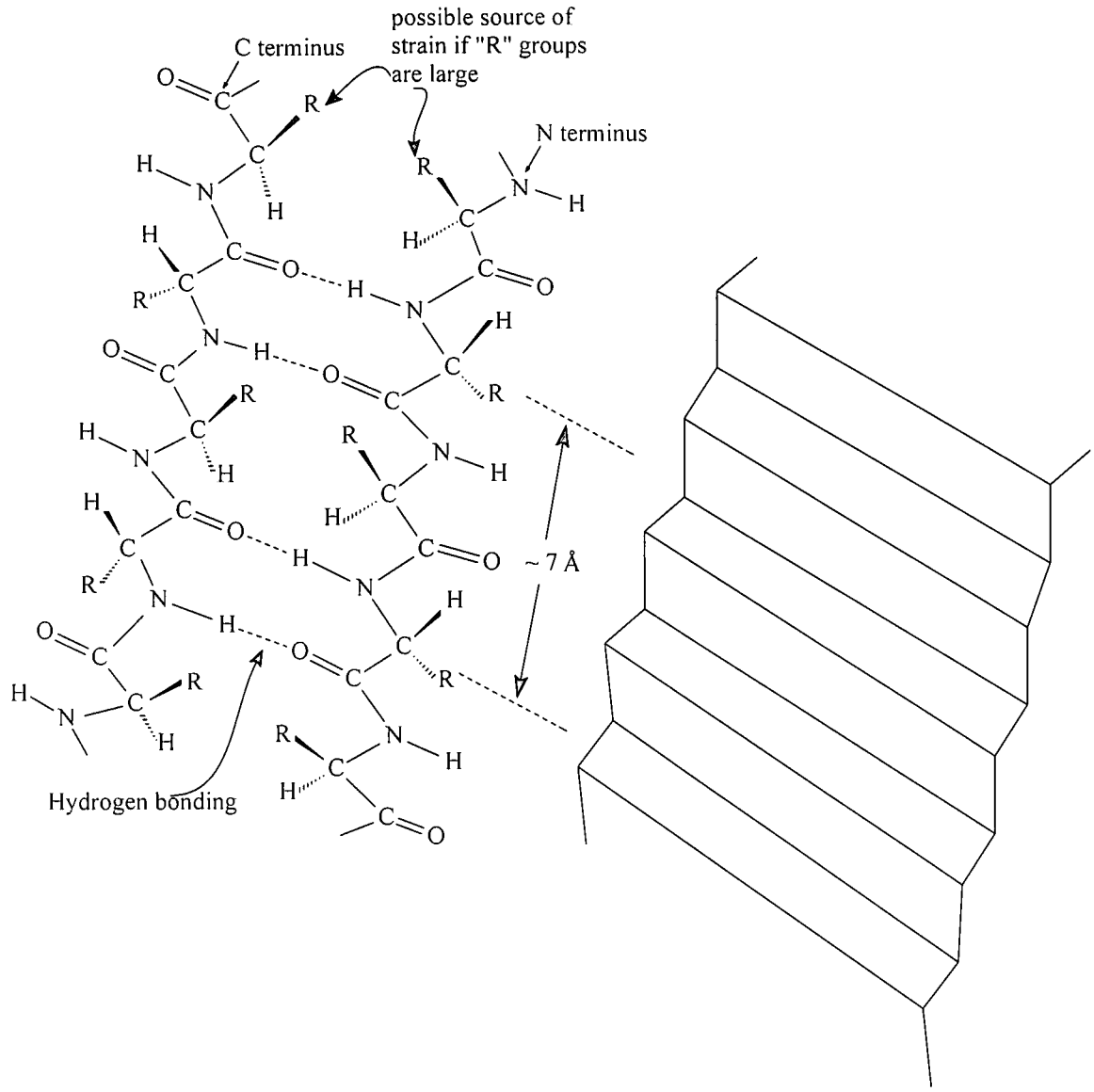


Fig. 3.5 Partial structure of a β-pleated sheet[48]

3.2. RANDOM METHOD OF CONFORMATIONAL ANALYSIS[37,43]

A random search is in many ways the complete opposite of a systematic search. In a random search it is not possible to predict the order in which conformations will be generated, whereas a systematic search explores the energy surface of the molecule in a predefined and logic fashion. A random search can move from one

region of the energy surface to a completely unconnected region in a single step and is able to explore conformational space by randomly changing either the atomic cartesian coordinates or the torsion angles about the rotatable bonds. Another important difference is that the random search does not keep the initial constraints for the remainder of the optimization. The result is that the overall structure gets trapped in the local minimum nearest to the initial starting configuration. Since the procedure theoretically should explore all possible starting conformations/dihedral angles, the global minimum should be reached at some stage. In practice a reasonable picture of the conformational preferences of the peptide is achieved much quicker. A plot of the dihedral angles obtained after minimization on a typical Ramachandran plot will in most cases exhibit three to four regions quite quickly which then become more and more populated with time.

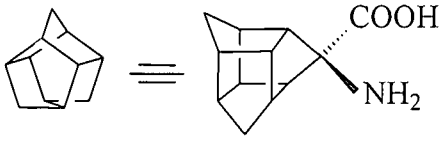
A systematic search reaches finality when all possible combinations of bond rotations have been considered however, in a random search, there is no natural endpoint. In other words, the endpoint is reached asymptotically. This leads to one never being absolutely sure that all of the minimum energy conformations have been found. Thus the strategy employed is to generate conformations until a statistical analysis[43] show that the chances to find new structures are very slim. This usually requires each structure to be generated many times, and inevitably leads to each region of the conformational space to be more extensively explored as explained above.

This project was conducted using the systematic method approach since one can be assured that the entire conformational space has been explored to the maximum in the shortest time.

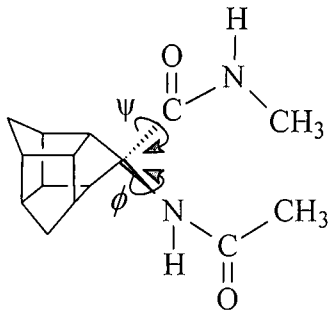
CHAPTER 4

COMPUTATIONAL DETAILS AND METHODOLOGY

The main thrust of this research project was to perform a complete exploration of the conformational profile of the Trishomocubane-amino acid cage peptide. The whole Ramachandran plot of Ac-Trish-NHMe dipeptide, compound **4b**, was investigated at the Hartree-Fock level using the 6-31G* basis set, and the low energy conformers were identified.



4a



4b

The results were subsequently used as a reference to generate molecular mechanics parameters for the Trishomocubane-amino acid cage RESIDUE, compatible with the parm94 set of parameters of the AMBER force-field. The force-field parameters are included in the appendix, after chapter 5.

All computational chemistry programs used in this study were implemented on a Silicon Graphics O2 UNIX workstation. The high-level *ab-initio* calculations were performed on a CRAY-YMP supercomputer at the Centre de Supercomputació, Catalunya (CESCA), Barcelona, Spain.

4.1 THE GAUSSIAN 94 REV. E 4.1 COMPUTER PROGRAM[36]

The GAUSSIAN series of programs deal mainly with Gaussian-type orbitals, as described in Chapter 2. They are capable of performing *ab-initio* Hartree-Fock (HF) molecular orbital calculations based on the linear combination of atomic orbitals (LCAO) approach. The GAUSSIAN programs are designed to perform a variety of molecular mechanical, semi-empirical and *ab-initio* molecular orbital (MO) calculations. GAUSSIAN computations can be carried out on systems in the gas phase or makes use of methods to approximate solution effects. It can also calculate structures and energies in their ground states or in an excited state. In this study the GAUSSIAN 94 Rev. E 4.1 program was also used in the characterization of parameters for the tris-amino acid cage peptide. This involved the geometry optimization and a subsequent calculation of molecular electrostatic charges from quantum mechanical wave functions,[49] using the 6-31G* basis set. Literature studies revealed that *ab-initio* calculations at the HF level using medium size basis sets have proven in the past to be very useful in describing the conformational features of amino acid residues.[40,41] Thus calculations to map the Ramachandran plot of the Ac-Trish-NHMe cage peptide were carried out at the HF/6-31G* level[38,42] in a similar way as described in §3.1 above.

In order to simplify the computational model, this study involved only gas phase calculations. The GAUSSIAN programs can also be used for various types of investigations such as exploring areas of chemical interest like substituent effects, reaction mechanisms, potential energy surfaces, and excitation energies.[33] Representing the compound under investigation in the form of cartesian coordinates or the Z-matrix notation are two acceptable methods for GAUSSIAN input data. This input data consists of the molecular charge and multiplicity followed by the symbols of the constituent atoms and a definition of the molecular structure. This information further defines the molecular geometry in term of bond lengths, bond angles and torsion angles. Further specification must also be given to whether a single-point calculation, geometry optimization or frequency calculation must be

carried out. In addition, the appropriate basis set and the level of theory, either the Hartree-Fock (HF) or the second order Møller-Plesset (MP2) level of theory must also be specified.[50,51]

4.2 THE AMBER 5.0 COMPUTER PROGRAM[24]

AMBER is a collective name for a suite of programs that allow users to carry out molecular mechanics calculations, specifically designed for peptides and proteins. The term *amber* is also used to describe the force-field parameters contained in the database, implemented in the AMBER 5.0 computer program.

For the purposes of this study, new parameters for the Trishomocubane-amino acid cage residue were developed and subsequently incorporated into the AMBER program. This was accomplished by the PREP, LINK, EDIT and PARM modules which are part of the preparatory programs within the AMBER program

4.2.1 Preparatory programs in AMBER

A brief description of each of the modules is provided below and a sample input file used in each module is presented at the end of this chapter (see Appendix).

PREP

This module is used mainly for residues not already defined in the standard AMBER database. Using the topology/parameter information the PREP file is able to create or add to a residue database. The standard AMBER database contains relevant information of molecules such as the coordinates, atom types, and atom charges for the 20 naturally occurring amino acids.[24]

LINK

The LINK file deals only with the topology information. The user informs LINK of the residue sequence of the molecule, and LINK extracts the topology information for each residue from the standard AMBER database. The topology for each residue is linked together to form the topology of the system. This is written to a binary file (lnkbin being the default name), which is read by EDIT.[24]

EDIT

One of the primary purposes of EDIT is to read PDB coordinates and apply them to the system generated by LINK. If there are any coordinates for atoms that are missing from the PDB file, EDIT will automatically generate these coordinates by using the stored internal coordinates in the link binary file, lnkbin. The EDIT file is advantageous in that it can change a molecules environmental conditions by solvating it in water.[24]

PARM

Credit can be giving to PARM when determining which bonds, angles, torsion angles, and atom types exist in a system. The appropriate parameters are extracted from the force field file. PARM then writes the final coordinates and topology files needed by all other AMBER programs. This results in two output files being generated, viz., topology file, (default name prmtop) and a coordinate file (default name prmcrd).[24]

4.2.2 Energy programs in AMBER[24]

The suite of modules contained within the AMBER program, is often referred to as the energy programs, viz, GIBBS, SANDER and SPASMS. They are all very unique modules, and one must know what information is needed when working with the energy programs. A summary of information needed by all the energy programs is given below:

Cartesian coordinates for each atom in the system is required. Model-building programs or some analytical techniques such as X-ray crystallography and NMR spectroscopy usually provide this information. For the purposes of this investigation, the structures were drawn using Gauss View.

Data about the connectivity, atom names, atom types, residue names, and charges are also required. This "topology" information comes from the AMBER 5.0 database, and is often referred to as db94.dat. This db94.dat file contains data for the standard amino acids as well as N and C-terminal charged amino acids, DNA and RNA. The database contains default internal coordinates for these monomer units. However, coordinate information is usually obtained from the protein database (PDB) files. In the case of the PCU cage residue, topology information was not found in the standard database, and therefore a new database was generated.

The database (db94.dat) directory includes all parameters for bonds, angles, torsions and atom types in the system. This is often referred to as the parm94 force field.[52]

Providing the desired procedures and parameters to the energy programs. This information is found in three separate files where the first contains the coordinates, the second contains the topology and parameters, and is called the "topology file" and the third contains the command or input file.

For the purposes of this study, only the SANDER module was used and a brief description of this module is given below.

SANDER[24]

The acronym stands for **S**imulated **A**nnealing with **N**MR-**D**erived **E**nergy **R**estraints. SANDER carries out energy minimization, molecular dynamics, and NMR refinements. However, this module is used for a variety of simulations that have nothing to do with NMR refinements. This program causes the structure to relax by iteratively moving the atoms down the energy gradient until a sufficiently low average gradient is obtained. SANDER provides standard protocols for energy minimization and molecular dynamics, and is used for everything except Gibbs free energy calculations. Structures are generally minimized before a molecular dynamics simulation can take place. The molecular dynamics portion generates configurations of the system by integrating the Newtonian equations of motion as described in Chapter 2.

The minimized structure was subsequently used as starting geometry, after changing the corresponding backbone torsion angles on a grid of points on the (ϕ,ψ) space at 15° intervals, ranging from −180 degrees to +180 degrees.[1,15] At each point of the grid, the geometry was optimized by keeping the backbone torsion angles, ϕ and ψ, constrained during the minimization process, while allowing all other variables to optimize.

These data were used to compute the AMBER Ramachandran maps, *in vacuo*, depicted in Chapter 5. After the Ramachandran maps were computed, the geometries of the different minima were identified. Local minima were identified by full optimization of the geometry (i.e., no constraints were imposed in the optimization).

Finally, the quality of the Ramachandran maps were assessed, by selecting four cross-sections on the potential energy surface of the Ramachandran maps computed at the *ab initio* level. For this purpose, the four sections on the Ramachandran map corresponding to ψ=75°, ψ=-45°, ϕ=60°, ϕ=-90° were chosen and are graphically presented in Chapter 5. These sections were chosen because they correspond to cross-sections cutting through the energy minima on the potential energy surface. Furthermore, in order to establish comparisons between the *ab initio* results with those of the AMBER computed results, energy calculations were performed using the AMBER force field, as well as at the HF/6-31G* level for each cross section. A pictorial representation of these results is presented in Chapter 5.

CHAPTER 5

RESULTS AND DISCUSSION

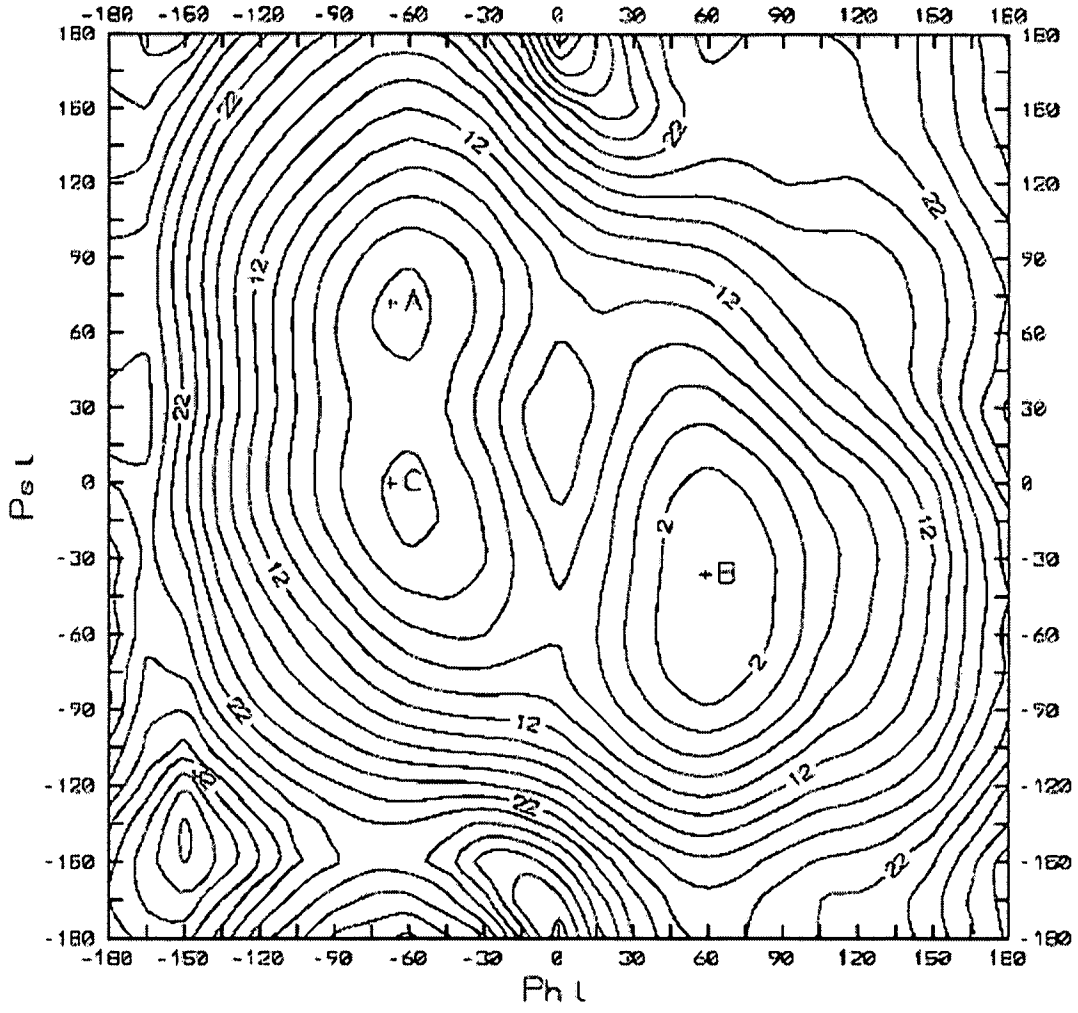
In order to explore the conformational space of the Ac-Trish-NHMe dipeptide, new atomic charges (parameters) were computed as mentioned in Chapters 1 and 4 respectively. The best way to test the new parameters was to compare both Ramachandran plots for the AMBER and *ab-initio* methods of calculation.

5.1 AC-TRISH -NHME

The Ramachandran plots of the Ac-Trish-NHMe dipeptide were computed at the molecular mechanics and *ab initio* levels by varying the torsional angles at 15° intervals. The Ramachandran plots computed using the AMBER force fields, and at the HF/6-31G* level are shown in Figures 5.1 and 5.2.

5.2 CHARACTERIZATION OF LOW ENERGY CONFORMERS OF AC-TRISH –NHME USING RAMACHANDRAN MAPS

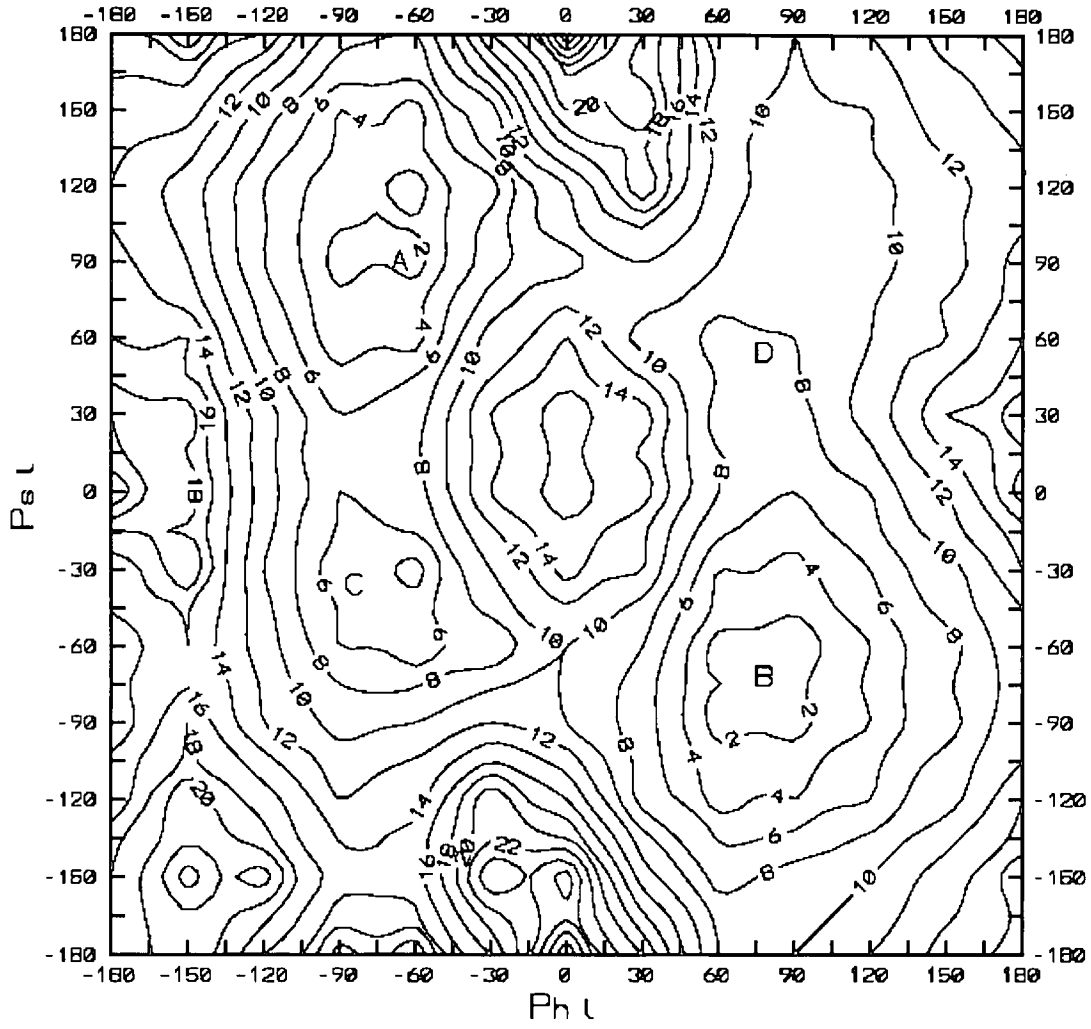
Figures 5.1 and 5.2 represent the Ramachandran maps generated from the amber and ab initio methods of computation. Important information such as the low energy regions for specific conformations and unique structure for each conformer can be identified.



Energies are expressed in kcal mol^{-1} relative to the lowest energy minimum, and the contours are plotted every 2 kcal mol^{-1}.

A represents the low energy C_{7eq} conformation;
B represents the low energy C_{7ax} conformations;
C represents the low energy 3_{10} conformation

Fig. 5.1 Ramachandran maps for Ac-Trish-NHMe using the parm94 force field

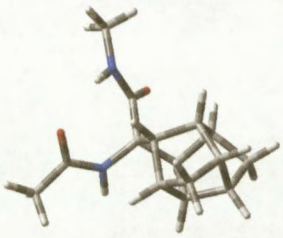


Energies are expressed in kcal mol^{-1} relative to the lowest energy minimum, and the contours are plotted every 2 kcal mol^{-1}.

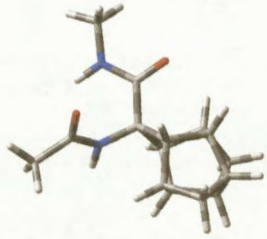
A represents the low energy C_{7eq} conformation;
B represents the low energy C_{7ax} conformations;
C represents the low energy 3_{10} conformation; and
D represents the low energy α_L conformation.

Fig. 5.2 Ramachandran maps for Ac-Trish-NHMe at the HF/6 31G* level

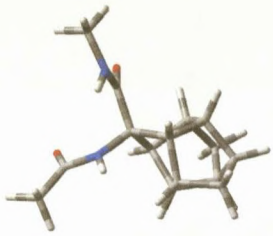
In the AMBER computed Ramachandran map, Figure 5.1, three unique minima, viz., C_{7eq}, 3_{10} and the C_{7ax} region (labelled A, B, and C) were observed. However the result of the Ramachandran plot obtained at the ab initio level (HF/6-31G*), reveals an additional minima, α_L. These conformers are presented in Figure 5.3 below.



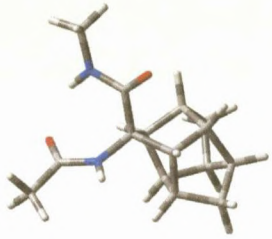
C_{7eq} conformer



3_{10} conformer



α_L conformer



C_{7ax} conformer

Fig. 5.3 Minima conformers obtained from the Ramachandran plot calculated at the *ab initio* HF/6-31G* basis set[#]

Comparison of the AMBER with the *ab initio* computed results reveals slight differences in energy between the C_7 and 3_{10} conformers. A summary of the relative energies computed at the different theoretical levels is presented in Table 5.1.

[#] The Cartesian coordinates for all structures presented throughout are available on the CD at back of book.

Table 5.1 Calculated relative energies for Ac-Trish-NHMe

	Relative Energies/kcal mol^{-1}			
	C_{7eq} **(A)**	C_{7ax} **(B)**	3_{10} **(C)**	α_L **(D)**
Parm 94	0.0	0.0	1.5	N/A
HF/6-31G*	0.1	0.0	2.3	8.0

These results indicate that the AMBER force field results were able to reproduce the local minima found at the *ab initio* level for the C_{7ax} and C_{7eq} conformations. It can also be seen that the order of the relative energies predicted using the AMBER force field is $C_{7ax} = C_{7eq} < 3_{10}$ whereas at the *ab initio* level it is: $C_{7ax} < C_{7eq} < 3_{10} < \alpha_L$. It should be noted that an energy difference of 0.1 kcal mol^{-1} is close to the error of the both programs and in reality almost negligible. The energy profile of butane is a good example.[53] The difference between the staggered (lowest) and the anti conformation (highest) is about 4.5 kcal mol^{-1}, while it is experimentally known that free rotation is maintained at room temperature. Molecules have enough thermal and kinetic energy at room temperature to allow processes requiring between 15 to 20 kcal/mol.[54]

Energy profiles of Ramachandran cross sections provide greater insight into the differences between *ab initio* and molecular mechanics calculations. Four cross sections on both Figure 5.1 and Figure 5.2 were then performed which is duly represented in Figures 5.8 to 5.11 below.

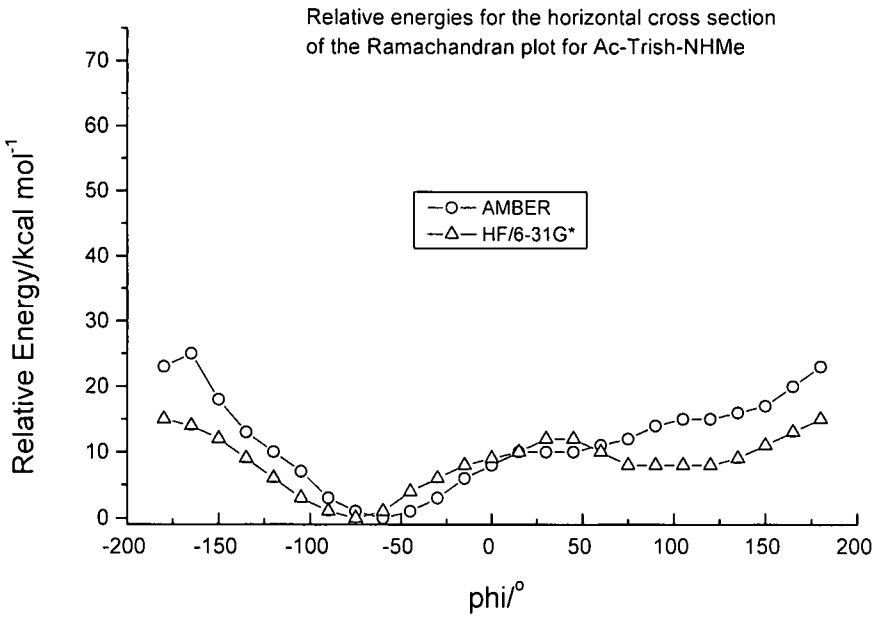


Fig. 5.4 Relative energies of the cross-sections of the Ramachandran map at $\psi = 90°$

Figure 5.4 represents a horizontal cross-section of the Ramachandran maps, computed at the parm94 level and at the HF/6-31G* level. The backbone torsion angles were fixed at $\psi = 90°$ and the corresponding $\phi = \pm 180°$ at 15° intervals. The significance of this graph is that at both computational levels, the energy minima correspond to $\phi = -60°$, $\psi = 75°$ and $\phi = -75°$, $\psi = 90°$ for the amber and ab initio plots respectively. This corresponds to the letter **A** in Figures 5.1 and 5.2 representing the C_{7eq} conformation.

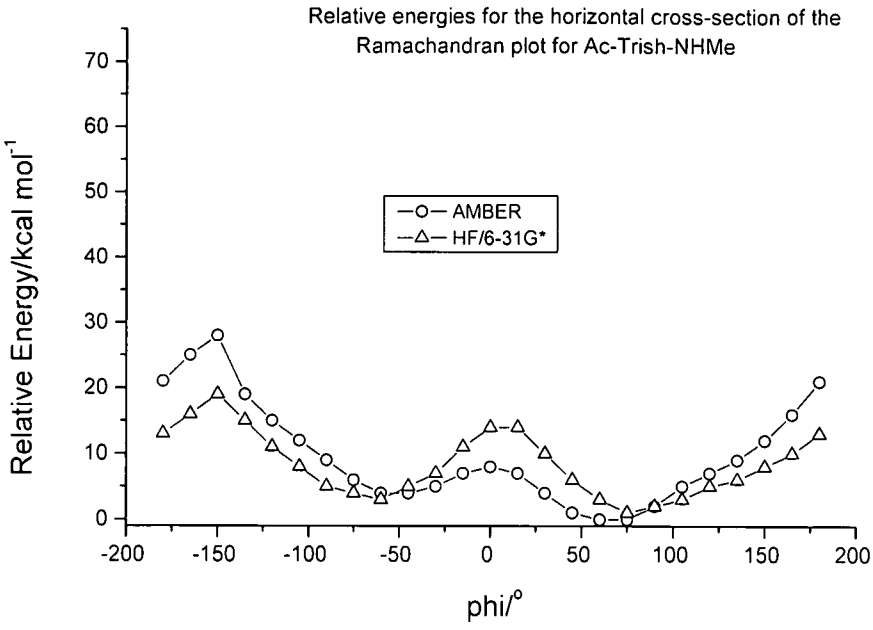


Fig. 5.5 Relative energies of the cross-sections of the Ramachandran map at $\psi = -45^{o}$

Figure 5.5 represents the second horizontal cross-sections of the Ramachandran maps, computed at the parm94 level and at the HF/6-31G* levels. In this case psi was kept constant at $\psi = -45°$ and the corresponding $\phi = \pm 180°$ at 15° intervals. The significance of this graph is that at both computational levels, the energy minima correspond to $\phi = 75°$ and $\psi = -65°$ for both the amber and ab initio plots, which corresponds to the letter **B** in Figures 5.1 and 5.2, representing the C_{7ax} conformation.

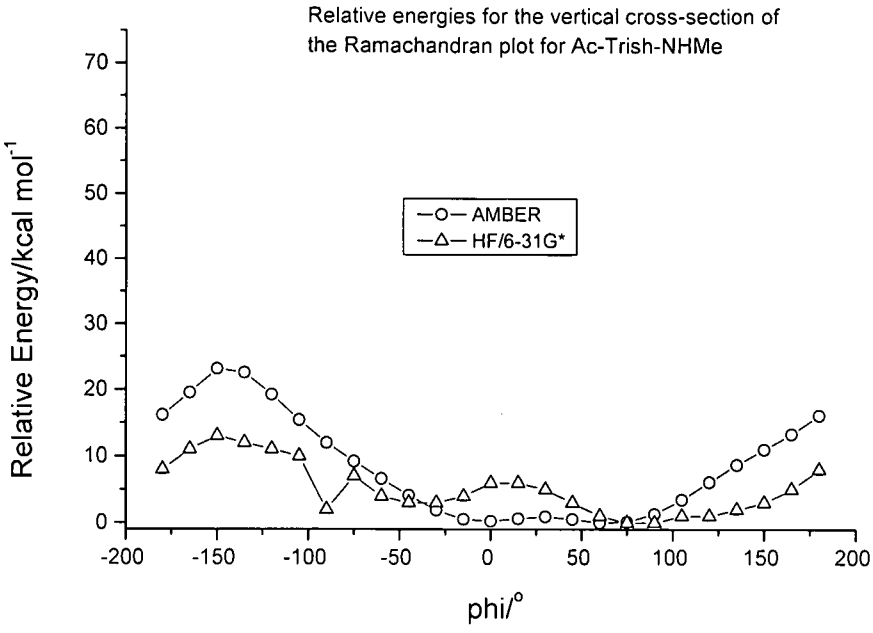


Fig. 5.6 Relative energies of the cross-sections of the Ramachandran map at ϕ = -75°

Figure 5.6 was obtained from a vertical cross-section of the Ramachandran maps, computed at the parm94 level and at the HF/6-31G* level. Again, geometry optimizations were performed at both levels at ϕ = -75° and the corresponding ψ = ±180° at 15° intervals. The significance of this graph is that at both computational levels, the energy minimum corresponding to ϕ = -60° and ψ = -15° for the ab initio and amber plots, which in turn corresponds to the region representing the 3_{10} conformation labelled **C** in Figures 5.1 and 5.2.

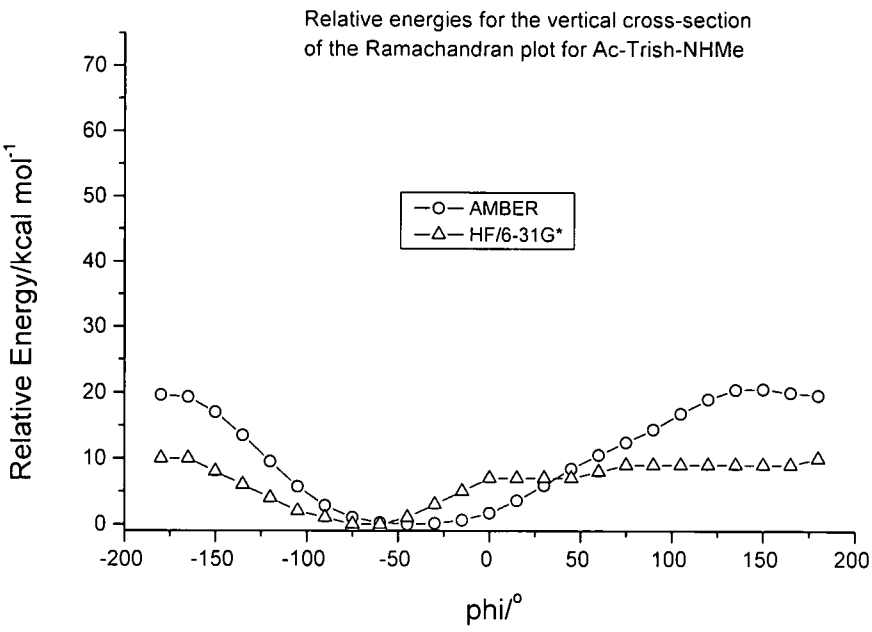


Fig. 5.7 Relative energies on the cross-sections of the Ramachandran map at $\phi = 75°$

Figure 5.7 represents the second vertical cross-section of the Ramachandran map computed at the parm94 level and at the HF/6-31G* level. Geometry optimizations were performed at both levels, at $\phi = 75°$ and the corresponding $\psi = \pm 180°$ at 15° intervals. The significance of this graph is that at both computational levels, only one energy minima was found corresponding to $\phi = 75°$ and $\psi = 60°$. This energy minima corresponds to the region representing the α_L conformation labelled **D** in Figure 5.2.

Performing geometric optimizations through the amber and *ab initio* methods show a somewhat reasonable trend of the location of the minima on the potential energy surfaces.

A summary of the conformational characteristics, along with the relative energies for these conformers using the molecular mechanics approach, incorporating the

parm94 force field along with those obtained at the various *ab initio* levels, is listed in Tables 5.2 to 5.5 respectively. The *ab initio* calculations were performed on the four low energy conformers only, and the main purpose is to assess the validity of the geometrical parameters as a function of the theoretical level.

Table 5.2 Calculated hydrogen bond parameters for Ac-Trish-NHMe

	C_{7eq} **(A)**		C_{7ax} **(B)**		3_{10} **(C)**		α_L **(D)**	
	d/Å	A/^{0}	d/Å	A/^{0}	d/Å	A/^{0}	d/Å	A/^{0}
Parm 94	1.93	140.4	1.88	154.8	2.48	109.7	N/A	N/A
HF/6-31G*	2.18	138.0	2.03	145.1	2.63	105.5	3.96	81.5

Key: A, B, C and D as numbered on the Ramachandran maps Figures 5.1 and 5.2

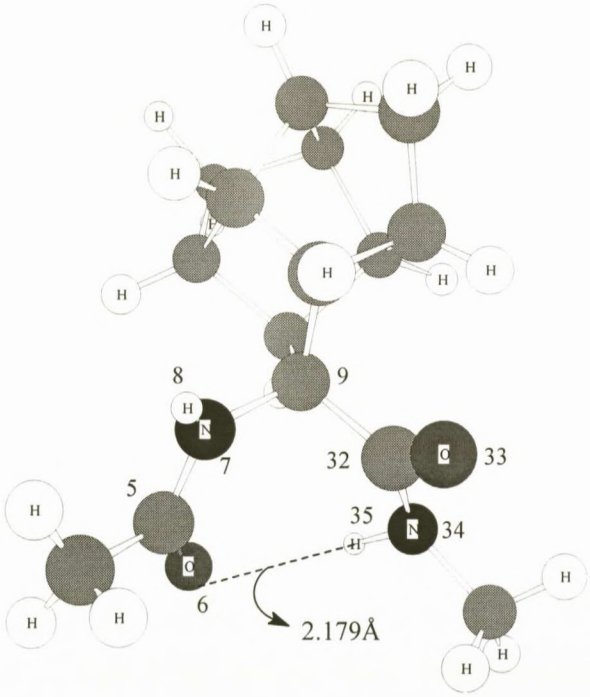


Fig.5.8 C_{7eq} conformer

Differences in the hydrogen-bonded characteristics for structures optimized using the parm94 force field along with those obtained at the HF level using the 6-31G* basis set are presented in Table 5.2. Figure 5.8 is used as a reference to record data

such as the hydrogen bond lengths between O6-H35 and the angles between NH-O. In general, as can be seen by looking at the results, the intramolecular hydrogen bond parameters at these theoretical levels are in agreement with each other. These results are similar to those computed by Bisetty et.al for the naturally occurring amino acids.

Table 5.3 Calculated torsional angles for Ac-Trish-NHMe

	Torsional Angles/^{0}							
	C_{7eq}		C_{7ax}		3_{10}		α_L	
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
Parm 94	-60.2	74.2	72.0	-59.1	-59.4	-14.1	N/A	N/A
HF/6-31G*	-60.0	90.0	75.0	-60.0	-60.0	-15.0	75.0	60.0

Table 5.3 shows that the differences of ϕ and ψ between the AMBER and the HF/6-31G* level falls between the range 2^o to 15^o.

Using the three-dimensional structure, Figure 5.8, of the Ac-Trish-NHMe dipeptide, information such as bond lengths and bond angles were obtained. These results are recorded in Tables 5.4 and 5.5.

The following bond lengths and angles were obtained for the Ac-Trish-NHMe peptide by following the number system adopted in Figure 5.8 above.

Table 5.4 Calculated bond lengths in Å for Ac-Trish-NHMe

	O6-C5	**N7-C5**	**H8-N7**	**C9-N7**	**C32-C9**	**O33-C32**	**N34-C32**	**H35-N34**
				$C_{7\,eq}$				
HF/6-31G*	1.200	1.350	0.990	1.460	1.540	1.200	1.340	0.990
				$C_{7\,ax}$				
HF/6-31G*	1.200	1.340	0.990	1.460	1.540	1.200	1.340	0.990
				3_{10}				
HF/6-31G*	1.200	1.350	0.990	1.460	1.540	1.200	1.340	0.980
				α_L				
HF/6-31G*	1.196	1.360	0.990	1.460	1.540	1.190	1.360	0.990

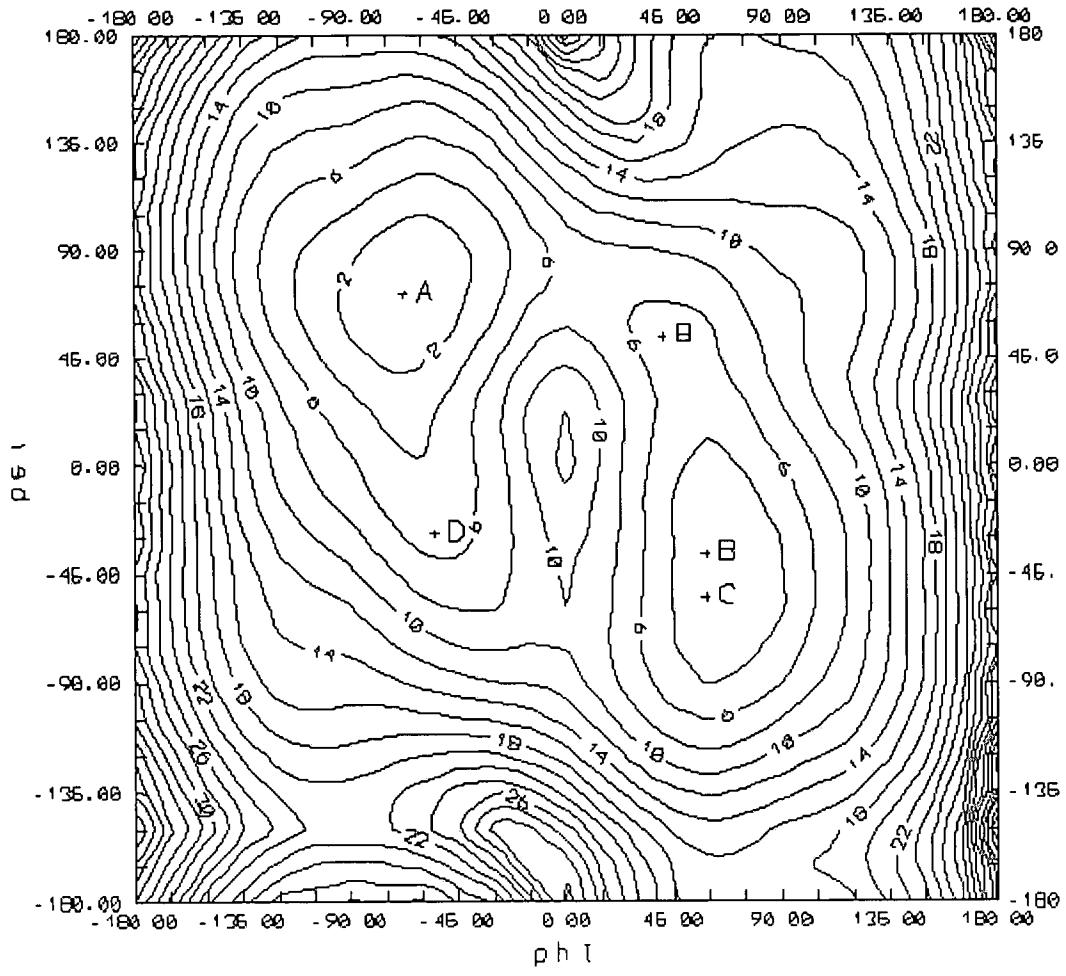
Table 5.5 Calculated bond angles in degrees for Ac-Trish-NHMe

	O6C5N7	C5N7H8	N7C9C32	C9C32C34
			$C_{7\,eq}$	
HF/6-31G*	123.19	115.81	109.14	116.90
			$C_{7\,ax}$	
HF/6-31G*	123.79	116.58	110.89	116.36
			3_{10}	
HF/6-31G*	122.87	117.67	113.24	118.39
			α_L	
HF/6-31G*	123.11	116.62	107.87	114.30

A close inspection of these results shows a good comparison with the literature data[6], Chapter 1- Figure **1,** given for the typical bond lengths and angles for naturally occurring peptides.

5.3 AC-PCU-NHME[21]

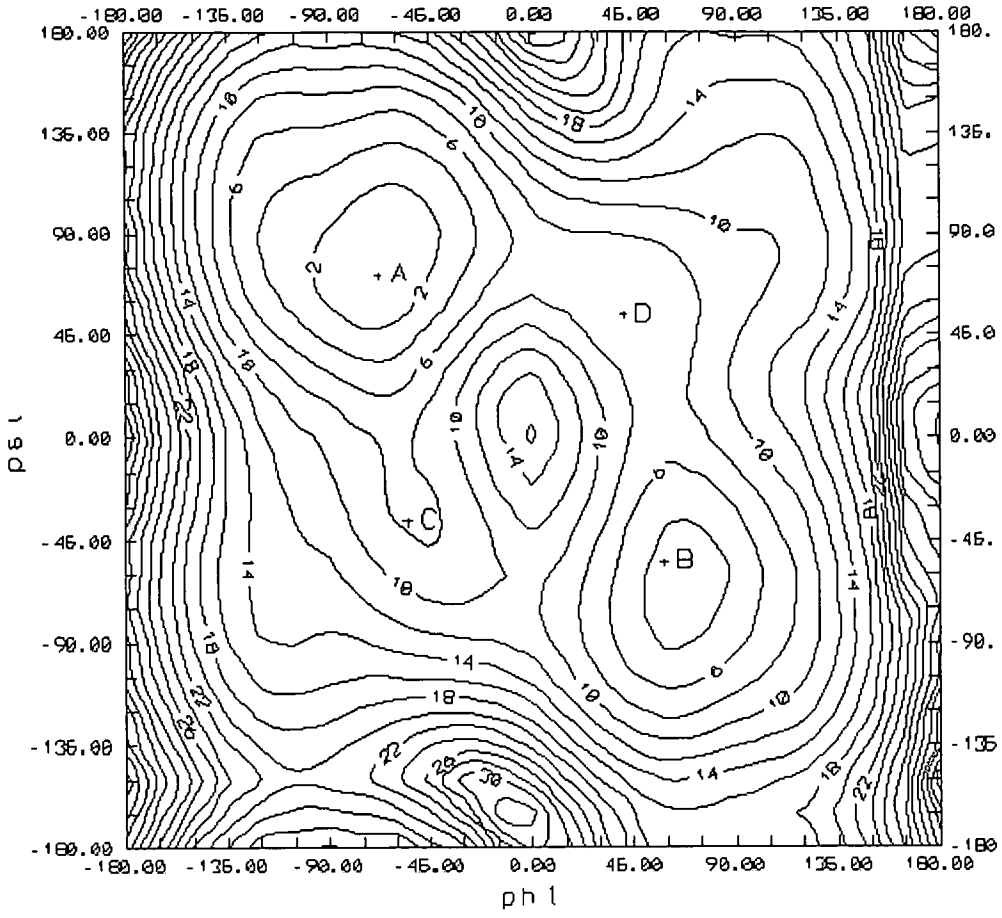
The following results presented below were obtained from Bisetty et.al's[21] investigation of the dipeptide Ac-PCU-NHMe. The Ramachandran plots for the Ac-PCU-NHMe amino peptide was also computed using the AMBER force-fields, and at the HF/6-31G* level (see figures 5.9 and 5.10). These maps and all other necessary information required to perform a comparison between the two dipeptides are presented below.



Energies are expressed in kcal mol^{-1} relative to the lowest energy minimum, and the contours are plotted every 2 kcal mol^{-1}.

A represents the low energy C_{7eq} conformation;
B and C represents the low energy C_{7ax} conformations;
D represents the low energy 3_{10} conformation; and
E represents the low energy α_L conformation.

Fig. 5.9 Ramachandran maps for Ac-PCU-NHMe using the parm94 force field



Energies are expressed in kcal mol^{-1} relative to the lowest energy minimum, and the contours are plotted every 2 kcal mol^{-1}. The letters A to D denote the locations of the low energy minima, where

A represents the low energy C_{7eq} conformation;
B represents the low energy C_{7ax} conformations;
C represents the low energy 3_{10} conformation; and
D represents the low energy α_L conformation.

Fig. 5.10 Ramachandran maps for Ac-PCU-NHMe at the HF/6 31G* level

The Tables below represent results obtained from the effects of changes in conformation on the local geometry of the peptide, Ac-PCU-NHMe.

Table 5.6 Calculated relative energies for Ac-PCU-NHMe

	C_{7eq} **(A)**	C_{7ax} **(B)**	3_{10} **(C)**	α_L **(D)**
Parm 94	0.0	2.65	5.17	4.99
HF/6-31G*	0.86	0.0	3.81	6.32

Table 5.7 Calculated hydrogen bond parameters for Ac-PCU-NHMe

	C_{7eq} **(A)**		C_{7ax} **(B)**		3_{10} **(C)**		α_L **(D)**	
	d/Å	A/^{0}	d/Å	A/^{0}	d/Å	A/^{0}	d/Å	A/^{0}
Parm 94	2.02	131.7	1.94	142.8	2.71	96.3	3.39	73.65
HF/6-31G*	2.14	139.9	2.07	144.3	3.33	87.6	3.25	89.6

Table 5.8 Calculated torsional angles for Ac-PCU-NHMe

	Torsional angles/degrees							
	C_{7eq}		C_{7ax}		3_{10}		α_L	
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
Parm 94	-61.1	74.9	69.2	-68.8	-50.3	-28.6	44.8	54.3
HF/6-31G*	-79.8	79.6	80.2	-69.8	-66.5	-39.9	65.8	43.6

All the data in Tables 5.9 and 5.10 were obtained by using the labelled three-dimensional Ac-PCU-NHMe structure (Figure 5.11).

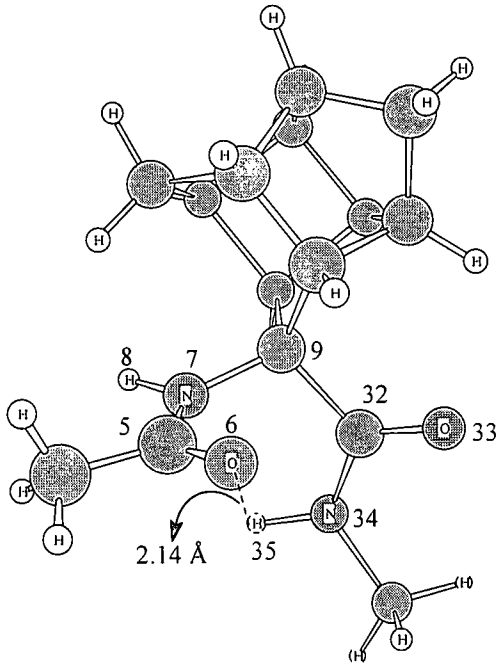


Fig. 5.11 C_{7eq} conformer for the Ac-PCU-NHMe

The numbering in Figure 5.11 was systematically adopted to obtain the data for the peptide. This data is presented in the Tables below.

Table 5.9 Calculated bond lengths in Å for Ac-PCU-NHMe

	O6-C5	N7-C5	H8-N7	C9-N7	C32-C9	O33-C32	N34-C32	H35-N34
					$C_{7\,eq}$			
HF/6-31G*	1.207	1.348	0.994	1.465	1.550	1.204	1.347	0.997
					$C_{7\,ax}$			
HF/6-31G*	1.207	1.347	0.994	1.464	1.549	1.204	1.345	0.997
					3_{10}			
HF/6-31G*	1.198	1.360	0.993	1.458	1.547	1.200	1.353	0.992
					α_L			
HF/6-31G*	1.198	1.359	0.993	1.460	1.552	1.198	1.356	0.992

Table 5.10 Calculated bond angles in degrees for Ac-PCU-NHMe

	O6C5N7	C5N7H8	N7C9C32	C9C32C34
			$C_{7\,eq}$	
HF/6-31G*	123.4	117.7	105.6	115.8
			$C_{7\,ax}$	
HF/6-31G*	123.5	117.3	107.7	115.8
			3_{10}	
HF/6-31G*	122.4	117.7	108.9	115.8
			α_L	
HF/6-31G*	122.5	117.8	107.5	114.8

5.4 A COMPARATIVE STUDY OF AC-TRISH-NHME WITH AC-PCU-NHME

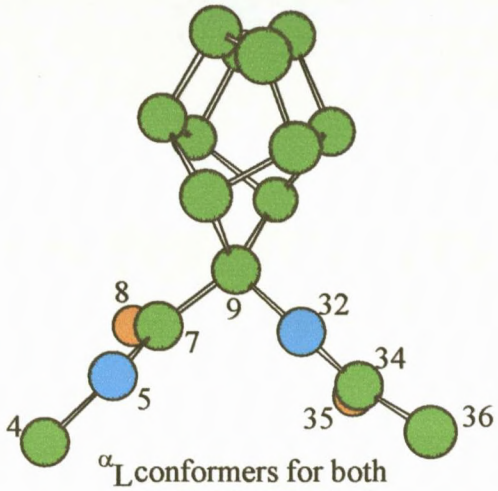
5.4.1 Ramachandran Plots

AMBER AND HF/6-31G*

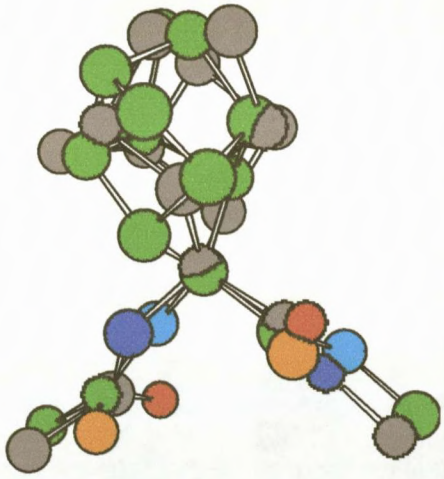
With reference to Figures 5.1, 5.2, 5.9 and 5.10, the similarities observed are the region of local minima obtained for both peptides. The conformational search revealed the C_{7ax}, C_{7eq}, 3_{10} and α_L conformers as being common to both peptides of interest with the methods of calculations used. Careful inspection of the Ramachandran plots show that both peptides have local minima appearing at similar coordinates. It is therefore expected that both cage structures will induce the same conformational constraints when experimentally incorporated in non-natural peptides.

5.4.2 Bond lengths, Bond Angles, Torsional Angles and Hydrogen Bond Parameters

In order to show some similarity and differences between the two previously mentioned peptides an overlay of one dipeptide over the other was conducted. The importance of this overlay could show how close each dipeptides actual conformational preferences are to each other. The following Figures given below show the overlays of both dipeptides.



$^{\alpha}$L conformers for both Ac-Trish-NHMe & Ac-PCU-NHMe



Overlay of both structures

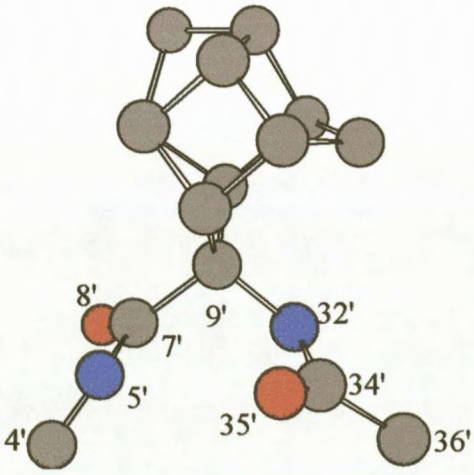
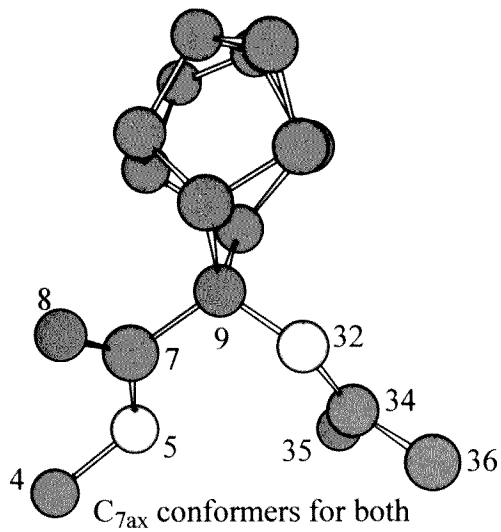
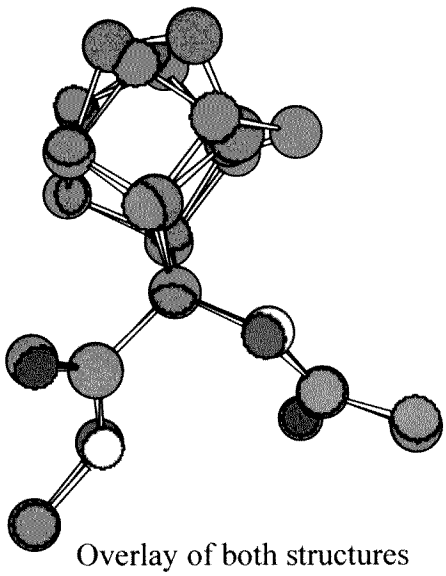


Fig.5.12 Comparison of the α_L conformers for both Ac-Trish-NHMe & Ac-PCU-NHMe



C_{7ax} conformers for both Ac-Trish-NHMe & Ac-PCU-NHMe



Overlay of both structures

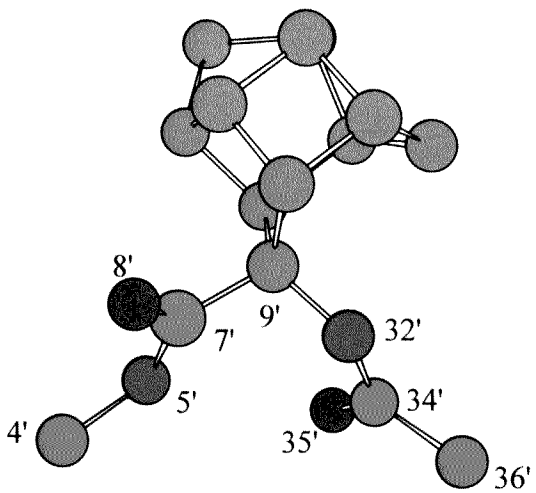
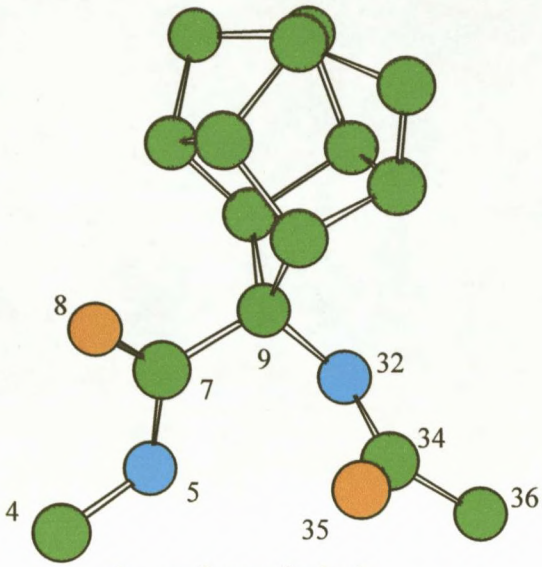
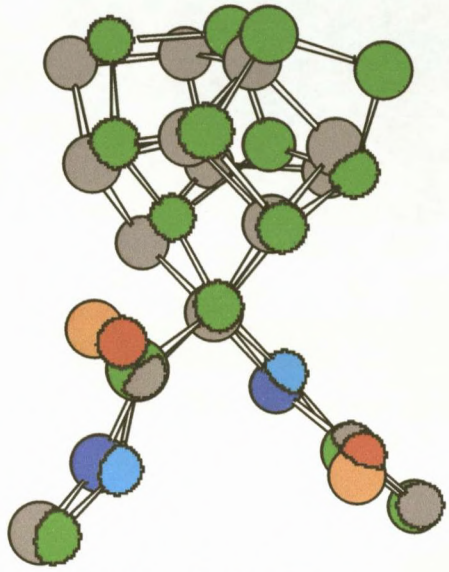


Fig.5.13 Comparison of the C_{7ax} conformers for both Ac-Trish-NHMe & Ac-PCU-NHMe



3_{10} conformer for both
Ac-Trish-NHMe &
Ac-PCU-NHMe



Overlay of both structures

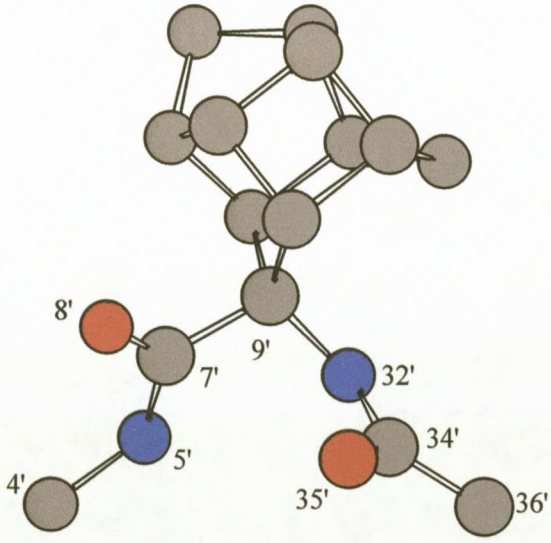
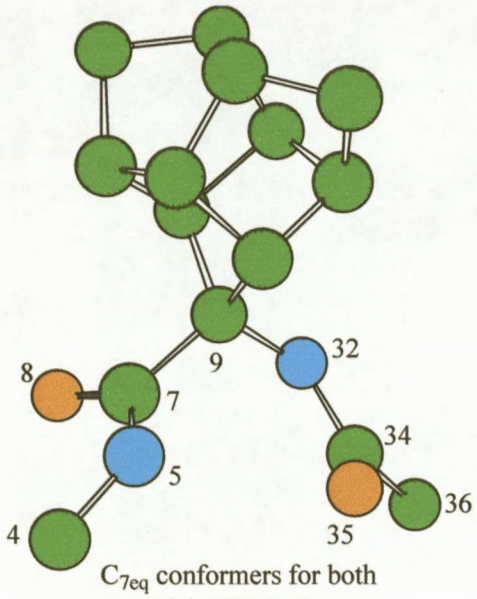
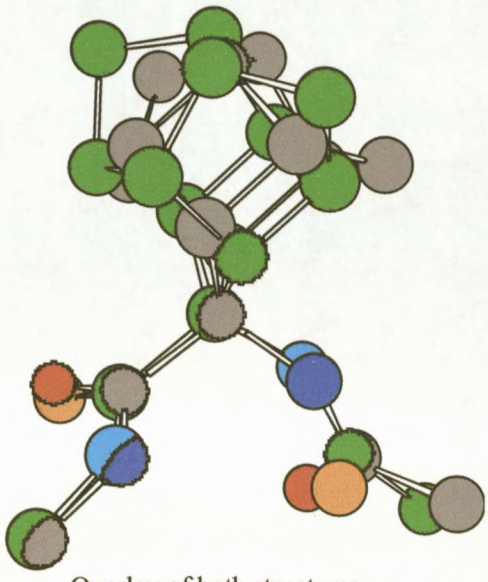


Fig.5.14 Comparison of the 3_{10} conformers for both Ac-Trish-NHMe & Ac-PCU-NHMe



C_{7eq} conformers for both Ac-Trish-NHMe & Ac-PCU-NHMe



Overlay of both structures

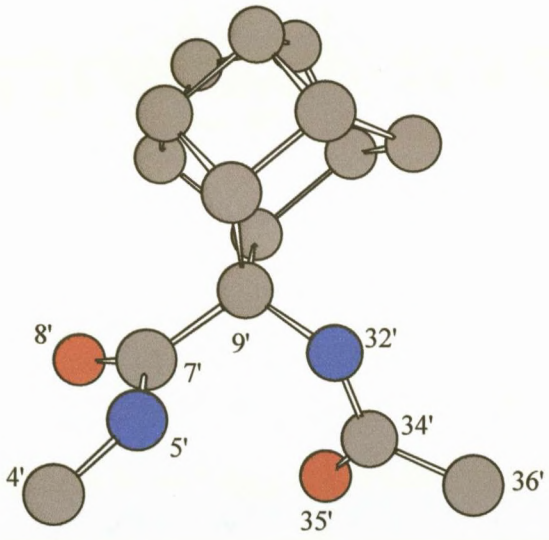


Fig.5.15 Comparison of the C_{7eq} conformers for both Ac-Trish-NHMe & Ac-PCU-NHMe

A closer inspection of all the results presented thus far shows a relative small difference, in torsion angles, bond lengths, bond angles and hydrogen bond parameters of both dipeptides.

The following data was recorded by viewing the distance of each element of the back-bone structure of each peptide to that of the other peptide, presented in the Table 5.11 below.

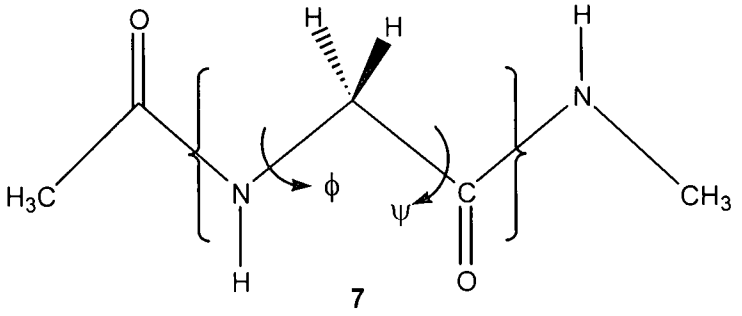
Table 5.11 Distance between the backbone elements of Ac-Trish-NHMe to Ac-PCU-NHMe for the conformers α_L, C_{7ax}, 3_{10} and C_{7eq} at the ab-initio level

	C4-C4'	N5-N5'	C7-C7'	O8-O8'	C9-C9'	N32-N32'	C34-C34'	O35-O35'	C36-C36'
α_L	0.477	0.418	0.128	0.501	0.078	1.094	0.280	2.496	1.212
C_{7ax}	0.078	0.154	0.007	0.141	0.007	0.084	0.087	0.053	0.068
3_{10}	0.204	0.291	0.085	0.457	0.171	0.208	0.068	0.276	0.133
C_{7eq}	0.050	0.129	0.114	0.252	0.199	1.153	0.142	2.266	1.345

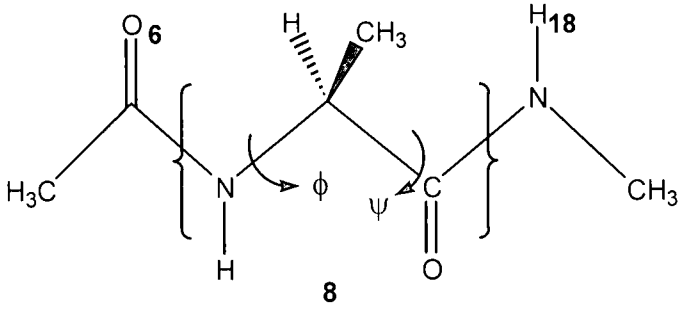
This information may seem a bit trivial but it gives one some indication of the similarity of the behaviour of these cage peptides. From the results presented thus far it can be deduced that the bulkiness of the substituent on a C_α carbon will cause the back-bone of the peptide to either partake in hydrogen bonding or move to a different conformation thus resulting in some stability.

5.5 CONFORMATIONAL STUDY USING MODEL PEPTIDES, AC-GLY-NHME, AC-ALA-NHME AND AC-AIB-NHME

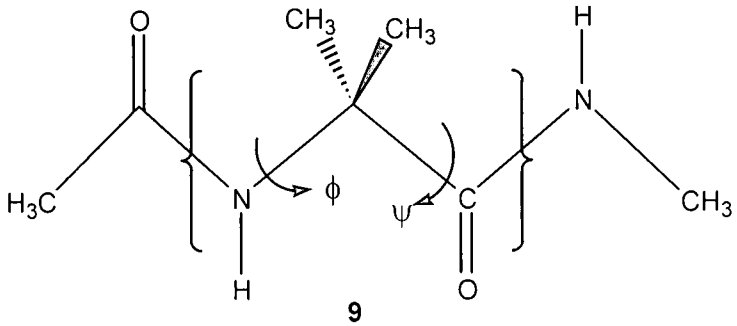
The following three-dimensional structures **7-9**, represent the naturally occurring amino acids glycine, alanine and aminoisobutyric acid respectively.



7



8



9

These model peptides are the building blocks that make up globular proteins. They are known to form stable hydrogen-bonded conformations and because of their simplistic structures tend to have great flexibility. This attribute contributes to the understanding of their structures and function as proteins. Bisetty et al[21] examined the preferred conformations of these building blocks, by looking at the geometric characteristics and came up with the following: -

Glycine (7) – Since glycine lacks a side chain in its structure it exhibits greater flexibility of conformational freedom. Calculations performed at both the molecular mechanics and *ab initio* level showed that both C_7 and C_5 conformations display structural features, which are characteristic of hydrogen-

bonded systems. The relative stability of these conformers can be represented as $C_7<C_5$.[21]

Alanine **(8)** – Three unique minima, C_{7ax}, C_{7eq} and C_5 were identified. The ordering of the energies of these conformers can be shown as $C_{7eq}< C_5<C_{7ax}$. The intramolecular hydrogen bond between H18---O6 for the C_{7ax} and C_{7eq} structure was 2.03Å and 2.26Å respectively[21], while the C_5 conformation showed a hydrogen bond between H8---O16 of 2.22 Å. The C_{7eq} conformer shows a tendency towards a gamma (γ) turn in peptides and the C_5 leans towards a β-sheet structure.

Aminoisobutyric acid **(9)** – Since similar results were obtained for the parm 91, parm 94 and parm 96 amber force field levels of calculation only the parm 94 and parm 96 force fields were used for Aib. Yet again the C_5 and C_7 conformer was immediately identified. In the structure **9** it can be seen that the C_α carbon is substituted by two alkyl groups as compared to structures **7** and **8**. This disubstitution has an affect on the flexibility of the peptide conformer. As a result the aminoisobutyric residue has a tendency to exhibit helical structures.[21]

These peptides are important to study since these fragments show conformational variations which are similar to proteins. A similarity can be drawn with Aib and Ac-Trish-NHMe since both peptides have a C_α carbon substituent. Summarily it can be seen that this substitution poses a restriction on the flexibility of the peptide thereby causing them to favour helical conformations for greater stability.

A comparison of Ac-Trish-NHMe with alanine glycine and aminoisobutyric acid reveals, glycine tends to favour extended conformations, alanine slightly on the helical side while aminoisobutyric acid conformation is very similar to that exhibited by the Ac-Trish-NHMe. Interestingly, the intramolecular hydrogen bond predicted for the C_7 conformation of the Aib dipeptide (H---O bond length

of 2.01 Å and N---H---O angle of 148.4^0) closely resembles the C_{7ax} conformation obtained at a similar computational level[21] for Ac-Trish-NHMe.

Ac-Trish-NHMe residue has the tendency to adopt helical conformations. Peptides that have large substituents on their alpha carbons result in structures that stabilise themselves through intramolecular hydrogen bonding (Figure 3.4).

CHAPTER 6

CONCLUSION

The conformational profile of Ac-Trish-NHMe (compound **4b**) was successfully explored and similar results were obtained from both the amber and *ab initio* methods employed. Comprehensive analysis of the Ramachandran maps performed at the different levels of calculation revealed the tendency of Ac-Trish-NHMe to prefer helical conformations *viz,* C_{7ax}, C_{7eq}, 3_{10} and α_L as indeed was the case of Ac-PCU-NHMe.[21] In summary the results obtained for Ac-Trish-NHMe and that of Ac-PCU-NHMe obtained by Bisetty et.al,[21] revealed that the highly strained cyclic cage peptide has a tendency to form helical structures. As discussed in Chapter 5, this means that the peptide will stabilise itself through intramolecular hydrogen bonding, resulting in helical structures.

Although X-ray diffraction and modern 2D NMR techniques are experimental probes used to study the conformational profile of peptides, the results of this theoretical investigation showed that both amber and *ab initio* calculations can be very useful in the exploration of the conformational profile of potential cases especially in the absence of experimental results. The results obtained show a satisfactory evaluation of the proposed aim of this project.

APPENDIX

AMPREP FILE

```
#!/bin/csh -f
set DIR="/usr/people/amber/AMBER"
echo removing old output..
/bin/rm -f  $1.out $1.par    .

echo PREP:
${DIR}/exe/prep   -i $1.in \
            -o $1.out \
            -p $1.par         || goto error

echo ""
exit(0)

error:
echo "  ====================================> Program error"
exit(1)
```

AMLINK FILE

```
#!/bin/csh -f
set DIR="/usr/people/amber/AMBER"
echo removing old output..
/bin/rm -f  $1"_lnk.out" $1"_lnk.bin"

echo LINK:
${DIR}/exe/link   -i $1"_lnk.in" \
            -o $1"_lnk.out" \
            -p cagedb94.dat\
            -l $1"_lnk.bin"           || goto error

echo ""
exit(0)
NEWCAGE_LNK FILE
```

```
Conformational study of Ac-Trish-NHMe

DU

    0    1    1    1    0
    acetyl-X-NMe
P   0    0    1    3    0
ACE 2YYY  NME

QUIT
```

AMEDIT FILE

```
#!/bin/csh -f
echo " "
echo "  Amber 5.0 EDIT "
echo " "

set DIR="/usr/people/amber/AMBER"

rm -f $1_edt.out $1_edt.bin $1.pdb

echo EDIT:
${DIR}/exe/edit -i gen_edt.in \
            -o $1_edt.out \
            -l $1_lnk.bin \
            -po $1.pdb \
            -e $1_edt.bin           || goto error

echo ""
exit(0)

error:
echo "  ====================================> Program error"
exit(1)
exit(0)
```

AMPARM FILE

```
#!/bin/csh -f
echo "  Amber 5.0 PARM"
echo " "

set DIR="/usr/people/amber/AMBER"



echo file 1..

/bin/rm -f  $1_prm.out $1.crd $1.top

${DIR}/exe/parm -i gen_prm.in \
                        -o $1_prm.out \
                        -e $1_edt.bin \
                        -f parm94.dat \
                            -c $1.crd \
                        -p $1.top
```

AMSANDER FILE

```
acc
 &cntrl imin=1, maxcyc=4400, nrun=0, nsnb=50, idiel=1, cut=8.0,
nmropt=0,
 scee=1.2,
 &end
 &wt type="END", &end
 &rst iat(1)=5, iat(2)=7, iat(3)=9, iat(4)=32,
 r1=-210, r2=-180, r3=-180, r4=-150,
 rk2=50.0, rk3=50.0, &end
 &rst iat(1)=7, iat(2)=9, iat(3)=32, iat(4)=34,
 r1=-210, r2=-180, r3=-180, r4=-150,
 rk2=50.0, rk3=50.0, &end
 &rst iat=0
 &end
```

REFERENCES

1. J.Gomez-Catalan, J.J. Perez, A.I. Jimenez and C. Cativiela, J. Peptide Sci., 1999, 251.
2. H. Wenschuh, M. Beyermann, H. Haber, J.K. Seydell, E. Krause, M. Bienert, L.A. Carpino, A. El-Faham and F. Albericio, J. Org. Chem., 1995, 60, 405.
3. D.K. Chalmers and G.R. Marshall, J. Am. Chem. Soc., 1995, 117, 5927.
4. C. Toniolo, Biopolymers, 1989, 28, 247.
5. V.J. Hruby, F. Al-Obeidi and W. Kazmierski, J. Biochemistry, 1990, 268, 249.
6 S. Eğe, 'Organic Chemistry:-Structure and Reactivity', 4th Ed., Houghton Mifflin Co., USA, 1999, pp. 1006-1007.
7 T.W.G. Solomons, 'Organic Chemistry', 6th Ed., John Wiley and Sons, USA, 1996, pp. 1170.
8. C.J. Noren, S.J. Anthoney-Cahill, M.C. Griffith and P.G. Schultz, Science, 1989, 224, 82.
9. J.R. Roesser, M.S. Chorghade and S.M. Hecht, 'Biochemistry', 1986, 25, 6361.
10. S. Hanessian, G. McNaughton-Smith, H.G. Lombart and W.D. Lubell, Tetrahedron, 1997, 53, 12789.
11. K.B. Brookes, P.W. Hickmott, K.K. Jutle and C.A. Schreyer, S.Afr. J. Chem., 1992, 45, 8.
12. D.W. Oliver, T.G. Dekker and F.O. Snykers, Eur. J. Med. Chem., 1991, 26, 375.
13. A.I. Jimenez, C. Cativiela, , A. Aubry and M. Marraud, J. Am. Chem. Soc., 1998, 120, 9452.
14. B.C. Wilkes and P.W. Schiller, Biopolymers, 1994, 34, 1213-1219.
15. C. Thurieau, M. Feetou, P. Hennig, E. Raimbaud, E. Canet and J.L Fauchere, J. Med. Chem.,1996, 39, 2095-2101.
16. B. Vitoux, A. Aubry, M.T. Cung and M. Marraud, Int. J. Peptides Protein Res., 1986, 27, 617-632.
17. F.J.C. Martins, A.M. Viljoen, H.G. Kruger, L. Fourie, J. Roscher, A.J. Joubert and P.L Wessels, Enantioselective Synthesis of Amino Acids from Pentacyclo[5.4.0.02,6.03,10.05]undecane-8,11-dione, Tetrahedron, 2000, 57(8), 1601.
18. T. Govender, M.Sc. dissertation, *University of Natal, Durban*, **2001**
19. T. Govender, H.K. Hariprakasha, H.G. Kruger and T. Raasch, Synthesis of novel Trishomocubane amino acid derivatives, Submitted for publication in Tetrahedron.

20. T Raasch, M.Sc. dissertation, *University of Natal, Durban*, **2003**.

21. K. Bisetty, Ph.D. thesis, *University of Natal, Durban*, **2002.**

22. K. Bisetty, J.Gomez-Catalan, C. Aleman, E. Giralt, H.G. Kruger and J.J. Perez, accepted for J. Peptide Sci., 2003.

23. G.D. Rose, L.M. Gierasch and J.A. Smith, Adv. Protein Chem, 1985, 37, 1-89.

24. D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley, D.M. Ferguson, R.J. Radmer, G.L. Seibel, U.C. Singh, P.K. Weiner and P.A. Kollman (1997), AMBER 5.0, University of California, San Francisco.

25. I.N. Levine, 'Quantum Chemistry', 3rd Ed., Allyyn and Bacon, Boston, 1983.

26. S.S. Mader, 'Biology', The McGraw Hill Co., New York, 1985, pp. 45-46.

27. S.R. Buxton and S.M. Roberts, 'Guide to Organic Stereochemistry', Addison Wesley Longman Limited, 1996, pp. 20-21.

28. C. Ramakrishnan and G.N. Ramachandran, J. Biophys., 1965, 5, 909.

29. A.N. Voldeng, C.A. Bradley, R.D. Kee, E.L. Knight and F.L. Melder, J. Pharm. Sci., 1968, 57, 1053.

30. Y. Inamoto, K. Aiyami, N.Tasaishi and Y.J. Fujikura, J. Med. Chem., 1967, 19, 536.

31. Y. Inamoto, K. Aiyami, T. Kadono, H. Makayama, A. Takatsuki and G. Tumura, J. Med. Chem., 1977, 20, 1371.

32. W.J. Hehre, L. Radom, P.von R. Schleyer and J.A. Pople, 'Ab initio Molecular Orbital Theory', John Wiley and Sons, New York, 1985, pp. 10-14.

33. J.B. Foresman and A. Frisch, 'Exploring Chemistry with Electronic Structure Methods: A Guide to Using Gaussian', Gaussian, Inc., Pittsburgh, PA, 1993, pp. 3-8.

34. W.G. Richards and D.L Cooper, 'Ab initio Molecular Orbital calculations for Chemists', 2nd Ed., Clarendon Press, Oxford, 1983, pp. 9-20.

35. Reference 25, pp. 203-211.

36. Gaussian 98, Revision A.3: M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, M. Head-Gordon, E.S. Replogle and J.A. Pople, Gaussian, Inc., Pittsburgh, PA, 1998.

37. H.D. Höltje and G. Folkers, 'Methods and Principles in Medicinal Chemistry', VCH Publishers Inc., New York, USA, 1996, 5, pp. 81-88.

38. N.L. Allinger, Y.H. Yuh and J.H. Lii, J. Am. Chem. Soc., 1989, 111, 8551.

39. T. Head-Gordon, M. Head-Gordon, M.J. Frisch, C.L Brooks III and J.A.Pople, J. Am. Chem. Soc., 1991, 113, 5989.

40. H.J. Böhm and S. Brode, J. Am. Chem. Soc., 1991, 113, 7129.

41. C.H. Hu, M. Shen and H.F. Schaefer III, J. Am. Chem. Soc., 1993, 115, 2913.

42. P.C. Hariharan and J.A. Pople, Theor. Chim. Acta., 1973, 28, 213.

43. A.R. Leach, 'Molecular Modelling: Principles and Applications', Addison Wesley Longman limited, 1996, pp. 131-133.

44. K.B. Lipkowitz and D.B Boyd, 1991, pp. 1-2, volume 2.

45. A.K. Ghose, G.M. Crippen, G.R. Revankar, D.F. Smee, P.A. McKernan and R.K. Robins, J. Med. Chem., 1989, 32, 746-756.

46. G.N. Ramachandran and V. Sasisekharan, Adv. Prot. Chem., 1968, 23, 283-437.

47. I.R. Gould, W.D. Cornell and I.H. Hillier, J. Am. Chem. Soc., 1994, 116, 9250.

48. H. Hart, L.E. Craine and D.J. Hart, 'Organic Chemistry:A Short Course', 10th Ed, 1999, Houghton Mifflin Company, pp. 510

49. G.G. Ferenczy, C.A. Reynolds and W.G. Richards, J. Comput. Chem., 1990, 11, 159-169.

50. T. Clark, 'A Handbook of Computational Chemistry', 'Ab initio Molecular Orbital Theory', John Wiley and Sons, New York, 1985, pp. 12-20.

51. W.J. Richards and J.A. Horsley, 'Ab initio Molecular Orbital Calculations for Chemists', Oxford University Press, Oxford, 1970, pp. 1-10.

52. S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, J. Am. Chem. Soc., 1984, 106, 765.

53. P.Y. Bruice, 'Organic Chemistry', 3rd Ed, 2001, Prentice Hall, Upper Saddle River, pp. 92.

54. J. B. Henrickson, D.J. Cram and G. S. Hammond, 'Organic Chemistry', 3rd Ed, 1970, McGraw-Hill Book Company, New York.