UNIVERSITY OF LONDON THESIS

Degree PhD Year 2006 Name of Author TOROZ, Dimitrios

Exploration of the conformational energy landscape of small peptides using electronic structure methods

Dimitrios Toroz

University College London

Supervisors:
Dr. Tanja van Mourik
Prof. Jonathan Tennyson

A thesis submitted for the degree of Doctor of Philosophy
of the University College London.

London
September, 2006

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

Exploration of the conformational energy landscape of small peptides using electronic structure methods

Dimitrios Toroz
University College London

Abstract

Small peptides work as neurotransmitters or hormones in the body. For example, the pentapeptide enkephalin is involved in a wide variety of physiological processes such as mediation of pain and respiratory depression. In order to understand the biological function of these molecules it is necessary to examine their molecular shape and structural preferences. The large flexibility of peptides makes them difficult to be characterized by experimental and theoretical methods.
A method was developed in order to explore the conformational preferences of a polypeptide using electronic structure methods based on hierarchical selection criteria. The strategy was to vary all the torsion angles of the peptide and to create all possible conformers. The conformers were assessed according to the number of hydrogen-bonding interactions in their structure. Calculations were performed at increasingly higher levels of theory. Only a number of the most stable conformers were taken through to the next level.
This hierarchical selection method was used to explore the conformational features of the dipeptide Tyr-Gly (the first two amino acids of the pentapeptide enkephalin: Tyr-Gly-Gly-Phe-Met).The conformational preferences of Tyr-Gly were also explored with a stepwise rotation method. The hierarchical selection method seemed to be superior. The most stable conformers found for Tyr-Gly are characterized by a characteristic hydrogen-bonding interaction (O-H•••O) between the hydroxyl hydrogen of glycine and the carboxyl oxygen of tyrosine. The optimized structures obtained with DFT differ from the structures obtained by MP2 geometry optimizations. MP2 optimizations make the structures more folded.
The method has also been used to study the conformational features of the Tyr-Gly-Gly tripeptide, Tyr-Gly-Gly-Phe tetrapeptide, Tyr-Gly-Gly-Phe-Leu pentapeptide and Gly-Gly-Gly tripeptide. The most stable conformers obtained for the Tyr-Gly-Gly tripeptide are characterized by folded structures with a characteristic hydrogen-bonding interaction between the (-OH) phenyl group of tyrosine and the carboxyl oxygen of glycine (3). The most stable conformers obtained for Tyr-Gly-Gly-Phe and Tyr-Gly-Gly-Phe-Leu are characterized by folded structures with a characteristic hydrogen-bonding interaction between the (-OH) phenyl group of tyrosine and the carboxyl oxygen of phenylalanine (4).
For Tyr-Gly, Tyr-Gly-Gly, Tyr-Gly-Gly-Phe and Tyr-Gly-Gly-Phe-Leu a large variation has been observed between DFT and MP2 orders of stability of the conformers. DFT fails to describe the dispersion effects arising from interactions involving the aromatic residues. MP2 would probably describe more accurately the conformational preferences of these peptides. However, the large basis set superposition error in MP2 calculations means that also MP2 may not be suitable to characterize the conformational preferences of these peptides.
The hierarchical selection method developed has been shown to be a useful method to study small peptides.

Acknowledgments

Firstly many thanks to my parents and to my brother for all their support and encouragement to complete this project all of these years.

I would like to thank my supervisor Dr Tanja van Mourik for her great ideas and her encouragement during this project.

Many thanks due to Professor Jonathan Tennyson for his great support during this project.

Many thanks are also due to Emmy for all the help and support all of these years.

Thanks are also due to Georgia Makridaki for her help and encouragement for many years.

I would like to thanks all my colleagues from room 105 in the chemistry department at UCL.

Finally thanks due to Chemreact and EPSRC for the financial support all of these years.

TABLE OF CONTENTS

1

INTRODUCTION

Molecular modeling is defined as anything that requires the use of a computer to describe any aspect of the properties of a molecule. Methods that have been implemented so far include techniques for the prediction of properties of large molecules such as docking of ligands in the active site of protein structures. In addition, this discipline includes methods which are used in computational chemistry such as computation of the energy of the molecular system and geometry optimization of a molecular system. Thus, computational chemistry can be described as the nucleus for molecular modeling.

A large number of studies have been carried out in the past to describe the biological function of peptides i.e. to understand the molecular mechanism of the specific biological activity of a peptide[1-6]. The aim of these studies is the design of new active molecules that can be used successfully as drugs. However in order to evaluate the biological function of a peptide it is necessary to describe the energetics of a peptide such as its conformational preferences and the intramolecular and intermolecular interactions that dominate especially in the low-lying conformers of a peptide. Furthermore, due to the fact that the accuracy of a simulation is not always precise, when it is possible, the computational results obtained from simulations have to be compared with experimental results to confirm the accuracy of the simulation model.

1.1. Geometry optimization

Geometry optimization is the process of starting with an input structure and finding a stationary point on the potential energy surface. The stationary point found will be the one closest to the input structure, not necessarily the global minimum. A minimum is local when it is the lowest point in some limited region of the potential energy surface. When the minimum is the lowest point on the potential energy surface it is called the global minimum. An efficient geometry optimization requires a sophisticated algorithm. It is necessary to know in which

direction to move, and how far in that direction. The first derivative of the energy (the gradient) provides information on the direction, whereas the second derivative (the curvature) provides information on the step size. In general it is not possible to go from the input structure to the minimum in just one step, but modern geometry optimization algorithms commonly reach the minimum in about 10 steps, given a reasonable input geometry. The most widely-used algorithms for geometry optimizations use the first and second derivatives of the energy with respect to the geometric parameters. In practice to understand the geometry optimization procedure we can consider the one-dimensional PES of a diatomic molecule (Figure 1)[7]. The input structure is at the point P_i (E_i,r_i) (E_i is the energy of the diatomic molecule and r_i is the bond length of the diatomic molecule at the input structure) and the approximate minimum, corresponding to the optimized structure, is at the point P_o (E_o,r_o). Before the optimization has been carried out the values of E_o and r_o are unknown.

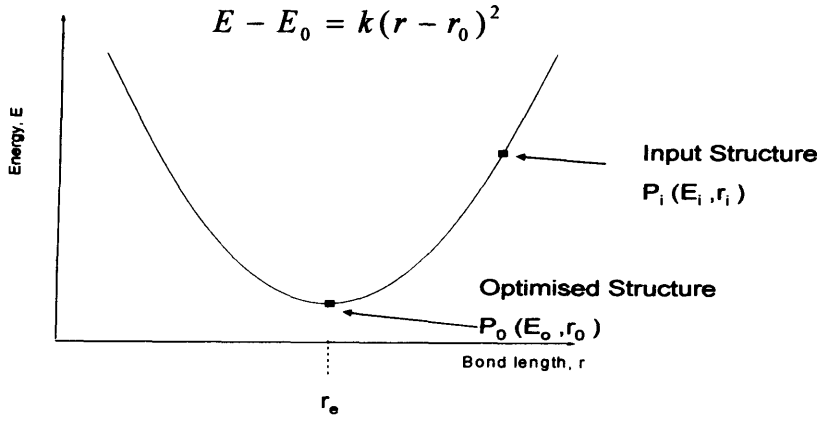


Figure 1. The potential energy of a diatomic molecule near the equilibrium geometry is approximately a quadratic function of the bond length. From an input structure the optimized structure can be found in one step if the function is strictly quadratic.

The potential energy surface is a quadratic function of r:

$$E - E_0 = k(r - r_0)^2 \qquad (1.1.1)$$

At the point corresponding to the input structure the first derivative gives:

$$(dE/dr)_i = 2k(r_i - r_0) \qquad (1.1.2)$$

For any arbitrary point, the second derivative is given by:

$$d^2E/dr^2 = 2k \qquad (1.1.3)$$

From the equations (1.1.2, 1.1.3) we get,

$$(dE/dr)_i = d^2E/dr^2 /(r_i - r_0) \qquad (1.1.4)$$

$$r_0 = r_i - (dE/dr)_i /(d^2E/dr^2) \qquad (1.1.5)$$

Equation 1.1.5 shows that if we know dE/dr, the slope or gradient of the potential energy surface (PES), at the point of the initial structure, d^2E/dr^2 the curvature of the PES and r_i, the initial geometry, we can calculate r_o, the optimized geometry in one step. The potential energy surface is not generally exactly quadratic and therefore we need more steps to get a minimum. Ideally, an optimization has converged when the gradient and the step (curvature) will be zero at the optimized minimum. In practice the optimization converged when the the gradient and the step (curvature) are sufficiently small. In most algorithms the second derivatives are only calculated approximately. Thus in order to get a minimum on the potential energy surface an iterative procedure is needed because the PES energy surface is not exactly quadratic and the second derivatives are calculated approximately. Some potential energy surfaces are far from the quadratic and this can lead to problems finding convergence to the geometry optimization.

Once a stationary point has been found it is desirable to check whether it is a minimum or a transition state. This is done by calculating the vibrational frequencies. Such a calculation involves finding the normal-mode frequencies

(these are the simplest vibrations of the molecule). The normal-mode frequency is given by the equation,

$$v = \frac{1}{2\pi c}\left(\frac{k}{\mu}\right)^{1/2} \qquad (1.1.6)$$

where v is the vibrational frequency (wavenumber in cm^{-1}), c is the velocity of light, k is the force constant for the vibration, μ is the reduced mass of the molecule. The force constant k of a vibrational mode is a measure of the stiffness of the molecule towards the vibrational mode - the harder it is to stretch or bend the molecule in the manner of that mode, the bigger is the force constant. For a minimum all the vibrations are real, while a transition state has one imaginary vibration corresponding to motion along the reaction coordinate.

1.2 Hartree-Fock theory

Quantum mechanics for any system begins with the Schrödinger equation:

$$H\Psi = E\Psi \qquad (1.2.1)$$

H is the Hamiltonian operator and E is the system energy. The wavefunction,Ψ, describes the position of all atomic nuclei and electrons present.

The Hamiltonian operator for the nuclei and electrons in a molecule is given by:

$$H = -\sum_{i=1}^{N_{el}} \frac{\nabla_i^2}{2} - \sum_{A=1}^{N_{at}} \frac{\nabla_A^2}{2M_A} - \sum_{i=1}^{N_{el}} \sum_{A=1}^{N_{at}} \frac{Z_A}{r_{iA}} + \sum_{i>j}^{N_{el}} \frac{1}{r_{ij}} + \sum_{A=1}^{N_{at}} \sum_{B=1}^{N_{at}} \frac{Z_A Z_B}{R_{AB}} \qquad (1.2.2)$$

for a system with N_{el} electrons and N_{at} nuclei, M_A is the nuclear mass of atom A in unit of electron mass, Z_A is the charge on nucleus A, r_{iA} is the distance of electron i from nucleus A, r_{ij} is the distance between electrons i and j. and R_{AB} is the distance between A and B.

The Hamiltonian can be simplified using the Born-Oppenheimer approximation, which derives from the observation that nuclei are much heavier that electrons, and consequently move much more slowly than electrons. From this approximation the Hamiltonian H becomes:

$$H = -\sum_{i=1}^{N_{el}} \frac{\nabla_i^2}{2} - \sum_{i=1}^{N_{el}} \sum_{A=1}^{N_{at}} \frac{Z_A}{r_{iA}} + \sum_{i>j}^{N_{el}} \frac{1}{r_{ij}} \qquad (1.2.3)$$

A two-electron wavefunction can be expressed as a product of one-electron molecular orbitals.

$$\Psi(\phi_1,\phi_2) = \frac{1}{\sqrt{2}}[\phi_1(1)\phi_2(2) - (\phi_1(2)\phi_2(1)] \qquad (1.2.4)$$

This can be expressed as a Slater determinant which ensures that the wavefunction is anti-symmetric with respect to electron exchange, producing the two-electron wavefunction[8,9]

$$\Phi_e = \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_1(1) & \phi_2(1) \\ \phi_1(2) & \phi_2(2) \end{vmatrix} \qquad (1.2.5)$$

The above determinant can be generalized for N electrons by the following equation:

$$\Phi_e = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & ..... & \phi_K(1) \\ \phi_1(2) & \phi_2(2) & ..... & \phi_K(2) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \phi_1(N) & \phi_2(N) & ..... & \phi_K(N) \end{vmatrix} \qquad (1.2.6)$$

Swapping two electrons corresponds to interchanging two rows of the determinant. This will have the effect of changing the sign of the determinant which is required for antisymmetry.

Molecular orbitals are orthogonal,

$$\int \phi_i^* \phi_i d\tau = 1 \qquad (1.2.7)$$

$$\int \phi_i^* \phi_j d\tau = 0 \quad , i \neq j \qquad (1.2.8)$$

i.e. the integral over all space for an MO with itself is 1 but with any other is 0.

At the Hartree-Fock theory the Schrödinger equation becomes:

$$H_e \Phi_e = E_e \Phi_e \qquad H_e = T_e + V_{ne} + V_{ee} \qquad (1.2.9)$$

Here the wavefunction for the system is a Slater determinant and the Hamiltonian is a sum of the electron kinetic energy, T_e, the potential energy for the interaction of the electrons with the nuclei, V_{ne}, and the potential energy from electron-electron interactions V_{ee}.

By pre-multiplying by Φ_e in equation 2.8 the system energy of the system E_e can be solved forming the integrals:

$$\langle \Phi_e | H_e | \Phi_e \rangle = E_e \langle \Phi_e | \Phi_e \rangle \qquad (1.2.10)$$

The wavefunction is written as a Slater determinant which multiplies out to:

$$\Phi_e = \frac{1}{\sqrt{N!}} (\phi_1(1)\phi_2(2).....\phi_N(N) - \phi_2(1)\phi_1(2).....\phi_N(N) + ...) \qquad (1.2.11)$$

The integral on the right hand side of equation 2.9 will contain terms like:

$$\begin{aligned} &\langle \phi_1(1) | \langle \phi_2(2) | .....\langle \phi_N(N) || \phi_1(1) \rangle | \phi_2(2) \rangle ..... | \phi_N(N) \rangle \\ &= \langle \phi_1(1) | \phi_1(1) \rangle \langle \phi_2(2) | \phi_2(2) \rangle .....\langle \phi_N(N) | \phi_N(N) \rangle \end{aligned} \qquad (1.2.12)$$

and terms like:

$$\begin{aligned} &\langle \phi_1(1) | \langle \phi_2(2) | .....\langle \phi_N(N) || \phi_1(2) \rangle | \phi_2(1) \rangle ..... | \phi_N(N) \rangle \\ &= \langle \phi_1(1) | \phi_2(1) \rangle \langle \phi_2(2) | \phi_1(2) \rangle .....\langle \phi_N(N) | \phi_N(N) \rangle \end{aligned} \qquad (1.2.13)$$

For normalized molecular orbitals the first of these is 1 and the second 0.

There are N! integrals of the first type and so the overlap integral of the total wavefunction becomes :

$$\langle \Phi_e | \Phi_e \rangle = \frac{1}{\sqrt{N!}}(1-0)(1-0)\sqrt{N!} = 1 \qquad (1.2.14)$$

1.3 The Fock operator

The Hamiltonian can be written in terms of operators that only operate on one electron at a time[10]:

$$H_e = \sum_i^N h_i + \sum_i^N \sum_{j>i}^N J_{ij} - \sum_i^N K_{ij} = \sum_i^N F_i \qquad (1.3.1)$$

Where the Fock operator, F_i, is defined as:

$$F_i = h_i + \sum_j^N J_{ij} - K_{ij} \qquad (1.3.2)$$

The Fock operator allows each electron to experience an average field from the nuclei and all other electrons. This allows its molecular orbital to be calculated. However the Fock operators will change for all other electrons if a molecular orbital is altered and therefore new molecular orbitals need to be calculated. In practice molecular orbitals are calculated until the whole solution is self consistent.

1.4 Basis sets

Molecular orbitals are commonly built from atomic orbitals[11]. Coefficients are introduced to control how much atomic orbitals should be present in a molecular orbital:

$$\phi_i = \sum_{\mu=1}^N c_{\mu i} x_\mu \qquad (1.4.1)$$

where the coefficients $c_{\mu i}$ are known as the molecular orbital expansion coefficients. One of the most common basis sets is denoted by 6-31G[12]. This means that, for a first row atom, six Gaussian functions are fit to the core 1s orbital. Each valence orbital is then represented by two functions, one that is a set

of 3 Gaussians and a second function that is a single Gaussian function. Additional polarization and diffuse functions may be added to the basis set and are indicated by an asterisk (*) and a plus sign respectively (+). A 6-31+G* basis set would indicate all heavy atoms have additional polarization and diffuse functions.

1.5 Møller-Plesset perturbation theory

Møller–Plesset perturbation theory is a method that includes electron correlation and provides adequate treatment of the correlation between the motions of electrons for a molecular system especially with electrons of opposite spin[13, 14]. Møller-Plesset perturbation theory is more expensive than Hartree-Fock but is a more accurate method as it qualitatively adds higher excitations to Hartree-Fock[15, 16].

In perturbation theory[17] the Hamiltonian H can be divided in terms of a perturbation of $H^{(0)}$.

$$H = H^{(0)} + \lambda V \tag{1.5.1}$$

Where $H^{(0)}$ is the unperturbed Hamiltonian, λV is a perturbation applied to $H^{(0)}$, and λ is a parameter. The perturbation is expressed as a wavefunction ψ_k and the energy E_k as a power series in the perturbation strength λ:

$$\begin{aligned} \psi_\kappa &= \sum_{n=0}^{\infty} \lambda^n \psi_\kappa^{(n)} \\ E_k &= \sum_{n=0}^{\infty} \lambda^n E_k^{(n)} \end{aligned} \tag{1.5.2}$$

The zeroth-order wavefunctions and energies are given by the following equations:

$$\psi_\kappa^{(0)} = \Phi_\kappa \qquad \text{and} \qquad E_k^{(0)} = E_k^{\,0} \tag{1.5.3}$$

The first-order energy correction is given in terms of the zeroth-order wavefunction as:

$$E_k^{(1)} = \langle \Phi_\kappa | V | \Phi_\kappa \rangle \tag{1.5.4}$$

The first-order wavefunction $\psi_k^{(1)}$ expressed in terms of the complete set of unperturbed functions Φ_j is:

$$\psi_\kappa^{(1)} = \sum_{j \neq k} <\Phi_j | V | \Phi_\kappa > /[E_k^0 - E_j^0] | \Phi_j > \qquad (1.5.5)$$

The second-order energy correction is expressed as follows:

$$E_k^{(2)} = \sum_{j \neq k} |<\Phi_j | V | \Phi_\kappa >|^2 /[E_k^0 - E_j^0] \qquad (1.5.6)$$

And the second-order correction to the wavefunction is expressed as

$$\psi_\kappa^{(2)} = \sum_{j \neq k} [E_k^0 - E_j^0]^{-1} \sum_{l \neq k} \{ < \Phi_j | V | \Phi_l > -\delta_{j,l} E_k^{(1)} \} \qquad (1.5.7)$$

$$< \Phi_l | V | \Phi_\kappa > [E_k^0 - E_j^0]^{-1} | \Phi_j >$$

The calculation of the energy corrections $E_k^{(n)}$ and wavefunction corrections $\Psi_k^{(n)}$ involves the unperturbed wavefunctions Φ_k, the perturbation V and the unperturbed energies E_j^0.

MP2 gives better results for the bond angles and torsions in comparison to the Hartree-Fock method due to the fact that MP2 includes electron correlation. In MP2 the Hartree-Fock wavefunction is perturbed to the second order in order to predict a better energy of the system. An important benefit of MP2 is that the size of the system examined does not have an effect on the quality of the calculated energy. In this thesis the second order perturbation will be useful because MP2 describes accurately long-range dispersion effects and can predict better results for peptides that contain aromatic residues.

1.6. Dispersion forces

Dispersion forces are weak attractive forces that even occur between atoms or nonpolar molecules caused by an instantaneous dipole moment of one atom or molecule that induces a similar temporary dipole moment in adjacent atoms or molecules[18]. The magnitude of the dispersion forces varies with molar mass. A typical magnitude of the potential energy of the dispersion interaction is about 2 kJ mol^{-1}. It has be found that for water dimer the hydrogen-bonding strength (water dimer interaction energy) is approximately (20kJ/mol)[19]. Thus dispersion is

an order of magnitude smaller than hydrogen-bonding. Dispersion is the weakest intermolecular force between neutral molecules whereas hydrogen bonding is the strongest intermolecular force. The interaction exists whenever two atoms or molecules are in proximity, whether or not other interactions are also present. The polarizability of an atom or molecule is the ease with which the electron distribution is distorted by an external electric field. The greater their polarizability, the greater the dispersion forces between two molecules. Dispersion forces are stronger for larger and heavier atoms and molecules than for smaller and lighter ones as size and polarizability increase with molar mass. For two atoms the dispersion interaction varies according to the inverse sixth power of the distance between the two atoms:

$$E_{disp} = \frac{-B_{ij}}{r_{ij}^6} \qquad (1.6.1)$$

The factor B_{ij} depends on the nature of the pair of atoms interacting (in particular their polarizability).

1.7 Density functional theory

Density functional theory is a method that includes an approximate treatment of electron correlation[20]. For this reason it should be a more accurate method than Hartree-Fock. Unlike Hartree-Fock theory that is based on wavefunctions and orbitals, density functional theory is based on the electron density (although usually orbitals are employed to get the density). In contrast with traditional quantum chemistry where the complicated many-body wavefunction is calculated, within DFT the ground state approximation is expressed as a functional of the ground state density.

The total electronic energy is a functional of the electron density in a sense that for a given function there is a single corresponding energy and is given by:

$$E = E^T + E^V + E^J + E^{XC} \qquad (1.7.1)$$

where E^T is the kinetic energy term (arising from the motion of the electrons), E^V describes the potential energy of the nuclear-electron attraction and of the

repulsion between pairs of nuclei, E^J is the electron-electron repulsion, and E^{XC} is the exchange–correlation term and includes the remaining part of the electron-electron interactions.

All terms except the nuclear-nuclear repulsion are functions of ρ, the electron density. E^J is given by the following expression:

$$E^J = \frac{1}{2}\int\int \rho(r_1)(\Delta r_{12})^{-1}\rho(r_2)dr_1dr_2 \qquad (1.7.2)$$

$E^T + E^V + E^J$ corresponds to the classical energy of the change distribution ρ.
The exchange-correlation (E^{XC}) energy functional term in equation 3 accounts for all electron correlation of the individual electrons and for the exchange energy that arises from the quantum mechanical wavefunction.

E^{xc} is usually divided into the exchange and correlation part:

$$E^{XC}(\rho) = E^x(\rho) + E^c(\rho) \qquad (1.7.3)$$

All three terms are functionals of the electron density; $E^x(\rho)$ and $E^c(\rho)$ correspond to the exchange functional and correlation functional respectively.

The E^{XC} is determined by the electron density ρ. A simple approximation for the exchange-correlation functional is the local density approximation (LDA)[21]. In the local density approximation the exchange correlation functional E_{XC} is constructed for regions of a system where the charge density is slowly varying and the exchange correlation energy at that point can be considered the same as that for a locally uniform electron gas of the same charge density and is given by:

$$E_{xc}^{LDA}[\rho] = \int \rho(r)\varepsilon_{xc}(\rho(r))dr \quad \text{and} \quad \varepsilon_{xc}(\rho(r)) = \varepsilon_x(\rho(r)) + \varepsilon_c(\rho(r)) \qquad (1.7.4)$$

where ε_{xc}, is the exchange correlation energy density.

Although LDA has been found to work well for some molecular systems, most chemical applications do not satisfy the restriction of slowly varying electron density and at this point LDA fails. However, much better results are obtained when density fluctuations were taken into account via the gradient of the density. This led to the development of various generalized gradient approximations (GGAs). An example is the popular functional given by Becke[22]:

$$E_{XC}^{Becke} = -\beta\rho^{1/3}\frac{x^2}{(1+6\beta x\sinh^{-1})} \qquad (1.7.5)$$

$$\mathbf{x}=|\nabla\rho|/\rho^{4/3}$$

This functional contains a parameter β which was chosen so that the sum of the LDA and Becke exchange terms accurately reproduce the exchange energies of six gas atoms, and has the value of β=0.0042. A large number of other GGA functionals for both correlation and exchange have been developed, of which the most popular probably the 4 parameter LYP[23] correlation functional.

A systematic error that Hartree-Fock theory has is the lack of electron correlation which results in higher energies than it should be. On the other hand, in DFT the combination of the HF exchange and the functional exchange (so-called hybrid functionals) eliminates this error. An example is the popular B3LYP functional which has the form[24]:

$$E_{Becke3LYP}^{XC} = E_{LDA}^{X} + c_0(E_{HF}^{X} - E_{LDA}^{X}) + c_x\Delta E_{B88}^{X} + E_{VWN3}^{C} + C_C(E_{LYP}^{C} - E_{VWN3}^{C}) \qquad (1.7.6)$$

where C_0=0.20, C_x=0.72, and C_c=0.81 are semi-empirical coefficients to be determined by an appropriate fit to experimental data. The B3LYP functional can be used to get better structure and energy estimations of the optimized structures resulting from Hartree-Fock method but can not take into account dispersion effects.

1.8. Basis set superposition error (BSSE)

The basis set superposition error in molecular orbital calculations is an error that arises from the incompleteness of the basis set used in the expansion of wavefunctions. Complete basis sets require a large number of basis functions. This means that in practice calculations are always done with incomplete basis sets. Thus errors in the energy from basis set incompleteness can be quite large. BSSE can occur both between two separate molecules (intermolecular BSSE) as well as between two different fragments in the same molecule (intermolecular BSSE). It has been observed that the magnitude of the intramolecular BSSE can be similar to that observed in intermolecular complexes (few Kcal/mol)[25].

In general the interaction energy between two molecules is calculated by the following expression:

$$\Delta E_{AB}(R) = E_{AB}(R) - E_A - E_B \qquad (1.8.1)$$

where $E_{AB}(R)$ is the total energy of the system AB and E_A and E_B are the energies of the subsystems A and B. The problem with the interaction energy computed by equation (1.8.1) is that it is subject to basis set superposition error. At large intermolecular distances the monomers A and B do not interact and the total energy $E_{AB}(R)$ reduces to the sum of the two monomers E_A and E_B. However, at smaller distances, the basis functions of one of the monomers in the complex AB can provide a better description of the electronic structure of the other monomer, providing an unphysical lower energy for the system *AB*. While other approaches to correct the BSSE error have been discussed in the literature[26, 27], the counterpoise procedure (CP) proposed by Boys and Bernardi continues to be the most prominent to correct for BSSE[28, 29]. In this procedure the monomer energies are evaluated in the complete dimer basis set. The counterpoise-corrected interaction energy then becomes:

$$\Delta E^{CP}(R) = E_{AB}(R) - E_A^{\{AB\}}(R) - E_B^{\{AB\}}(R) \qquad (1.8.2)$$

The monomer energies are computed in the dimer basis set. The equation (1.8.2) represents the uncorrected interaction energy $\Delta E^{noCP}(R)$. ΔE_{BSSE} is usually defined as the difference between ΔE^{CP} and ΔE^{noCP}:

$$\Delta E_{BSSE} = E_A^{\{AB\}}(R) + E_B^{\{AB\}}(R) - E_A^{\{A\}} - E_B^{\{B\}} \qquad (1.8.3)$$

The energy required to bring monomer A to a particular intermolecular geometry $\mathbf{r_A}$ is given by the deformation energy:

$$\Delta U_A^{def}(r_A) = E_A^{\{A\}}(r_A) - E_A^{\{A\}}(r_e) \qquad (1.8.4)$$

with a similar expression for B. r_e represents the equilibrium geometry of molecule A. The deformation energies should be taken into account when calculating the CP-corrected interaction energy.

The total counterpoise-corrected interaction energy, with intermolecular geometrical parameters R and intermolecular parameters r_A and r_B can then be written as:

$$\begin{aligned}\Delta E^{CP}(R,r_A,r_B) = {} & E_{AB}(R,r_A,r_B) - E_A^{\{AB\}}(R,r_A,r_B) - E_B^{\{AB\}}(R,r_A,r_B) \\ & + \Delta U_A^{def}(r_A) + \Delta U_B^{def}(r_B)\end{aligned} \qquad (1.8.5)$$

1.9 Semi-empirical methods

A disadvantage of ab initio methods and density functional theory is that they are computationally expensive especially when they are used to characterize a large molecule. To overcome this problem an alternative method, the so called semi-empirical methods have been proposed which solve the Schrödinger equation with empirical adjustments and are much less computationally expensive[30]. A disadvantage of semi-empirical methods is that that are not accurate enough in comparison with the classic ab initio methods. For this reason semi-empirical methods have been used as a method that can predict a first guess of the structural preferences of a molecule before carrying out an ab initio calculation.

Most of the semi-empirical strategies are based on the Zero Differential Overlap approximation (ZDO), which neglects all products of basis functions assuming the overlap to be zero. Over the years, a large number of semi-empirical methods with different choices have been published, including CNDO/2, CNDO/S, INDO, MINDO/3, INDO/S, SINDO1, MSINDO, MNDO, NDOC, AM1, PM3, SAM1, MNDO/d, PM3/tm, and NDDO[31].

The approximations that have been most widely used today are based on MNDO[32], modified neglect of differential overlap which is based on the NDDO (neglect of diatomic differential overlap) method.

The Austin model 1 (AM1)[33] was designed to eliminate the problems of MNDO. The repulsions between atoms separated by distances approximately equal to the sum of their van der Waals radii were overestimated with MNDO. To solve this problem in AM1 Gaussian functions have been used. Both attractive and repulsive Gaussian functions were used. The attractive Gaussian functions were designed to overcome the repulsion and they were applied in the regions where the repulsions were too large. Repulsive Gaussian functions were used at regions with smaller separations between the atoms.

PM3 is a semi-empirical method devised by Stewart and is also based on MNDO[34]. The design of PM3 contains the same elements as AM1. The difference of PM3 with AM1 is that the parameterization for the PM3 method was automated whereas in AM1 the parameterization was obtained by applying chemical knowledge. This has the effect that some parameters have significantly different values in AM1 and PM3, even though both methods use the same functional and the accuracy, for both methods, to predict structural properties is approximately at the same level.

1.10 Force fields

Unlike quantum mechanical approaches, in molecular mechanics the electrons and nuclei of the atoms are not explicitly included in the calculations[35]. Molecular mechanics considers a molecule to be a series of atom types that define the characteristics of an element in a specific environment. For example, a carbon atom in a carbonyl form is treated differently from a carbon that is bonded with three hydrogens. In force field calculations the electronic structure is not calculated at all. This simplification allows the creation of a fast computational approach that can be applied to molecules of any size.

In force field calculations the total energy is minimized with respect to the atomic coordinates, and consists of a sum of different contributions that compute the deviations from equilibrium bond lengths, angles and torsions plus non-bonded interactions:

$$E_{tot}=E_{str}+E_{bend}+E_{tors}+E_{vdw}+E_{elec}+\ldots \qquad (1.10.1)$$

where 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, and E_{elec} is the electrostatic energy term. The majority of the studies that describe the properties of large peptides have been performed using force field methods due to the fact that these are much less expensive than the quantum mechanics calculations. However in this thesis we will not describe in detail the theory behind force field calculations as our target is to describe the conformational preferences of a peptide using higher and more accurate quantum mechanics methods.

1.11 Conformational analysis and forces that determine peptide structure

A major problem in biophysical chemistry is the identification of the most stable structures of a peptide on the potential energy surface. This problem takes on particular interest because variations of the conformations have an effect on the biological activity of these molecules. Peptide conformations are determined by

different factors. First a peptide conformation is determined by geometric factors such as bond lengths and bond angles of the peptide. Second other factors determining peptide conformations are the characterized hydrogen bonds, as well as the intermolecular interaction when a solvent is treated with the peptide. In addition conformational analysis of peptides can be studied with the application of restrictions to the conformers that are characterized by steric interactions.

1.11.1 Introduction to peptides, internal rotation bonds and angles

A peptide is defined as a molecule that is the product of a reaction between two or more amino acids where the carboxyl group of one amino acid binds to the amino group of another amino acid to form a peptide bond. In order to specify the three-dimensional conformation of a peptide, it is necessary to specify the internal rotation angles or torsion angles.

The torsion angle of a flexible bond of a peptide is specified by four atoms:

Given fixed bond lengths and bond angles, the torsion angle φ describes the relationship between the C-D and A-B bonds. This angle is defined as the angle between the plane specified by bonds B-C and C-D and the plane specified by bonds A-B and B-C. The value $\varphi = 0°$ corresponds to the conformation in which the bonds A-B and C-D are *cis* (figure 2a). When the bonds are *trans* in the torsion angle $\varphi = 180°$ (figure 2b).

Rotations around a B-C single bond will lead to different torsion angles. The torsion angle is frequently called τ (tau). By convention, following the definition proposed by W. Klyne and V. Prelog[36] a positive value of the torsion angle A-B-C-B is assigned to the clockwise rotation of up to 180° necessary to bring the

front atom into an eclipsed position with the rear atom. The definition for the dihedral angles used in our study uses positive value as for torsion angles that had a negative value as for example -120° was taken to be +240°.

When A, B, C and D are main chain atoms (i.e. the carboxylic carbon, C_1; the alpha carbon, C_2 or C_α; and the amide group nitrogen, N), then are three torsion angles along the backbone chain called φ, ψ, and ω.

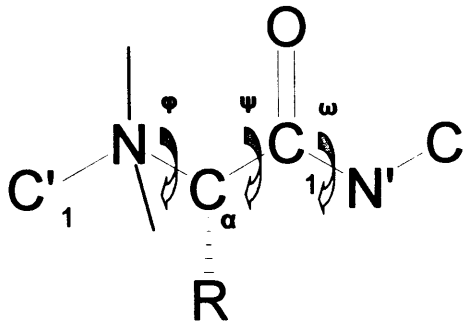


Figure 3. Definition of the position of φ, ψ, and ω torsion angles in a polypeptide chain.

The ω torsion angle characterizes the peptide bond (C_α-C_1-N'-C). The ψ angle defines the bond that is formed between the C_α and the carbon that is involved in the peptide bond (N-C_α-C_1-N'). Finally the φ angle defines the bond between the amide nitrogen and the carbon of the first amino acid (C_1'-N-C_α-C_1).

1.11.2 Hydrogen bonding

A hydrogen bond generally exists between a donor group X-H and an acceptor Y when two molecules associate and the slightly positive hydrogen atom of the donor X is attracted to the slightly negative atom of the acceptor Y. The existence of hydrogen-bonding formation can be observed in a spectroscopic experiment. For example, frequency shifts of infrared and Raman bands can indicate a hydrogen-bonding formation. Results from computational calculations are necessary for complete structural assignments of the spectra.

1.11.3 Hydration of peptides

It is well known that the properties of a biological system (e.g. peptide) when it is treated in a solvent environment can change significantly. Particularly, the interaction of a molecule with water has been the subject of several experimental and theoretical physical chemistry studies[37-39].

1.11.4 Ramachandran or steric contour diagrams

In theory, a peptide chain can adopt an essentially infinite variety of backbone conformations, each corresponding to a unique set of values for the various backbone rotation angles. However, many of these hypothetical conformations can be excluded from consideration on the basis of unfavourable steric overlaps.

The best known conformational distributions in peptides is provided by the Ramachandran map[40], a plot of the allowed values of the torsion angles, φ (the N-C_α bond) and ψ (the C_α-C bond), of the main chain of the polypeptide.

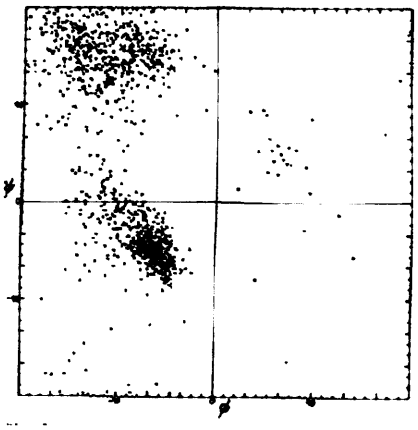


Figure 4.1

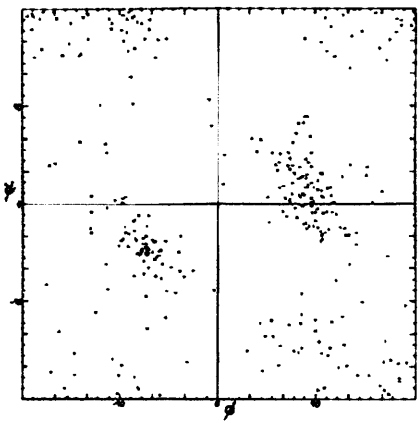


Figure 4.2

Figure 4.1 Plot of φ and ψ angles experimentally determined for approximately 1000 non-glycine residues in eight proteins whose structures have been refined at high resolution taken from reference 30.

Figure 4.2 Plot of dihedral angles φ and ψ experimentally determined for the glycine residue in 20 high-resolution protein structures taken from reference 30.

Figure 4.2 shows the distribution of a glycine residue whereas in figure 4.1 the distribution of all other residues is presented[41]. The dark zones show the main regions where the low-lying conformers are observed, whereas the light zones show the regions where high-lying conformers are likely to be observed.

1.12 Strategies for conformational analysis of peptides and large molecules

The most common problem in the conformational analysis for peptides is their flexibility. Additionally the number of possible conformations is increased as the peptide chain gets larger. Due to computational constraints it is impossible to consider all of the possible structures of the peptide. This increases the risk to miss low-lying conformers. Therefore, one must simplify the system for a detailed study. The essential strategy that is commonly used involves the isolation of a part of the structure. With this an aspect of the molecule can be analyzed and understood. In order to carry out this strategy a few basic ways have been proposed: The use of a smaller part of a peptide and the comparison of two systems that are almost identical[42].

1.12.1 Using a smaller system of a peptide

As the chain of a peptide increases it is more difficult to examine in detail the conformational preferences of such a peptide. In order to study the conformational features of a large peptide it is easier to study a small part of this peptide. The results of such a study can be used to explore the energy landscape of the larger system. The structural preferences of the smaller system can be used as an indication of the structural preferences that the larger molecule may have. In this thesis we apply this method in order to explore the conformational features of the pentapeptide enkephalin. For example the conformational preferences of the pentapeptide in the gas phase can be explored gradually by exploring the conformational features of the corresponding dipeptide, tripeptide and tetrapeptide in the gas phase.

1.12.2 Comparing two systems that are almost identical

The comparison of a system with a closely related molecule can be used in order to identify specific structure variables not found from the study of the current molecule. From this comparison, information of structural features of the peptide that remain constant can be isolated. Additionally critical features of the peptide can be identified when structures are compared that are similar but they do not have the same function. In this thesis the results of our studies on peptides will be compared with related structures.

References

1. Smith, D. D.; Hruby, V. J.; Chow, M., *Annu. Rev. Pharmacol. Toxicol.,* **1990,** 30, 501-534.

2. Barrana, P. E.; Polfera, N. C.; Campopiano, D.; Clarke, D. J.; Langridge-Smith, P. R. R.; Langley, R. J.; Govan, J. R. W.; Maxwell, A.; Dorin, J. R.; Millard, R. P.; Bowerse, M. T., *Int. J. Mass Spectr.* **2005,** 240, 273-284.

3. Jining, L.; Makagiansar, I.; Yusuf-Makagiansar, H. V.; Chow, T. K.; Siahaan, T. J.; Jois, S. D. S., *Eur. J. Biochem.,* **2004,** 271, 2873-2886.

4. Rao, S.; Bhola, N.; Dhawan, M. D., *Opioid Peptides: An Update*. National Institute on Drug Abuse, 1988; p 10-19.

5. Preecea, N. E.; Nguyena, M.; Mahataa, M.; Mahataa, S. K.; Mahapatraa, N. R.; Tsigelny, I.; O'Connora, D. T., *Regulatory Peptides.,* **2004,** 118., 75–87.

6. Bolla, M. L.; Azevedo, E. V.; Smith, J. M.; Taylor, R. E.; Ranjit, D. K.; Segall, A. M.; McAlpine, S. R., *Org. Lett.,* **2003,** 5, 109-112.

7. Lewars, E., *Computational Chemistry:Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. 2003.

8. Jensen, F., *Introduction to Computational Chemistry*. 1999.

9. Leach, A. R., *Molecular Modelling, Principles & Applications, 2nd ed.* 2001.

10. Doll, K.; Saunders, V. R.; M., H. N., *J. Quantum Chem.* **2001,** 82, 1.

11. Lucchese, R. P.; Conrad, M. P.; Schaefer, J., *J. Chem. Phys.,* **1978,** 68, 5292.

12. Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J., *J. Chem. Phys.,* **1988,** 89, 2193.

13. Møller, C.; Plesset, M. S., *Phys. Rev.,* **1934,** 46, 618.

14. Suhai, S.; Ladik, J., *J. Phys. C: Solid State Phys.,* **1982,** 15, 4327.

15. Seel, M.; Bagus, P. S.; Ladik, J., *J. Chem. Phys.,* **1982,** 77, 3123.

16. Cimiraglia, R.; Resta, R., *Int. J. Quant. Chem.,* **1981,** 19, 301.

17. Foresman, J. B.; Frisch, A., *Exploring Chemistry with electronic Structure Theory methods*. Second ed.; 1996.

18. Atkins, P. W., *Quanta,A Handbook of Concepts*. Second edition ed.; Oxford, 1991.

19. Dannenberg, J. J., *J. Phys. Chem. A* **1999,** 103, (11), 1640-1643.

20. Kohn, W.; Becke, A. D.; Parr, R. G., *J. Phys. Chem.* **1996,** 100, 12974.

21. Kohn, W.; Sham, L. J., *Phys. Rev.,* **1965,** 140, A1133.

22. Becke, A. D., *Phys. Rev. A* **1988,** 38, 3098.

23. Lee, C.; Yang, W.; Parr, R. G., *Phys. Rev. B.,* **1988,** 37, 785.

24. Becke, A. D., *Chem. Phys.,* **1993.,** 98, 5648-5662.

25. Jensen, F., *Chem. Phys. Lett.* **1999,** 261, 633-636.

26. Mayer, I.; Surjan, P. R., *Int. J. Quantum Chem.* **1989,** 36, 225.

27. Almlöf, J.; J., T. P. R., *J. Chem. Phys.* **1987,** 86, 553.

28. Boys, S. F.; Bernardi, F., *Mol. Phys.* **1970,** 19, 553.

29. Van Mourik, T.; Wilson, A. K.; Peterson, K. A.; Woon, D. E.; Dunning, T. H., *Adv. Quant. Chem.* **1999,** 31, 105-133.

30. Kulkarni, K. S.; Sarma, C. R., *J. Phys. B:Molec. Phys.* **1973,** 6, 2377.

31. Stewart, J. J. P., *Rev. Comput. Chem.,* **1990,** 1, 45.

32. Dewar, M. J. S.; Thiel, W., *J. Am. Chem. Soc.,* **1977,** 99, 4899.

33. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart., J. J. P., *J. Am. Chem. Soc.,* **1985,** 107, 3902.

34. Stewart, J. J. P., *J. Comput. Chem.,* **1989,** 10, 209.

35. Höltje, H.-D.; Wolfgang Sippl; Didier Rognan; Folkers, G., *Molecular Modeling. Basic Principles and Applications*. 2nd ed.; 2003.

36. Klyne, W.; Prelog, V., *Experientia* **1960,** 16, 521-523.

37. Robertson, E. G.; Simons, J. P., *Phys. Chem. Chem. Phys.,* **2001,** 3, 1.

38. Zwier, T. S., *J. Phys. Chem. A,* **2001,** 105., 8827.

39. Macleod, N. A.; Robertson, E. G.; Simons, J. P., *J. Chem. Phys.,* **2002,** 283, 221.

40. Ramachandran, G. N.; Ramakrishman, C. J., *J. Mol. Biol.,* **1963,** 7, 95.

41. Branden, C.; Tooze, J., *Introduction to protein structure,2nd ed.* 1999.

42. Cantor, C. R.; Schimmel, P. R., *The conformation of biological macromolecules,Part 1.* 1980.

2

INTRODUCTION TO OPIOID PEPTIDES

2.1 Introduction

Among the most common central nervous system neurotransmitters, acetylcholine, norepinephrine, serotonin and dopamine have been known for a long time. During the past years amino acids such as glycine, glutamate and γ-aminobutyric acid have been verified as major neurotransmitters[1]. More recently other peptides have become known as possible neurotransmitters or neuromodulators in the central nervous system. In particular attention has been focused upon the opioid peptides. The term "opioid peptides" has been usually applied to naturally occurring peptides with opiate-like biological properties. Known members of the group of opioid peptides can be divided into five main groups:

- The two pentapeptides methioninenkephalin (Met-enkephalin) and leucinenkephalin (Leu-enkephalin)[2].
- Peptides related to enkephalin such as its analogues. This category includes peptides arising from adrenal proenkephalin[3] such as Met-Enkephalinyl-Arg-Phe.
- β-Endorphin and the related α- γ- and δ-endorphins[4, 5]. It should be noted that the terms "endorphin" and "opioid" peptide have often been used synonymously.
- Peptides present in body fluids e.g. anodynin (in blood)[6].
- Various other peptides whose opiate-like properties do not arise from direct interaction with opiate or opiate-like receptors, e.g. kyotorphin[7].

2.2 Properties of enkephalins

Enkephalins are known as morphine-like neurotransmitters and they are found in the brain, in the limbic system and in the spinal cord. Enkephalins were first isolated in 1975 from pig brain[6]. From the isolation of enkephalin was also found that enkephalin is a mixture of two pentapeptides, Tyr-Gly-Gly-Phe-Leu and Tyr-

Gly-Gly-Phe-Met, i.e. Leu-enkephalin and Met-enkephalin. Enkephalin and its fragment Tyr-Gly-Gly-Phe bind much more strongly than morphine to their common target, the opiate receptor. The extreme flexibility of enkephalin compared with morphine, necessitates both experimental and theoretical studies to locate possible electronic structures that could help structure-activity studies.

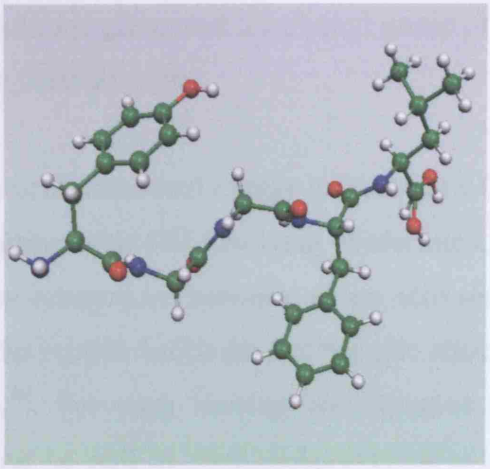


Figure 5. The brain peptide leu-enkephalin (Tyr-Gly-Gly-Phe-Leu)

During the last few years the conformation of enkephalin has been studied more extensively than that of any other biologically active peptide[8-20]. These studies provided information about the conformation(s) of the two pentapeptides in different states: In the isolated state (theoretical calculations), in the crystal form (X-ray analysis), in solution (spectroscopic techniques) and the receptor-bound state (characterization of conformational analogues).

2.3 Conformational studies of enkephalins

Isogai et al. performed a conformational analysis of Met-enkephalin in the gas phase with unchanged NH_2 and COOH-terminal groups and without taking into account solvent effects using the Empirical Conformational Energy Program for Peptides (ECEPP)[21]. With the ECEPP[22] algorithm the geometry of the amino acid residues and the potential functions for inter-atomic interactions were characterized using a set of internally consistent and standard parameters. The most stable conformer resulting from this study is stabilized by a hydrogen bond

between the tyrosyl hydroxyl group and the carbonyl group of Gly (3). This conformer was compatible with NMR data obtained in aqueous solution. Thus it was concluded by the authors that this structure might represent the preferred conformation of Met-enkephalin in solution. However, the results of tyrosine fluorescence quantum yield measurements performed with Met-enkephalin in water and butanol solution indicated[23] that conformers with a hydrogen bond between the tyrosyl hydroxyl group and a carbonyl group of the peptide backbone are not predominant in these solvents.

De Coen performed a conformational energy calculation of the zwitterionic form of Met-enkephalin starting with 400 low-lying conformers, which were obtained by not considering any interactions between amino acid residues and by varying the torsion angles of the peptide backbone and the side chains by step sizes of 20° and 30°, respectively[24]. For each starting conformation, the energies of the available conformations as well as their probability of existence at 25° C were then calculated using a strategy based on semi-empirical averaged potential functions for the energy calculations. Interestingly, only 15 out of 400 conformers examined accounted for 97% of the total probability. Among these, both extended structures and folded conformations, including various β-turns centered on Gly (2)-Gly (3) and Gly (3)-Phe (4), were found.

Premilat and Maigret used the Monte Carlo method to obtain statistical samples of conformations of Leu-enkephalin[25]. For the zwitterionic form of the peptide, an end-to-end distance distribution with a peak of 5 Å was obtained indicating a predominance of folded conformations stabilized by ionic interactions between the terminal NH_3^+ and COO^- groups. In a subsequent study a clustering analysis was performed, the target of this study was to reduce the statistical Monte Carlo sample into a limited number of conformational groups[26]. From this analysis 5 different groups were investigated in the zwitterionic form and the results obtained showed improved agreement with the experimental values. Thus the authors suggested that the behaviour of enkephalin conformers in solution may be represented by a limited number of structural classes of enkephalin.

Perez et al. performed a study[27] to characterize the low-energy conformational features for Met-enkephalin for the 92 low-lying conformers created using the ECEPP potential investigated by Isogai[21]. These structures were minimized using the CHARMM force field potential. The most stable conformer found is characterized by a β-type conformer and contains an OH---O interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of phenylalanine (4). The global minimum obtained from this study was one of the most stable conformers found from a study performed by Koča[28] where the conformational potential energy of Met-enkephalin was analyzed by a molecular mechanics method in combination with the programs DAISY[29] and ROSE[30]. The program ROSE was used to automatically create all possible conformations of the peptide. The conformational search is based on combinatorial composition of conformations of parts of the molecule, the so-called fragments and pseudofragments. The program has been useful for conformational analysis of molecules with known conformations of fragments. The DAISY program was used to obtain information on the potential energy surface of the peptide. First all possible interconversions between conformers were predicted. Second the energy computations were performed based on molecular mechanics calculations. Furthermore, Wales and Evans[31] performed a study to describe the energy landscape and dynamics of Met-enkephalin based on a discrete path sampling method. Through this method the time scales for motion between different regions and in particular predicting that the peptide can fold from extended conformations to a compact low-energy β-type conformer, were calculated. The β-type conformer was shown to have a significantly higher rate during folding than the extended conformers.

The results of a computational study exploring the conformational features of linear and cyclic enkephalins have been published by Kříž et al[32]. In this study the software package CIDADA[33] was used for the prediction of the low-energy minima on the potential energy surface using calculations based on molecular mechanics. From this study the tyrosine α-amino group and the aromatic properties of the tyrosine and phenylalanine residues are suggested to be the key elements essential for the particular bioactivity of the enkephalin molecule. The extended conformations of open chain structures of the enkephalins are preferred

in the X-ray structures while folded conformations are preferred in NMR solution data.

Jalkanen has published a study on the energetics of the Leu-enkephalin structures with vibrational frequencies, vibrational absorption, and vibrational circular dichroism and Raman intensities[34]. In this study the relative conformational energies of selected low-energy structures of Leu-enkephalin starting from either X-ray determined structures of all non-hydrogen atoms or from six different model structures of Leu-enkephalin were computed. For the resulting 30 different structures geometry optimizations were performed at the SCC-DFTB level with and without dispersion correction, and at the B3LYP/6-31+G*, AM1, PM3, and MM levels. The SCC-DFTB method[35] arises from density functional theory (DFT) by a second-order expansion of the DFT total energy functional and includes a pairwise dispersion for the description of long-range dispersion interactions. Only the B3LYP/6-31+G* results were discussed as they were thought to be the most accurate as B3LYP does not involve any empirical corrections. The most stable conformers resulting from the B3LYP/6-31+G* optimizations are stabilized by a hydrogen bond between the oxygen of the carbonyl C=O group of the tyrosine residue (1) and the hydrogen of the NH amide group of the phenylalanine residue. Another hydrogen bond is observed in this conformer between the oxygen of the carbonyl C=O group of glycine (3) and the hydrogen of the Leu-enkephalin residue. In all the most stable conformers obtained for Leu-enkephalin no interaction is observed for the hydroxyl group of the tyrosine (1) residue. From the comparison of the spectroscopic studies for this molecule, the group modes for the amide groups, the N-terminus and the C-terminus are very intense in the IR spectrum.

The quantum states of Leu-enkephalin conformations were described by Abdali et al[36]. In this study the barrier for the conformational transition between the single bend structure, which is characterized by an interaction between the C=O carboxyl of Phe (4) and the hydrogen amide of Gly (2), and the extended conformer was estimated to be about 12 kcal mol^{-1}. Furthermore molecular dynamics simulations were carried out using a double β-bend structure, which is characterized by an interaction between the carboxyl group of Leu (5) and the

hydrogen amide of Gly (3), as the starting conformation. The simulation showed that a single bend structure forms after 3-4 ns and remains stable up to 10 ns. For the conversion of the double-bend to a single-bend in the first 2 ns an extended structure is formed. Once the molecule establishes three to four hydrogen bonds, the single-bend backbone structure is energetically very favourable. Furthermore, it is clear from the simulation that Leu-enkephalin goes through several different conformations including an extended conformation, before relaxing to the final single-bend conformation. In a subsequent study three Leu-enkephalin conformers provided by X-ray crystallography were optimized using DFT at the B3LYP/6-31G* level. VA (vibrational absorption) and VCD (vibrational circular dichroism) measurements were also performed for Leu-enkephalin in DMSO solution[37]. A comparison of the VA spectrum and the B3LYP calculations suggest that a single β-bend conformation is present in this solution. However, the calculations found that different single β-bend conformations provide different low-energy structures. Therefore, it was necessary to carry out a VCD spectroscopy study for the lowest structures, due to the sensitivity of the VCD to the conformational changes in the molecule. A comparison between the measured and calculated VCD spectra determined which of the many single β-bend conformations is most likely present in such a solution.

In a subsequent study a conformational analysis at the B3LYP and the SCC-DFTB level of theory and a comparison of the results with vibration absorption spectra of Leu-enkephalin was described[38]. The SCC-DFTB optimized structures with dispersion correction provided a better fit to the measured spectrum of Leu-enkephalin than SCC-DFTB without dispersion correction when both frequency and intensity are taken into account. The calculated DFT spectra provide more features and a better fit when all regions of the measured spectrum are taken into account. In the case of SCC-DFTB and DFT calculations a single β-bend conformer that has an interaction between the C=O carboxyl of Phe (4) and the hydrogen amide of Gly (2) is suggested as the most stable structure, whereas a double β-bend structure that has an interaction between the carboxyl group of Leu (5) and the hydrogen amide of Gly (3), is suggested by calculations using SCC-DFTB with dispersion included.

A vibrational analysis of capped Leu-enkephalin has been previously conducted by Hirst and Watson[39]. Structure generation and initial optimizations for the peptide were performed using the CHARMM[40] molecular dynamics package with the CHARMM22[41] force field. The low–lying conformers resulting from the force field optimizations were further optimized at the EDF1/6-31+G* level. EDF1 is an empirical exchange-correlation functional which uses a set of experimental data[42]. The lowest-energy conformer found has a single-bend structure which is characterized by two main interactions. The primary interaction that the most stable conformer exhibits is a hydrogen bond between tyrosine (1) and the carboxyl oxygen of glycine (3). The second interaction is a hydrogen-bonding interaction between the backbone NH of the peptide and the aromatic ring of tyrosine (NH-π). Also in the most stable conformers the backbone N-H is pointed into the centre of the tyrosine ring, even when the tyrosine hydroxyl is too distant from the main chain to form a hydrogen bond. In a subsequent theoretical study the amide I mode of arbitrary polypeptides of Leu-enkephalin were computed using vibrational frequency calculations[43]. In this study the low-energy conformers of the pentapeptide were identified at the EDF1/6-31+G* level and then the vibrational frequencies were calculated directly from DFT calculations and from a simple oscillator model calculation. The model calculation was parameterized using small-molecule DFT data. From this study the comparison between amide I bands and model bands noted the importance of the side-chain interactions, especially when they are hydrogen-bonded to the backbone (for e.g. tyrosine). Another conclusion extracted from this study was that by performing DFT calculations with the model developed in this study some of these effects might be captured but at a great initial computational cost.

Gautham and Vengadesan recently published a study to explore the energy landscape of Met-enkephalin and Leu-enkephalin based on an Orthogonal Latin Squares Sampling [MOLS] method[44]. In the application of this method the ECEPP/3 force field[45] was used. In this study the 1500 low-lying conformers were considered. It was found that the fully-folded conformers are among the lowest-energy conformers whereas fully-extended conformers are located among the high-energy conformers. Partially-folded conformers were intermediate in energy between the extended and folded states. Similar studies[31, 46] showed that folded

structures of the two enkephalin pentapeptides are among the most stable ones whereas extended structures populated the high-energy conformers.

Chandrasekhar and co-workers performed a study on the conformational preferences of Leu-enkephalin in both neutral and zwitterionic form using NMR spectroscopy and Molecular Dynamics calculations at the surface of a hydrated DMPC (Dimyrstoylphosphatidylcholine) bilayer[47]. From this study, the most stable conformation found was a bend structure at Gly (3) where the ψ dihedral angle adopts a staggered conformation. In the presence of an aqueous interface the neutral peptide forms β-turn conformers at residues Gly (2) and Phe (4). The time behaviour of the φ and ψ angles of the five residues of the peptide in neutral and zwitterrionic form in an aqueous solution showed that the large residues Tyr (1), Phe (4) and Leu (5) adopt similar staggered conformations in all simulations whereas the glycine dihedrals showed fluctuations between staggered or skewed conformations.

From the previous studies discussed above some general conclusions can be derived. The results obtained from almost all the studies showed that Tyr (1) plays an important role in the conformational preferences of the peptide. Additionally the structures that seem to be dominant on the conformational energy landscape of the peptide consist of folded conformers in the neutral as well as in the zwitterionic form in the gas phase. Most of the low-lying conformers obtained are characterized by an interaction between the OH of Tyr (1) and either the carboxyl oxygen of Gly (3) or the carboxyl oxygen of Phe (4). For the zwitterionic form of the peptide the predominant conformers are stabilized by an interaction between the terminal NH_3^+ and the COO^- groups. On the other hand, conformational studies in solution for enkephalin showed that conformers with a hydrogen-bonding interaction between the tyrosyl hydroxyl group and the carbonyl group of the peptide backbone are not among the most stable ones. Most of the studies performed to explore the conformational features of enkephalin use methods based on molecular mechanics calculations. Although some of the strategies include DFT geometry optimizations (with the B3LYP functional), these were performed from optimized structures that resulted from force field optimizations. Thus, in order to describe the conformational features of enkephalin, a strategy

that will include calculations only with the more accurate electronic structure theory methods is necessary to provide higher accuracy.

The aim of this study is to explore the conformational preferences of Leu-enkephalin using only electronic structure theory methods. As was mentioned in chapter 1, we will try to use a general method to describe the conformational features of a large system by exploring the conformational features of a smaller system of the peptide. Enkephalin is a very flexible peptide, consisting of up to 20 torsion angles that need to be varied resulting in a very large number of possible combinations. The strategy that will be employed in this thesis is to explore the conformational features of Leu-enkephalin by first exploring the conformational features of its dipeptide Tyr-Gly, tripeptide Tyr-Gly-Gly and its tetrapeptide Tyr-Gly-Gly-Phe. From the conclusions that will result from the exploration of the conformational features of the dipeptide Tyr-Gly, crucial information can be extracted that can be used in the analysis of the tripeptide Tyr-Gly-Gly. From the low-lying conformers of these smaller peptides important information can be extracted to reduce the number of the conformers that need to be considered to explore the potential energy surface of the pentapeptide enkephalin.

References

1. Solomon, H.; Snyder, S.; Steven, S. R., *Ann. Rev. Neurosci.* **1979,** 2, 35.

2. Hughes, J.; Smith, T. W.; Kosterlitz, H. W.; Fothergill, L. A.; Morgan, B. A.; Morris, H. R., *Nature* **1975,** 258, 577-579.

3. Rossier, J. Y.; Audigier, N.; Ling, J.; Cross, J.; Udenfriend, S., *Nature,* **1980,** 288, 88-90.

4. Bradbury, A. F.; Smyth, D. G.; Snell, C. R.; Birdsall, N. J. M.; Hulme, E. C., *Nature* **1976,** 260, 793-795.

5. Lin, N.; Burgus, R.; Guillemin, R., *Proc. Natl. Acad. Sci.* **1976,** 73, 3942-3946.

6. Pert, C. B.; Pert, A.; Tallman, J. F., *Proc. Natl. Acad. Sci.* **1976,** 73, 2226-2230.

7. Takagi, H.; Shiomi, H.; Ueda, H.; Amano, H., *Nature* **1979,** 282, 410-412.

8. Perez, J.; Villar, H.; Uyeno, E.; Toll, L.; Olsen, C.; Polgar, W.; Loew, G. H., *Int. J. Quantum Chem., Quantum Biol. Symp.* **1993,** 20, 147-160.

9. Perez, J.; Villar, H.; Loew, G. H., *Int. J. Quantum Chem.* **1992,** 44, 263-275.

10. Zaman, M. H.; Shen, M. Y.; Berry, R. S., *J. Phys. Chem. B.* **2003,** 118, (11), 5143-5156.

11. Loew, G. H.; Cometta-Morini, C.; Perez, J.; H., V., *Int. J. Quantum Chem., Quantum Biol. Symp.* **1991,** 18, 165-181.

12. D'Alagni, M.; Delfini, M.; Di Nola, A.; Eisenberg, M.; Paci, M.; Roda, L. G.; Veglia, G., *Eur. J. Biochem.* **1996,** 240, 540-549.

13. Janecka, A.; Kruszynski, R., *Current Medicinal Chemistry* **2005,** 12, (4), 471-481.

14. Blomberg, D.; Hedenstrom, M.; Kreye, P., *J. Org. Chem.* **2004,** 69, (10), 3500-3580.

15. Malicka, J.; Czaplewski, C.; Groth, M., *Current topics in Medicinal Chemistry* **2004,** 4, (1), 123-133.

16. Unruh, J. R.; Johnson, C. K., *Biophys. J.* **2003,** 84, (2), 289.

17. Shen, M. Y.; Freed, K. F., *J. Chem. Phys.* **2003,** 118, (11), 5143-5156.

18. Amodeo, P.; Naider, F.; Picone, D., *J. Pep. Sci.* **1998,** 4, (4), 253-265.

19. Schwonek, J.; Sanders, C., *Biophys. J.* **1993,** 64, (2), 288.

20. Nayeem, A.; Vila, J.; Scheraga, H., *J. Comp. Chem.* **1991,** 12, (5), 594-605.

21. Isogai, Y.; Nemethy, G.; Scheraga, H., *Proc. Natl. Acad. Sci.* **1977,** 74, (2), 414-418.

22. Momany, F. A.; McGuire, R. F.; Burgess, A. W., *J Phys. Chem.* **1975,** 79, (22), 2361-2381.

23. Schiller, P. W.; Yam, C. F.; Lis, M., *Biochemistry* **1977,** 16, 1831-1838.

24. De Coen, J.; Humblet, C.; Koch, M., *FEBS Lett.* **1977,** 73, 38-42.

25. Premilat, S.; Maigret, B., *J Phys. Chem.* **1980,** 84, 293-299.

26. Maigret, B.; Premilat, S., *Biochem. Biophys. Res. Commun.* **1982,** 104, 971-976.

27. Perez, J.; Villar, H.; Loew, G. H., *J. Comput. Aided Mol. Des.* **1992,** 6, (2), 175-190.

28. Koca, J.; Carlsen, P. H. J., *J. Mol. Struct. (Theochem)* **1995,** 337, 17-24.

29. Koca, J.; Carlsen, P. H. J., *J. Mol. Struct. (Theochem)* **1992,** 257, 105-130.

30. Koca, J.; Carlsen, P. H. J., *J. Mol. Struct. (Theochem)* **1992,** 257, 131-141.

31. Wales, D.; Evans, D., *J. Chem. Phys.* **2003,** 119, 9947-9955.

32. Kriz, Z.; Koca, J.; Carlsen, P. H. J., *J. Mol. Struct. (Theochem)* **2001,** 540, 231-250.

33. Koca, J., *J. Mol. Struct. (Theochem)* **1994,** 308, 13-24.

34. Jalkanen, K. J., *J. Phys. Condens. Matter* **2003,** 15, 1823-1851.

35. Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E., *J. Chem. Phys.* **2001,** 114, (12), 5149-5155.

36. Abdali, S.; Jensen, M.; Bohr, H., *J.Phys. Condens. Matter* **2003,** 15, 1853-1860.

37. Adbali, S.; Jalkanen, K. J.; Cao, X.; Nafie, L. A.; Bohr, H., *Phys. Chem. Chem. Phys.* **2004,** 6, 2434-2439.

38. Abdali, S.; Niehus, T. A.; Jalkanen, K. J.; Cao, X.; Nafie, L. A.; Frauenheim, T.; Suhai, S.; Bohr, H., *Phys. Chem. Chem. Phys.* **2003,** 5, 1295-1300.

39. Hirst, J. D.; Watson, T. M., *Phys.Chem.Chem.Phys.* **2004,** 6, 2580-2587.

40. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A., *J. Comp. Chem.* **1986,** 7, 230.

41. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J., *J. Comp. Chem.* **1983,** 4, 187.

42. Adamson, R. D.; Gill, P. M. W.; Pople, J. A., *Chem. Phys. Lett.* **1998,** 284, 6-11.

43. Hirst, J. D.; Watson, T. M., *Mol. Phys.* **2005,** 103, 1531-1546.

44. Gautham, N.; Vengadesan, K., *J. Phys. Chem. B* **2004,** 108, (30), 11196-11205.

45. Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scherag, H. A., *J.Phys. Chem.* **1992,** 96, 6472.

46. Sanbonmatsu, K. Y.; Garcia, A. E., *Proteins* **2002,** 46, 225.

47. Chandrasekhar, I.; van Gunsteren, W. F.; Zandomeneghi, G.; Williamson, P. T. F.; Meier, B. H., *J. Am. Chem. Soc* **2006,** 128, (1), 159-170.

3

THE STRUCTURE OF THE TYR-GLY DIPEPTIDE

3.1 Introduction

Theoretical calculations on gas-phase peptide structures are an invaluable complement to experiments: whereas infrared (IR) spectroscopy studies can identify structural features occurring in the observed conformers, comparison with computed data on the energetics and vibrational frequencies is crucial for conversion of the observed spectra into structural assignments. One major problem with computational studies of peptides is their size and flexibility, which makes it difficult to use high-accuracy electronic structure methods. The computational methods that have been most often applied include low-cost methods such us force field molecular mechanics. However, a truly quantitive prediction of peptide energetics using force fields is not yet available[1]. The aim of the current work is to develop an efficient method which uses electronic structure theory methods, would consider all possible conformations and locates all low-lying conformers of small peptides.

A wide range of biological molecules in the gas phase have been characterized with a combination of experimental and theoretical studies including the neurotransmitters tryptamine, serotonin, adrenaline and noradrenaline[2-5]. However not many studies are present studying dipeptides in the gas-phase. Dipeptides are the smallest chains that contain a peptide bond and thus they are the simplest systems to explore basic peptide properties. Spectroscopic and computational studies on dipeptides in the gas phase provide the opportunity to study the individual properties of such peptides free of solvent environment.

Recent studies have shown that just a few conformers were observed from experiments although many possible conformational combinations were predicted by theoretical studies[4, 6, 7]. For example, from the spectroscopic studies performed for the peptides Trp-Oet, Gly-Trp, Trp-Ser and Pro-Trp (Trp=tryptophan;

Oet=ethylester; Ser=serine; Pro=proline) the number of the conformers observed is surprisingly small[7]. Ab initio calculations have been introduced to provide all structural possibilities of these conformers. The spectroscopic experiments show which of these conformers are actually observed[8].

Most of the ab initio studies have been performed to investigate the potential energy surface of the peptide in terms of the torsional angles characterizing the conformation around the peptide bond, the so called Ramachandran angles[9]. Keefe and Pearson performed a study to investigate how other geometrical parameters can affect the energy and structure of peptides[10]. A conclusion extracted from this study was that the bond lengths and angles of the amide plane are fixed for various dipeptides. This means that in large peptide systems the amide-plane parameters do not need to be considered in geometry optimizations.

A number of spectroscopic studies have been performed on the three aromatic amino acids phenylalanine (Phe)[11, 12], tyrosine (Tyr)[13, 14] and tryptophan[15, 16] and peptides containing these aromatic amino acids[17, 18]. Most of the studies that have been performed for the peptides are using chemically protected (capped) groups (HC=O-NH N-terminus; C=O-OCH_3, C=O-NH_2 or C=O-$NHCH_2$ C-terminus)[19-24]. The IR spectrum of uncapped Trp-Gly, Gly-Trp and Trp-Gly-Gly in the gas phase has recently been published by Hünig and Kleinermanns[25]. In a subsequent study Bakker et al. published a study presenting mid-infrared spectra of these peptides[26]. Cohen et al. have published a study based on resonance-enhanced multiphoton ionization (REMPI) spectroscopy of jet-cooled dipeptides containing either tyrosine or phenylalanine as the aromatic chromophore[27].

In the current study we take into consideration all possible conformers of the dipeptide Tyr-Gly since experimental gas phase data are not available for this peptide. This results in a large number of low-lying conformers. In order to explore the conformational landscape of Tyr-Gly we employ a sequential stepwise rotation method to locate the most stable conformers. For the stepwise rotation method, large step sizes were employed increasing the risk to miss some low-lying conformers. For this reason, a hierarchical selection method was developed that considers all the possible conformations of the peptide and hierarchically

increases the electronic structure level of theory used. The results of the current study have been published in Molecular Physics[28].

3.2 Methodology

The conformational landscape of the Tyr-Gly dipeptide was explored using two different methodologies. The first method employed was a "stepwise rotation" procedure, whereas the second method uses a "hierarchical selection scheme". Figure 6a shows the flexible bonds of the peptide that were considered for the two different procedures. Figure 6b shows the atomic labelling of the peptide with the location of the Φ and Ψ angles. All the electronic structure calculations were performed with the software package Gaussian (versions 98[29] and 03[30]) on clusters of 1.7 and 2.8 GHz Pentium 4 and dual Xeon PCs running Linux at the Chemistry Department at University College London (UCL), as well as on a cluster of 900 MHz Sunfire V880 servers, at the HiPerSPACE Computing Centre at UCL.

3.2.1 Stepwise rotation of internal bonds

The conformational analysis of Tyr-Gly was performed via a sequential variation of the dihedral angles labelled (A-I) shown in Figure 6a. The initial structure of Tyr-Gly was built using Molden[31] with Ramachandran angles[32] φ and ψ (with -119° and 113° respectively). The conformational analysis of the Tyr-Gly dipeptide was performed with the systematic variation of the nine dihedral angles through six steps. The first two dihedral angles A and B were varied simultaneously. The resulting structures were optimized using the HF/6-31+G* level of theory and subsequently re-optimized at the Density Functional Theory (DFT) level using the B3LYP functional[33] and the 6-31+G* level of theory. Next dihedrals C-H were varied sequentially through six steps. In each step, the conformers created from optimizations at the HF/6-31+G* level of theory were re-optimized with B3LYP/6-31+G*. All optimized conformers were considered in each step except in step G: from the 77 conformers resulting from variation of dihedral G only 57 (with relative energies within 30 kJ/mol of that of the most stable conformer) were taken to next step (variation of dihedral H). The initial values of dihedrals A, B, C, D were set to 0, 90, 180, 270°, whereas the initial values of dihedrals F, G and H were set to 0, 120, 240°. The stepwise rotation method yielded 95 different Tyr-Gly conformers.

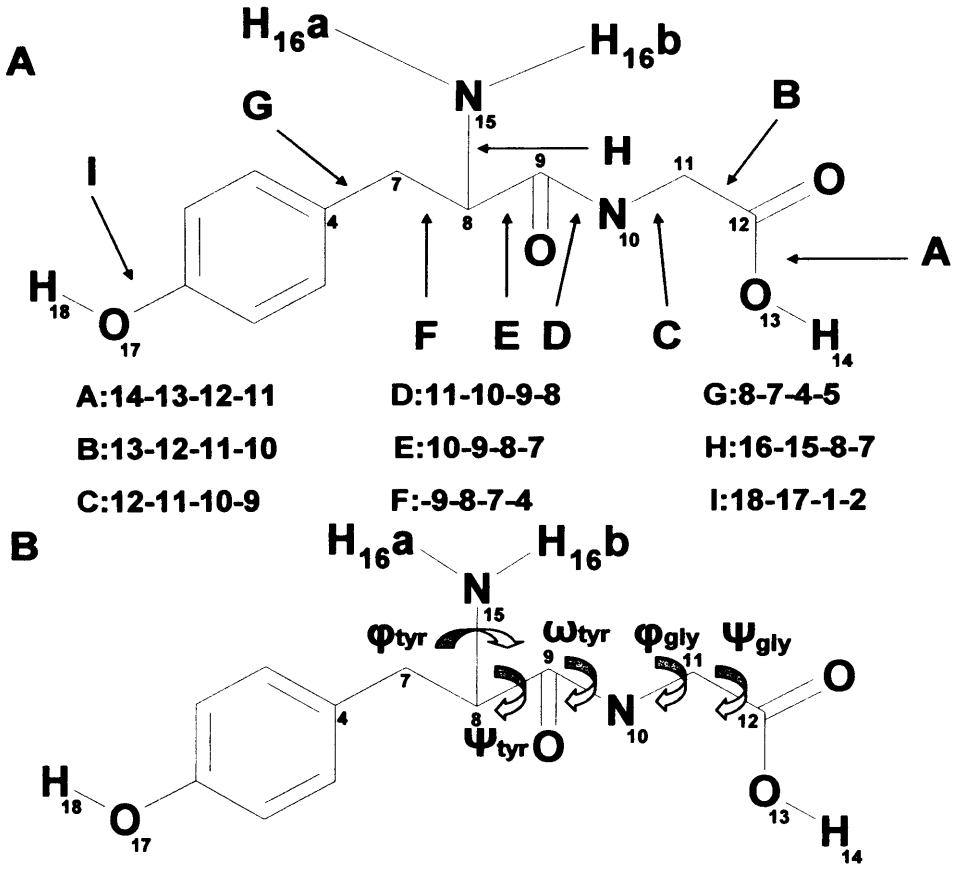


Figure 6. A. Atom labeling and definition of the dihedral angles considered in the hierarchical selection procedure.

B. Definition of the dihedral angles ω, φ, ψ.

The hydroxyl OH-group has two possible orientations (values of 0 and 180° for dihedral I). The variations of dihedral I should in principle not be necessary, as turning over the OH group has the same effect as flipping the entire aromatic ring, and this was already considered by variation of dihedral G. However, since we discarded conformers we considered it necessary to investigate hydroxyl flipping in the most stable conformers. Using the 20 most stable conformers, new structures were obtained by flipping the hydroxyl group and setting the initial value of dihedral I to 180°. B3LYP/6-31+G* optimizations of the new 20 structures yielded four new conformers. Thus, in total, the stepwise rotation procedure yielded 99 Tyr-Gly conformers.

3.2.2 Hierarchical selection scheme

In the stepwise rotation procedure large step sizes (90° and 120°) were used for rotation around the flexible bonds in order to keep the conformers that need to be optimized manageable, and thus, it is possible that low-lying conformers were missed. In this section we outline an alternative method to locate the most stable conformers of short peptides. This method is based on a hierarchical selection approach: first, all possible conformers are created and sorted according to their number of intramolecular hydrogen-bonding interactions; next at increasingly more accurate levels of theory, conformers are sorted according to their single point or optimized energy. Only the most stable conformers are taken through to the next level of calculation. We constructed a computer program that creates Tyr-Gly structures by varying dihedrals A-H; all possible combinations of these dihedrals were taken into account. The starting structure was the same as that employed in the stepwise rotation, except that the initial values of dihedrals A-H were replaced with those from the most stable conformer resulting from the stepwise rotation. We found it necessary to employ small step sizes for some dihedrals in order not to miss any low-lying conformers. Thus dihedrals A, D and E were varied with a step size 180°, dihedral H was varied with a step size of 120°, B and G were varied with a step size of 60° , and C and F were varied with a step size of 30°. Structures with overlapping atoms were discarded, which resulted in a total of 44344 conformers. The program subsequently identifies the hydrogen-bonding interactions in all structures, and groups the conformers according their number of hydrogen bonds. An X-H-Y atom group was considered to form a hydrogen-bond interaction if $R_{xy} \leq 2.6$ Å, (X-H-Y) $\geq 100°$, and Y$\neq$ (C or H). Next single-point energy calculations were performed using HF/3-21G*. For the most stable structures based on the single-point energies, geometry optimizations were performed using HF/3-21G*. Furthermore, the most stable conformers based on the optimized structures at the HF/3-21G* were optimized using B3LYP/6-31+G*.

The most stable structures obtained from the stepwise rotation form a characteristic O-H•••O ring hydrogen-bonding interaction (see below). However, single-point energy calculations underestimated the energy of these structures (as

indicated by the large increase in stability after HF geometry optimization), with the result that they ranked too low on the list of the most stable structures based on the HF/3-21G* single-point energies. As only the most stable conformers on this list are taken to the next step of the hierarchical selection, this could mean that important conformers are missed. Single-point energy calculations underestimate the energies of the O-H•••O ring structures because significant geometric relaxation is required to form this particular hydrogen-bonding motif and thus geometry optimizations are essential to obtain an accurate energy ranking of these structures. As it was too computationally expensive to perform HF/3-21G* geometry optimizations for all the O-H•••O ring conformers we optimized these structures using the semi-empirical PM3 method[34]. Test calculations showed that HF and PM3 give similar structures for these conformers. The first 200 conformers found from the HF/3-21G* single-point energy calculations, as well as the first 100 most stable conformers found from the PM3 optimizations, were optimized using the HF/3-21G* method. Finally the 50 most stable conformers obtained from these calculations were optimized using the B3LYP/6-31+G* method, which resulted in 49 unique conformers. The relative energies of the twenty most stable conformers according to the B3LYP/6-31+G* geometry optimizations were evaluated by single-point calculations at the second order Møller-Plesset perturbation (MP2) level of theory. Zero-point energies (ZPEs), scaled by 0.976[35], were calculated using B3LYP/6-31+G*. Additionally, the structures of selected conformers were optimized at the MP2/6-31+G* level of theory.

Large and complex systems such as proteins are composed from smaller repeating units, the amino acids. The twenty naturally occurring amino acids that comprise proteins are of the **L- form**[36]. The L and D prefix is given to the formal name of an amino acid with one centre to indicate that the ligand of reference resides either on the right side (D-form) or on the left side (L-form) of the vertical line in its Fischer projection[36]. In our study we constructed our dipeptide (Tyr-Gly) with Molden which automatically creates L-amino acids.

3.2.3 Code specification

The hierarchical selection scheme employed in the current chapter has been implemented in a computer program written in Fortran 90. The computer program has been designed to create all possible conformers of a peptide by rotating selected internal bonds with specified step sizes. This is implemented using a recursive subroutine. This recursive subroutine is structured as:

- Loop over all values of the current torsion angle
- For each value of this torsion there is loop over all other torsions by calling this recursive subroutine.

Structures with overlapping atoms are discarded (X-H $\leq$ 1.2 Å).

Information such as the number of the hydrogen-bonding interactions the conformers contain as well as the atoms that form H-bonding interactions $R_{xy} \leq 2.6$ Å, (X-H-Y) $\geq 100°$, and Y$\neq$ (C or H) are written on a file. Furthermore, the distance between the hydrogen and the H-acceptor (R_{HY}) as well as the X-H-Y angle are computed and written on this file. Finally information on the hydrogen-bonding interactions occurring in the peptide are collected such us the existence of a π-interaction with an aromatic residue. Also the existence of bifurcated H-bonds is being computed and written.

3.3 Results

3.3.1 Stepwise rotation method

Table 1 lists the relative energies obtained for the 20 most stable conformers obtained from the stepwise rotation method. The twenty most stable structures contain two hydrogen-bonding interactions in their structure. From their structures it can be observed that a characteristic hydrogen-bonding interaction occurs between the hydroxyl group of the glycine residue and the carbonyl oxygen of tyrosine (O-H---O). Some of the most stable conformers have a characteristic folded "book" structure (labeled "book" in table 1) resulting from a small (<90°) $C_4C_7C_8C_9$ torsion angle.

Conformers	Rel. Energies (kJ/mol)	φ_{gly} (degrees)	ψ_{tyr}	ω_{tyr}	ψ_{gly}	Structure
1	0.000	-73.69	-16.94	-171.84	57.43	OHO, anti
2	0.211	73.12	7.88	171.00	-58.71	book, OHO, anti
3	0.360	-73.36	11.11	-176.38	-57.45	OHO, anti
4	0.829	-73.27	10.69	-176.43	56.95	book, OHO, anti
5	1.676	-179.04	13.21	176.85	57.17	book, OHO, syn
6	2.358	74.47	12.41	171.09	-179.31	book, anti
7	2.400	-73.10	11.32	-176.58	-57.77	OHO, anti
8	2.448	74.51	12.58	171.21	-57.88	OHO, anti
9	2.451	-73.12	11.26	-176.50	57.62	OHO, syn
10	2.849	-179.44	12.60	177.15	57.58	OHO, syn
11	3.088	108.31	17.45	179.85	-179.34	Book, anti
12	3.090	177.83	12.66	176.85	-58.13	book, OHO, anti
13	3.298	177.45	12.88	176.83	-178.85	anti
14	3.498	106.79	17.65	-175.32	-178.90	syn
15	3.807	106.07	17.98	-175.20	179.73	anti
16	3.906	99.31	-17.06	-173.57	179.72	syn
17	6.892	99.36	-16.96	-173.61	-7.95	anti
18	7.078	-99.32	12.34	173.25	-8.14	anti
19	7.525	108.31	17.45	179.85	7.10	book, anti
20	8.390	104.30	16.49	-179.67	-5.40	book, anti

Table 1. Relative energies and values of φ, ψ, ω obtained for the 20 most stable structures for tyrosine-glycine obtained from the stepwise rotation method.

From the energies obtained for the twenty most stable structures of the dipeptide it can be observed that for most conformers, by flipping the dihedral $H_{18}O_{17}C_6C_1$ by 180° already existing conformers are obtained. For example, variation of dihedral $H_{18}O_{17}C_6C_1$ by 180° in conformer 7 gives the energy of conformer 9. However,

significant differences occur in the relative energies of conformers 1, 2, 4, 5, 6, 10, 11, 12, when the dihedral H_{18}-O_{17}-C_6-C_1 is varied by 180°. In all these structures, Dih4: C_4-C_7-C_8-C_9 is about 100° degrees larger than in the other structures and the side chain is perpendicular to the ring.

Conformers	D18: H18-O17-C6-C1 set to 0° Relative Energies/kJ/mol	D18: H18-O17-C6-C1 set to 180° Relative Energies/kJ/mol
1	0.000	0.203
2	0.211	2.920
3	0.360	0.332
4	0.829	1.676
5	1.676	0.829
6	2.358	3.088
7	2.400	2.451
8	2.448	2.849
9	2.451	2.400
10	2.849	2.448
11	3.088	2.358
12	3.090	0.211
13	3.298	3.498
14	3.498	3.298
15	3.807	3.906
16	3.906	3.807
17	6.892	7.078
18	7.078	6.892
19	7.524	8.579
20	8.390	9.164

Table 2. Relative energies of the 20 most stable conformers of Tyr-Gly varying dihedral D18: H18-O16-C6-C1 by 180 degrees according to the most stable conformers.

From the stepwise rotation method we can conclude that rotation of the hydroxyl group of the tyrosine residue changes the stability of the conformer. Conformers for which the hydroxyl group of the tyrosine residue of the dipeptide is lying towards the hydroxyl group of the glycine residue appear to be more stable than the corresponding conformer where dihedral $H_{18}O_{17}C_6C_1$ is rotated by 180°. These changes in stability may be explained by the long-distance interaction between the hydroxyl oxygen of the tyrosine residue and the carboxyl oxygen of the glycine residue.

3.3.2 Hierarchical selection method

From the stepwise rotation the risk to miss low-lying conformers was increased by employing large step sizes to rotate around the flexible bonds of the peptide. Thus, the hierarchical selection scheme was employed that considers all possible conformers of the peptide. Table 3 lists the relative energies obtained for the 20 most stable conformers obtained from the hierarchical selection method. In comparison with table 1 we can observe that some of the conformers obtained are different from those obtained with the stepwise rotation. This due to the fact that the hierarchical selection method found all low-lying conformers predicted by the stepwise rotation, and located several additional conformers not found with the stepwise rotation. The hierarchical selection scheme therefore appears to be the more reliable method for locating the most stable conformers of a peptide. For this reason we consider only the 20 most stable conformers of Tyr-Gly resulting from the hierarchical selection method in the rest of this chapter. For these conformers we performed single point energy calculations using the MP2/6-31+G* level of theory.

Conformers	Rel. Energies (kJ/mol)	Φ_{tyr}	Ψ_{tyr}	ω_{tyr}	Ψ_{gly}	Structure
1	0.000	-73.69	-16.97	-171.85	57.43	OHO, anti
2	0.183	73.11	7.90	171.01	-58.71	book, OHO, anti
3	0.193	-73.72	-17.01	-171.83	57.45	OHO, anti
4	0.326	73.28	-16.73	176.00	-57.5	OHO, syn
5	0.333	73.22	-17.07	175.89	-57.45	OHO, anti
6	0.822	-73.26	11.11	-176.39	56.95	book, OHO, anti
7	1.374	-176.08	-17.82	-178.07	179.54	anti
8	1.430	-176.52	-17.79	-178.09	179.53	syn
9	1.768	-73.36	10.67	-176.42	57.17	book, OHO, syn
10	1.785	-108.00	-20.56	174.46	-179.92	anti
11	1.972	-109.43	-20.43	174.52	179.93	syn
12	2.327	-179.11	13.23	176.85	-179.31	book, anti
13	2.391	74.47	12.43	171.09	-57.77	OHO, anti
14	2.425	74.51	12.63	171.21	-57.88	OHO, syn
15	2.459	-73.11	11.36	-176.60	57.62	OHO, anti
16	2.826	-73.13	11.28	-176.51	57.58	OHO, syn
17	2.882	100.65	-17.13	-170.64	-179.54	anti
18	2.893	101.57	-17.02	-170.66	-179.49	syn
19	3.053	-179.56	12.62	177.16	-179.34	book,syn
20	3.078	73.35	9.33	169.80	-58.13	book, OHO, syn

Table 3. Relative energies and values of φ, ψ, and ω obtained for the 20 most stable structures for tyrosine-glycine obtained from the hierarchical selection method.

Table 4 lists the 20 most stable conformers (based on single-point MP2 energies corrected by B3LYP zero-point energies) resulting from the hierarchical selection procedure. Their structures are presented in figure 7. Bold entries in table 1 indicate structures not found by the stepwise rotation method. The six most stable conformers have the folded "book" conformation (labeled "book" in table 4). These six conformers were also located by the stepwise rotation method. Most of the twenty most stable conformers contain an OH•••O hydrogen-bonding interaction between the C-terminal hydroxyl group and the carbonyl oxygen of tyrosine (indicated by 'OHO' in table 4). Some conformers just differ in minor structural differences, such as the orientation of the tyrosine OH group or the orientation of the C-terminal carboxyl group. For example, the conformers in the 1↔3, 2↔5, 4↔6, 7↔10, 8↔9, 11↔12, 13↔14, 15↔16 and 17↔19, 18↔20 pairs just differ in the orientation of the OH tyrosine group, whereas conformers 1↔2, 3↔5, 7↔11, 10↔12, 13↔16, 14↔15, 17↔20 and 18↔19 only differ in the orientation of the C-terminus. In addition, conformers 15 and 18 just contain a differently-oriented NH_2 group. All twenty conformers have ψ_{tyr}-values close to 0° (between -20° and +13°), which allows the formation of a short N10-H•••N15 interaction. The φ_{gly}-angles of the twenty most stable conformers fall into three groups: around -100°; -73°/+73°; -100°/+100°. The conformers with a φ_{gly}-angle around -73°/+73° contain the OHO H-bonding interaction in their structure, whereas conformers with a φ_{gly}-angle around -180° and -100°/+100° have either extended or "book" conformations.

Based on the structural elements listed in table 4, the twenty Tyr-Gly conformers can be grouped into four distinct families: book/OHO (conformers 1-3, 5), book (4, 6), extended (7-12), and extended/OHO (conformers 13-20). The stepwise rotation method located all book, OHO and book/OHO conformers listed in table 2, but missed all extended conformers. In the stepwise rotation method we did not account for all possible combinations of the torsion angles since we rotate the bonds with large step sizes so one risks that particular low-lying conformers were missed. A more thorough stepwise method, which would not discard any conformers along the way (but which would have to be done at a lower level of theory than B3LYP), may perform better. However as the sampling of conformers depends on the order of bond rotation, it is not straightforward to define a unique

algorithm for the stepwise rotation method, and we therefore we did not investigate this method further.

Conformers	Rel. Energies (kJ/mol)	ψ_{tyr}	ω_{tyr}	φ_{gly}	ψ_{gly}	Structure
1	0.000	7.90	171.01	73.11	-58.71	OHO, book, anti
2	3.410	11.11	-176.39	-73.26	56.95	OHO, book, anti
3	3.622	9.33	169.80	73.35	-58.13	OHO, book, syn
4	3.872	13.23	176.85	-179.11	-179.31	book, anti
5	4.169	10.67	-176.42	- 73.36	57.17	OHO, book, syn
6	4.543	12.62	177.16	-179.56	-179.34	book, syn
7	**5.557**	**-20.56**	**174.46**	**-108.00**	**-179.92**	**anti**
8	**5.566**	**-17.13**	**-170.64**	**100.65**	**-179.54**	**anti**
9	**5.656**	**-17.02**	**-170.66**	**101.57**	**-179.49**	**syn**
10	**5.815**	**-20.43**	**174.52**	**-109.43**	**179.93**	**syn**
11	**6.485**	**-17.82**	**-178.07**	**-176.08**	**179.54**	**anti**
12	**6.528**	**-17.79**	**-178.09**	**-176.52**	**179.53**	**syn**
13	7.031	-16.73	176.00	73.28	-57.5	OHO, syn
14	7.134	-17.07	175.89	73.22	-57.45	OHO, anti
15	7.135	-16.97	-171.85	-73.69	57.43	OHO, anti
16	7.414	-17.01	-171.83	-73.72	57.45	OHO, anti
17	10.487	12.63	171.21	74.51	-57.88	OHO, syn
18	10.524	11.36	-176.60	-73.11	57.62	OHO, anti
19	10.566	12.43	171.09	74.47	-57.77	OHO, anti
20	10.947	11.28	-176.51	-73.13	57.58	OHO, syn

Table 4. Relative energies ΔE (in kJ mol^{-1}), based on single point MP2/6-31+G* calculations with inclusion of scaled (0.976) B3LYP/6-31+G* zero-point energies; values of the Ramachandran angles φ, ψ and ω (in degrees); and structural elements of the twenty most stable Tyr-Gly conformers, obtained using the hierarchical conformational analysis method.

Entries in bold indicate conformers that were not found by the stepwise rotation.
$\Psi_{tyr} = \tau$ (N15-C8-C9-N10); $\omega_{tyr} = \tau$ (C8-C9-N10-C11); $\Phi_{gly} = \tau$ (C9-N10-C11-C12); $\Psi_{gly} = \tau$ (N10-C11-C12-O13);-see figure 6b
OHO: OH•••O hydrogen bonding interaction the C-terminal hydroxyl group and the carbonyl oxygen of tyrosine. Book: τ (C4-C7-C8-C9) < 90°. Syn/anti: Tyrosine OH syn or anti with respect to glycine OH. All structures have a favourable N10-H•••N15 (R(H•••N)) = 2.1 -2.2 Å; NHN angle> 100°) interaction.

Figure 8 shows the distribution of the glycine Ramachandran angles φ_{gly} and ψ_{gly} for the 115 conformers that were identified and optimized at the B3LYP/6-31+G* level of theory in the current work (union of the conformers resulting from the hierarchical selection scheme and the stepwise rotation). In agreement with the Gly Ramachandran plot by Lovell *et al.*[37] obtained using data from high-resolution proteins, structures with φ-values of 0 ± 60° are not stable conformers as they would suffer steric hindrance between the O atom of the peptide group and the carboxylic acid group of the C-terminus. For Tyr-Gly, structures of the α-type are nearly completely absent. This is not surprising, as a roughly +60°/+60° or -60°/-60° combination of φ/ψ would bring the tyrosine C=O group and the C-terminal C=O together, whereas +60°/-60° or -60°/+60° combinations (γ-type conformers) bring the tyrosine C=O and the C-terminal OH together, thereby allowing the formation of the OHO interaction.

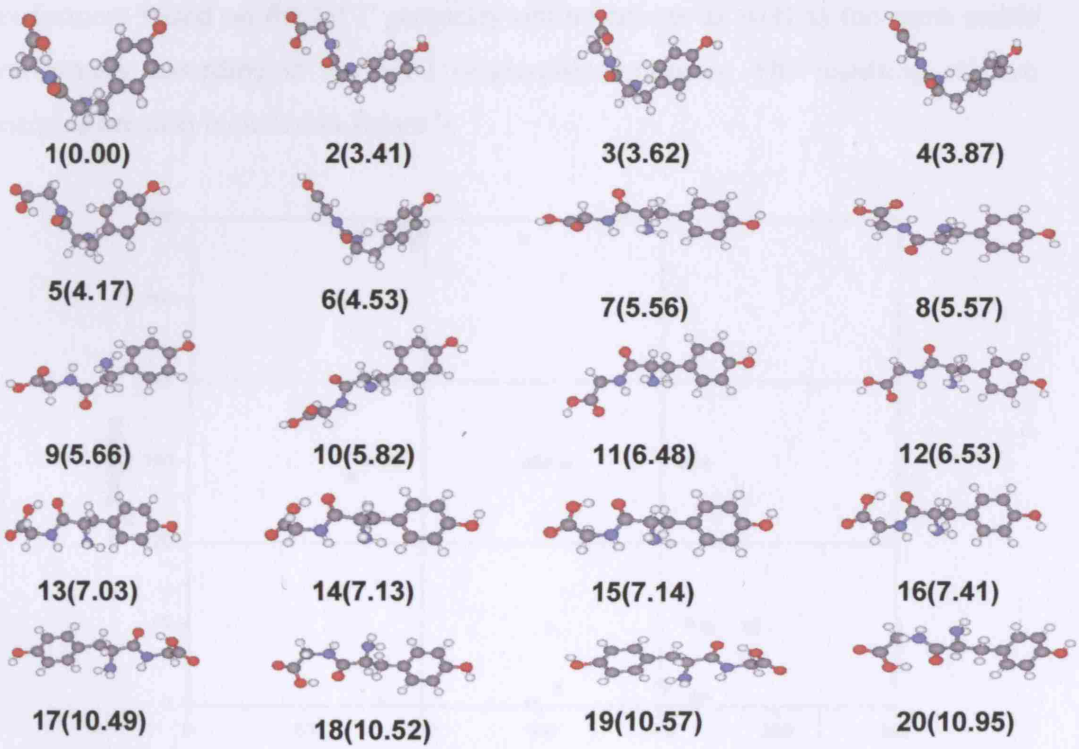


Figure 7. B3LYP/6-31+G* structures of the 20 most stable Tyr-Gly conformers. Relative energies (from single-point MP2/6-31+G* calculations with inclusion of scaled (0.976) B3LYP/6-31+g* ZPEs) are given in kJ mol^{-1}.

Figure 9 compares the relative energies of the Tyr-Gly conformers computed at the B3LYP and single-point MP2 levels of theory, with and without inclusion of (B3LYP, scaled) ZPEs. The ZPE corrections are largest for the OHO-type conformers, and the inclusion of ZPEs consequently favours the non-OHO conformers (4, 6-12). Figure 10 compares the order of stability of the 20 most stable conformers at the B3LYP and the MP2 level of theory. MP2 favours the book conformers (1-6). The order of stability resulting from the MP2 single point energy calculations is considerably different than the order of stability that B3LYP predicts for the most stable conformers of Tyr-Gly (see Figure 10). For example, the most stable conformer according to the MP2 single point energies is conformer 11 according to the DFT calculations. Overall, there is more variation in the MP2 relative energies than in the B3LYP relative energies. To test the reliability of the DFT structures, we performed MP2/6-31+G* geometry optimizations on conformers 1-7 and 11-16, which include the most stable conformers based on the DFT geometry optimizations as well as the most stable conformers according to the MP2 single-point energies. The resulting relative energies are also included in figure 9.

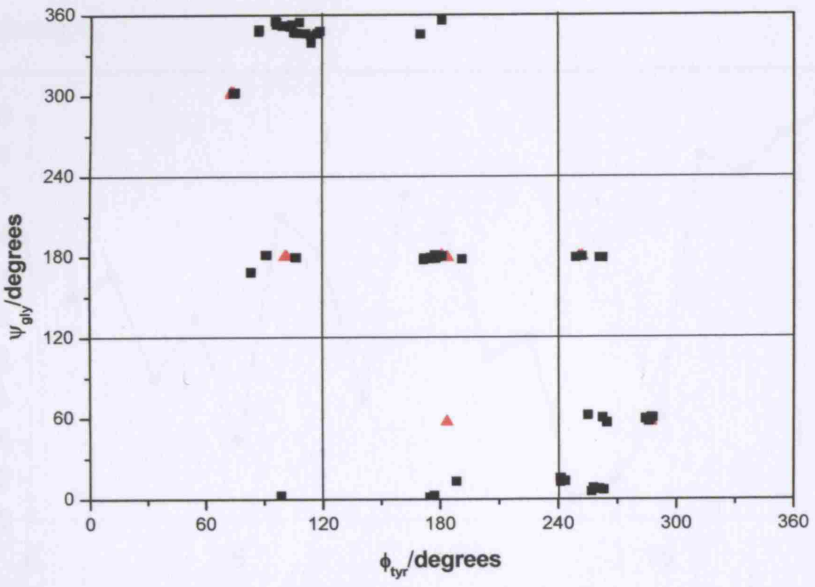


Figure 8. Ramachandran plot showing the φ_{tyr}/ψ_{gly} combination occurring in the 99 Tyr-Gly conformers identified and optimized with B3LYP/6-31+G* in the current work. The points belonging to the twenty most stable conformers (based on the B3LYP energies) are indicates by red colour.

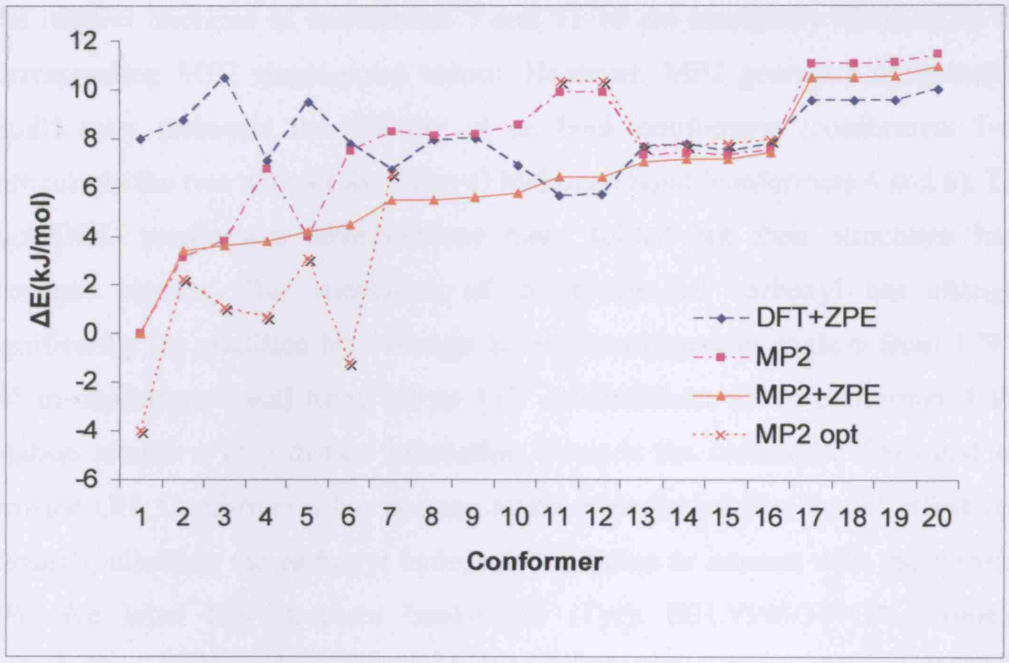


Figure 9. Comparison of the B3LYP, B3LYP+ZPE and MP2+ZPE relative energies of the twenty most stable (based on the B3LYP energies) Tyr-Gly conformers. The ZPEs were computed with B3LYP/6-31+G* and scaled by 0.976. The MP2+ZPE and DFT+ZPE curves have been made to coincide for conformer 15 (the most stable conformer based on the DFT geometry optimizations).

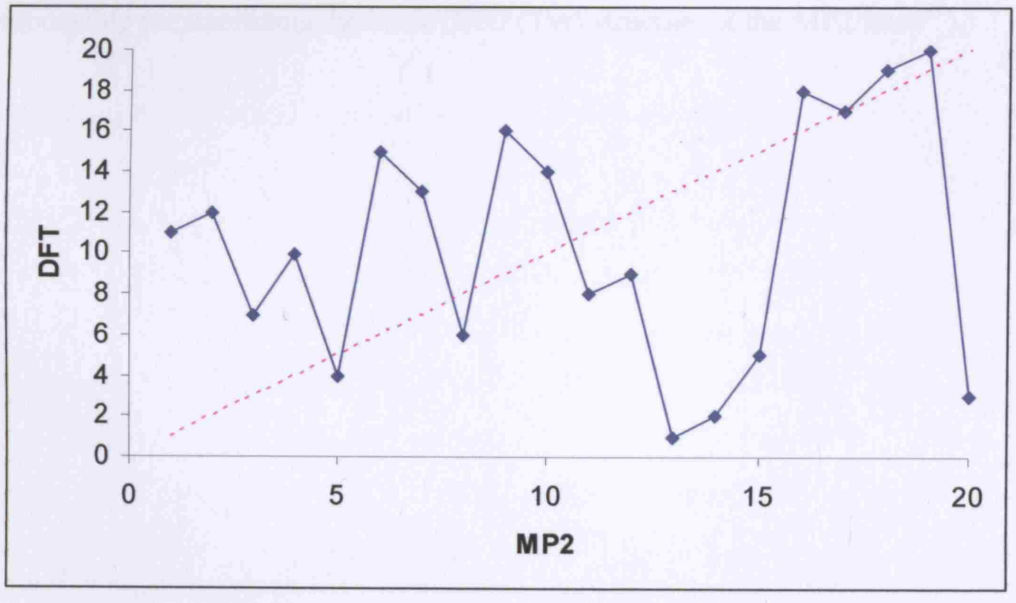


Figure 10. Comparison of the order of stability of the 20 most conformers at the B3LYP and the MP2 level of theory. The dotted line indicates ideal agreement.

The relative energies of conformers 7 and 11-16 are essentially identical to the corresponding MP2 single-point values. However, MP2 geometry optimization significantly increases the stability of all book conformers (conformers 1-6), particularly the two without the OH•••O hydrogen bond (conformers 4 and 6). The book/OHO conformers have become more folded but their structures have remained similar. The orientation of the C-terminal carboxyl has changed significantly (as qualified by a change in the Ramachandran angle φ from 179 to 74° in conformer 4 and from 180 to 113° in conformer 6). In conformer 4 this rotation allows a very distant interaction between the carboxylic C=O and the tyrosine OH. Conformer 6 has become much more folded (the 'book' effectively 'closes'), allowing the carboxyl hydroxyl of glycine to interact with the tyrosine OH. We label this structure book/OHO (Tyr). B3LYP/6-31+G* geometry optimization starting from this folded geometry returns the original B3LYP-optimized structure (structure 6 in figure 7), confirming that the folded book/OHO (Tyr) structure is not a minimum on the B3LYP potential energy surface. The increased degree of folding in the book and book/OHO conformers is likely due to dispersive interactions involving the aromatic ring, which are not correctly described by standard density functionals like B3LYP[38] (although there is indication that intramolecular basis set superposition errors are also partially responsible for stabilizing the book/OHO (Tyr) structure at the MP2 level[39]).

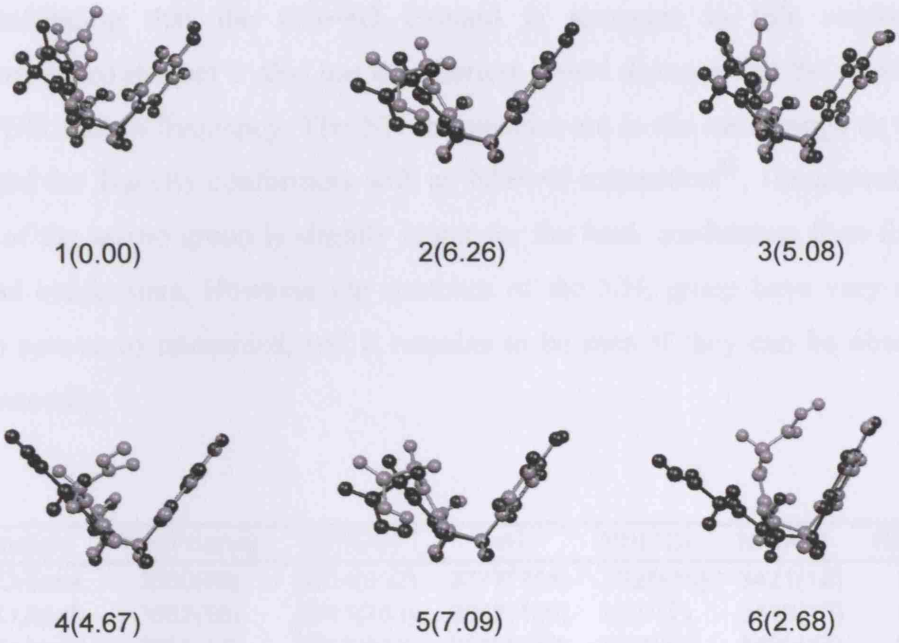


Figure 11. Comparison of the B3LYP/6-31+G* (black atoms) and MP2/6-31+G* (grey atoms) optimized structures of conformers 1-6. The aromatic rings of the DFT and MP2 structures have been made to overlap. Relative energies resulting from the MP2 geometry optimizations are given in kJ mol^{-1}.

Table 5 lists the harmonic vibrational frequencies of the OH and NH stretches of the 20 most stable conformers based on the MP2 single-point energies. The OH•••O interaction of the OHO-type conformers induces a considerable red shift of the OH stretch frequency of the glycine residue (~350-380 cm^{-1} compared to the OH mode of the non-OHO conformers), and the corresponding intensities are predicted to be very large. The glycine OH frequency shift is of similar magnitude as that observed for the OHO-containing Trp-Gly conformers considered by Hünig and Kleinermanns[25]. It is also of comparable magnitude to the frequency shift observed for the catechol OH stretch in noradrenaline-H_2O, when the catechol OH is H-bonding to a water molecule (the OH mode is shifted by 361 cm^{-1} compared to isolated noradrenaline[40, 41]. The OH frequency shift in the Tyr-Gly OHO conformers, resulting from intramolecular H-bonding, is therefore comparable in magnitude to that caused by intermolecular H-bonding in

noradrenaline-H_2O, signifying the considerable strength of the OH•••O interaction in Tyr-Gly. Conformer 1 has the shortest H•••O distance and the largest OH red shift, indicating that the OH•••O H-bond is strongest in this conformer. Additionally, conformer 1 also has the shortest H•••N distance and the most red-shifted NH stretch frequency. The NH frequencies are in the same range as those calculated for Trp-Gly conformers with an NH•••N interaction[25]. The asymmetric stretch of the amino group is slightly larger for the book conformers than for the extended conformers. However the stretches of the NH_2 group have very small (though non-zero) intensities, and it remains to be seen if they can be observed experimentally.

	Structure	OH(Phenyl)	OH(Gly)	NH	NH(SS)	NH(AS)	R(OH-O)	R(NH--H)
1	OHO,book	3660(59)	3214(602)	3377(114)	3328(10)	3421(18)	1.740	2.117
2	OHO,book	3662(58)	3241(753)	3382(109)	3327(7)	3419(16)	1.756	2.131
3	OHO,book	3661(57)	3223(610)	3379(113)	3328(9)	3420(17)	1.744	2.121
4	book	3662(53)	3598(76)	3388(92)	3327(3)	3415(10)		2.185
5	OHO,book	3662(59)	3243(749)	3384(108)	3327(7)	3419(16)	1.758	2.133
6	book	3662(54)	3598(76)	3389(91)	3327(3)	3415(10)		2.186
7	extended	3662(59)	3596(68)	3392(91)	3327(4)	3404(11)		2.158
8	extended	3662(58)	3595(65)	3396(87)	3328(6)	3404(10)		2.183
9	extended	3662(57)	3595(65)	3396(87)	3328(6)	3404(10)		2.182
10	extended	3663(58)	3596(68)	3392(90)	3327(4)	3404(11)		2.158
11	extended	3663(58)	3597(78)	3390(86)	3330(4)	3405(10)		2.227
12	extended	3661(57)	3598(78)	3390(85)	3329(4)	3405(11)		2.227
13	OHO	3662(62)	3240(706)	3386(103)	3330(9)	3405(15)	1.760	2.173
14	OHO	3662(63)	3240(709)	3386(103)	3331(9)	3405(15)	1.760	2.175
15	OHO	3661(63)	3240(761)	3386(90)	3331(6)	3406(16)	1.758	2.163
16	OHO	3661(62)	3240(769)	3385(105)	3331(8)	3405(16)	1.758	2.164
17	OHO	3662(60)	3248(733)	3388(117)	3328(2)	3417(4)	1.761	2.134
18	OHO	3662(61)	3248(755)	3389(113)	3329(2)	3417(4)	1.760	2.144
19	OHO	3662(61)	3248(735)	3388(117)	3329(2)	3417(4)	1.761	2.133
20	OHO	3662(60)	3249(756)	3389(113)	3329(1)	3417(4)	1.761	2.144

Table 5. Scaled (OH: 0.976; NH: 0.956) harmonic vibrational frequencies (in cm^{-1}) of the OH and NH stretch modes. Calculated intensities (km mol^{-1}) are given in round brackets. The R(OH•••O) and R(NH•••N) distances are given in Å. .

SS: Symmetric stretching mode of the two hydrogen atoms of the N-terminus (NH(SS)) AS: Assymetric stretching mode of the two hydrogen atoms of the N-terminus (NH(AS)).

3.4 Discussion

The major problem with conformational studies on peptides is their flexibility, leading to a large number of possible conformers. In the current study we explored the conformational features of the dipeptide Tyr-Gly (by performing single-point Hartree-Fock calculations on all possible conformers); however for larger peptides this would be a too expensive task. The current study shows that the most stable conformers all have several H-bonding interactions (all exhibit a short NH•••N and NH•••π contact; the three most stable conformers also have an OH•••O H-bond). In previous studies it was suggested that the number of hydrogen bonds is not directly correlated to the energy ordering of the conformers[42]. In our study we can observe that some correlation exists for the most stable conformers. All the most stable conformers form two hydrogen-bonding interactions.

The most stable conformers identified in the current work lie very close in energy. Whether all of these can actually be observed in supersonic-jet experiments also depends on the conversion barriers between similar conformers: barriers of 5 kJ mol^{-1} or greater are generally sufficient to trap a particular conformer in a potential well[43]. High conversion barriers for supersonic jet expansion relaxations would not result in significant changes in the conformation structure, whereas low barriers may result in conformers converting into other, more stable, conformations.

The best way to test the ability of our method to select the most stable conformers would be verification by experiment. Unfortunately, experimental gas-phase data on Tyr-Gly are not yet available, but there are some data on related dipeptides containing an aromatic residue at the N-terminal side. Resonance-enhanced multiphoton ionization (REMPI) spectra of jet-cooled dipeptides containing either tyrosine or phenylalanine as the aromatic chromophore (C) suggests that dipeptides of the form C-X (Tyr-Ala, Phe-Ala, Phe-Gly) form an interaction between the carboxyl terminus and the ring, associated with a folded gauche conformation[27]. In our six most stable conformers both the COOH and NH_2 group

adopt a gauche conformation with respect to the aromatic ring; however, the COOH group shows a stronger propensity to bind to the C=O of the peptide group (in the book/OHO conformers), or to the tyrosine OH (in conformer 6), rather than to the aromatic ring. Instead of a COOH-π interaction, the twenty Tyr-Gly conformers show an interaction between the N-terminal amino group and the π-electron cloud of the aromatic ring.

Hünig and Kleinermanns recently published far-IR spectra on Trp-Gly. Although four conformers were observed in the supersonic jet, the IR-UV spectrum of only one of these could be observed. This conformer is unfolded, exhibiting an extended side chain. It contains an NH•••π H-bonding chain, as also observed in all twenty Tyr-Gly conformers considered in the current study, but no COOH•••O=C (OHO) interaction. A subsequent mid-IR spectroscopic study[25], which presented spectra of two of the observed Trp-Gly conformers, confirmed this conclusion: whereas the most stable conformer based on the DFT calculations does contain an OHO hydrogen bond, the two observed conformers do not. Both exhibit an unfolded backbone conformation. In contrast to the experimental findings on Trp-Gly, the DFT and MP2 results presented in the current study indicate that Tyr-Gly prefers a folded (book) conformation with an OHO-type interaction. The two most stable conformers (according to the MP2 geometry optimizations) are conformers 1 and 6 (see figure 11). In conformer 1 the C-terminal carboxyl group H-bonds to the C=O of the peptide group (book/OHO), whereas in conformer 6 it binds to the tyrosine OH (book/OHO (Tyr)). The two structures are very similar otherwise. Based on the absence of OHO-type Trp-Gly conformers in the jet-cooled experiments, Tyr-Gly conformers with such an interaction may not be observed either, as it is unlikely that the formation of this interaction is much influenced by the nature at the aromatic residue. The OHO (Tyr) conformer, however, is unique to peptides containing a phenol ring. It would therefore be interesting to see if an OHO(Tyr) conformer is a stable minimum on the B3LYP potential energy surface from jet-cooled experiments, the presence of this conformer in the supersonic jet would indicate the inadequacy of B3LYP geometry optimizations on peptides containing aromatic residues.

There have been a number of experimental studies in the gas phase on aromatic peptides. A combination of ultraviolet and infrared ion-dip spectroscopic experiments, coupled with ab initio and DFT computation, on tryptophan[44] and phenylalanine[45] identified as the most stable conformers a structure that exhibits a 'daisy chain' of H-bonding interaction (carboxylic proton → amino group → aromatic ring). A similar H-bonding interaction is observed for all twenty most stable conformers identified for Tyr-Gly where all of them form an NH → amino group → aromatic ring H-bonding interaction.

Interaction of the peptide backbone with an aromatic ring has an effect on the stability of the Tyr-Gly dipeptide, as evidenced by the preference for book-type conformations, and the increased stability and foldedness after MP2 geometry optimization (which suggests that interactions with the π-electron cloud of the aromatic ring are responsible for the geometry change). However, apart from the dispersion effect resulting from the interaction of the peptide backbone with the aromatic ring, the basis set employed is also important as large basis sets are required in order to minimize/reduce intramolecular basis set superposition error (BSSE)[37]. BSSE is larger in MP2 than in DFT calculations. In the current method the B3LYP/6-31+G* method may not give reliable results for molecular systems where the intramolecular dispersion energy is likely to be the major factor that determines the conformation.

The large structural changes in the book-type conformers upon MP2 geometry optimization indicates that care should be taken in interpreting the DFT-computed IR frequencies. Cybulski and Seversen recently published a study employing twelve different combinations of exchange and correlation functionals and compared them with results obtained from ab initio calculations for two different configurations of the water dimer and three different configurations of the thymine-adenine complex[46]. From this study, it was concluded that none of the DFT functionals approach the reliability of MP2 method. Additionally, it was proposed that DFT optimizations should be followed by a single point ab initio (MP2 or CCSD(T)) calculation which for large systems could limit the choices only to the MP2 calculation. It would be better to calculate the IR spectra of the

book conformers with MP2, using the MP2-optimized geometries. However, due to computational restraints, we have not attempted this.

Experimental gas-phase data on Tyr-Gly would be very welcome to verify the structures identified in the current study. We would like to emphasize that a thorough structural characterization of small peptides is a difficult task. The large number of possible conformers, couple with the need for high-level quantum chemical methods that can account for dispersion (as shown in the current study), make it nearly impossible to find the most stable conformers without experimental guidance. A combination of different theoretical approaches to identify the low-lying conformers may reduce the risk missing any of these. In the next chapters we will employ the hierarchical selection scheme to the larger peptides Tyr-Gly-Gly and Tyr-Gly-Gly-Phe and Tyr-Gly-Gly-Phe-Leu (Leu-enkephalin) for which is impossible to do single-point HF calculations for all possible conformers.

References

1. Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A., *J.Am.chem. Soc.,* **1997,** 119, 5908.
2. Carney, J. R.; Zwier, T. S., *Chem. Phys. Lett.* **2001,** 341, 77.
3. van Mourik, T.; Emson, L. E. V., *Phys.Chem.chem. Phys.,* **2002,** 4, 5863.
4. van Mourik, T.; Emson, L. E. V., *Phys.Chem.Chem.Phys.,* **2004,** 6, 2827.
5. Van Mourik, T.; Früchtl, H. A., *Molec. Phys.* **2005,** 103, 1641.
6. Snoek, L. C.; van Mourik, T.; Simons, J. P., *Molec. Phys.* **2003,** 101, 1633.
7. Çarçabal, P.; Snoek, L. C.; van Mourik, T., *Molec. Phys. Chem. Phys.* **2005,** 103, 1239.
8. Hunig, I.; Seefeld, K. A.; Kleinermanns, K., *Chem. Phys. Lett.* **2003,** 369, 173.
9. Ramachandran, G. N., *Biopolymers* **1963,** 6, 1494.
10. Keefe, C.; Pearson, J., *J. molec. Struct.* **2004,** 675, 177.
11. Snoek, L. C.; Robertson, E. G.; Kroemer, R. T.; Simons, J. P., *Chem. Phys. Lett.* **2000,** 49, 321.
12. Lindinger, A.; Toennies, J. P.; Vilesov, A. F., *J. Chem. Phys.* **1999,** 110, 1429.
13. Li, L.; Lubman, D., *Appl. Spectrosc.* **1988,** 42, 418.
14. Rizzo, T.; Park, Y.; Levy, D. H., *J. Chem. Phys.* **1986,** 85, 6945.
15. Cable, J.; Tubergen, M. J.; Levy, D. H., *J.Am.Chem.Soc.* **1987,** 109, 6198.
16. De Vries, M. S.; Abo-Riziq, A. G.; Crews, B.; Bushnell, J. E.; Callahan, M. P., *J.Molec Phys.* **2005,** 103, 1491.
17. Wiedemann, S.; Metsala, A.; Nolting, D.; Weinkauf, R., *Phys. Chem. Chem. Phys.* **2004,** 6, 2641.
18. Unterberg, C.; Gerlach, A.; Schrader, T.; Gerhards, M., *J. Chem. Phys.* **2003,** 4, 1760.
19. Gerhards, M.; Unterberg, C., *Phys. Chem. Chem. Phys.* **2003,** 4, 1760.
20. Gerhards, M.; Unterberg, C.; Gerlach, A.; Jansen, A., *Phys. Chem. Chem. Phys.* **2004,** 6, 2682.
21. Chin, W.; Dognon, J. P.; Piuzzi, F.; Dimicolli, I.; Mons, M., *Molec. Phys.* **2005,** 103, 1579.

22. Chin, W.; Dognon, J. P.; Piuzzi, F.; Tardivel, B.; Dimicolli, I.; Mons, M., *J. Am. Chem. Soc.* **2005,** 127, 707.

23. Dian, B. C.; Longarte, A.; Zwier, T. S., *Science* **2002,** 296, 2369.

24. Chass, G. A.; Mirasol, R. S.; Setiadi, D. H.; Tang, T.-H.; Chin, W.; Mons, M.; Dimicoli, I.; Dognon, J. P.; Viskolcz, B.; Lovas, S.; Penke, B.; and Csizmadia, G., *J. Phys. Chem. A* **2005,** 109, 5289.

25. Hunig, I.; Kleinermanns, K., *Phys. Chem. Chem. Phys.* **2004,** 6, 2650.

26. Bakker, J. M.; Plützer, C.; Hünig, I.; Häber, T.; Compagnon, I.; Helden, G., V.; Meijer, G.; Kleinermanns, K., *ChemPhysChem* **2005,** 6, 120.

27. Cohen, R.; Brauer, B.; Nir, E.; Grace, L.; de Vries, M. S., *J. Phys. Chem. A,* **2000,** 104, 6351.

28. Toroz, D.; van Mourik, T., *Molec. Phys.* **2006,** 104, (4), 559-570.

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

30. Gaussian 03, Revision B.04, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, J., T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.;

Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, Q. C., J. V.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A., Gaussian, Inc., Pittsburgh PA, 2003.

31. Schaftenaar, G.; Noordik, J. H., *J. Comp-Aided Mol Design,* **2000,** 14, (2), 123-134.

32. Ramachandran, G. N.; Sasisekharan, V.; Ramakrishman, C. J., *J. mol. Biol.,* **1963,** 7, (95).

33. Becke, A. D., *J. Chem. Phys.,* **1993,** 98, 5648.

34. Stewart, J. J. P., *J. Comp. Chem.* **1989,** 10, 209.

35. Graham, R. J.; Kroemer, R. T.; Mons, M.; Robertson, E. G.; Snoek, L. C.; Simons, J. P., *J. Phys. Chem. A,* **1999,** 103, 9706.

36. Rodger, A.; and Rodger, M., *Molecular Geometry*. Oxford, 1995.

37. Lovell, S. C.; Davis, I. W.; Arendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., *Proteins* **2003,** 50, 437.

38. van Mourik, T.; Gdanitz, R. J., *J. Chem. Phys.,* **2002,** 116, 9620.

39. van Mourik, T.; Karamertzanis, P. G.; Price, S. L., *J.Phys. Chem. A,* 110, 8-12.

40. Snoek, L. C.; van Mourik, T.; and Simons, J. P., *Molec. Phys.* **2005,** 101, 1239.

41. Snoek, L. C.; van Mourik, T.; Carcabal, P.; Simons, J. P., *Phys. Chem. chem. Phys.* **2003,** 5, 4519.

42. Watson, T. M.; Hirst, J. D., *Phys. Chem. Chem. Phys.,* **2003,** 10, 2580.

43. Ruoff, R. S.; Klots, T. D.; Emilsson, T.; Gutowsky, H. S., *J. Chem. Phys.* **1990,** 93, 3142.

44. Snoek, L. C.; Robertson, E. G.; Kroemmer, R. T.; Simons, J. P., *Chem. Phys. Lett.* **2000,** 321, 49.

45. Snoek, L. C.; Kroemmer, R. T.; Hockridge, M. R.; Simons, J. P., *Phys. Chem. Chem. Phys.,* **2001,** 3, 1819.

46. Cybulski, S. M.; Seversen, C. E., *J. Chem. Phys.* **2005,** 122, 014177.

4

THE STRUCTURE OF THE TYR-GLY-GLY TRIPEPTIDE

4.1 Introduction

In the previous chapter we developed a hierarchical selection scheme to locate the preferred conformations of small peptides and applied it to the dipeptide Tyr-Gly (Tyr=tyrosine; Gly=glycine)[1]. In the current chapter we apply the hierarchical selection scheme to explore the conformational landscape of the neutral (non-zwitterionic) tripeptide Tyr-Gly-Gly, which contains the first three amino acid residues of the brain peptide Leu-enkephalin (Tyr-Gly-Gly-Phe-Leu).

Studies have been published in the past for protected amino acids and dipeptides by applying both gas-phase experimental measurements and theoretical methods.[2-12] Several recent studies are focused on the exploration of the conformational features of tripeptides with both experimental and theoretical studies[13-15]. The majority of studies that have been published are focused on the simplest peptides of glycine and alanine and their derivatives and these will be discussed in this thesis in chapter 7[16-20]. Tripeptides have been used for representing protein structure as they contain all necessary factors that can influence the behaviour of a larger biomolecular system. For example, the conformational preferences of the tripeptide Tyr-Gly-Gly can be used in order to determine the conformational preferences of the corresponding pentapeptide Tyr-Gly-Gly-Phe-Leu. Additionally, from the preferences of the Ramachandran torsional angles φ, ψ of the tripeptide, the Ramachandran angle preferences and allowed regions of the Ramachandran surface of the larger system can be predicted.

Recent studies are focused on tripeptides that contain at least one residue with an aromatic ring, because of the possible interaction between the π-electron clouds of the aromatic ring and the peptide backbone. A recent study showed that the interaction between the aromatic ring and the peptide backbone is surprisingly large (10 kcal mol^{-1})[21]. The simplest tripeptide with an aromatic residue is phenylalanine-glycine-glycine (Phe-Gly-Gly). The IR/UV spectrum as well as the

structure of Phe-Gly-Gly with the aid of ab initio quantum chemical and molecular dynamics calculations has been studied by Řeha et al[22]. From this study it was concluded that a folded structure stabilized by a dispersion attraction between the phenyl and the carboxylic group is the global minimum. Additionally empirical potentials such as AMBER[23] and CHARMM[24] as well as DFT methods fail to predict the energetics of this peptide. The semi-empirical SCC-DFTB-D[25] method that covers the London dispersion energy was able to predict the global minimum which is in agreement with the experimental data provided. Hünig and Kleinermanns recently published the spectrum of uncapped Trp-Gly-Gly in the gas phase[26]. In a subsequent study Bakker et al. published a study presenting mid-infrared spectra of this peptide[27]. Both studies predicted a folded structure that contains a strong hydrogen-bonding interaction between the carboxylic group of the peptide backbone and the indole NH bond of tryptophan (1). The characterization of tripeptides with electronic structure theory methods has the limitation that they are computational expensive. Thus the majority of the studies that have been published are using force field and Molecular Dynamics calculations to characterize the conformational preferences of such peptides[28, 29].

Because of the absence of experimental gas-phase data on Tyr-Gly-Gly, to guide the conformational search, in the current study we should take into consideration the complete potential energy surface of the peptide. However, due to computational constraints it is impossible to consider all possible conformers. The computational approach applied for the Tyr-Gly dipeptide extracted the conclusion that conformers with less than two hydrogen-bonding interactions are not among the most stable. This conclusion implies that we can neglect Tyr-Gly-Gly conformers with less than two hydrogen-bonding interactions in their structure. However, it is not possible to perform calculations for all conformers that contain two or more hydrogen-bonding interactions and therefore we did also not take into account conformers that have two and three hydrogen-bonding interactions. To reduce the risk to miss low-lying conformers, we employ a combination of three different strategies to explore the conformational preferences of the Tyr-Gly-Gly tripeptide. The first strategy is based on the hierarchical selection method. The second strategy is based on a stepwise addition procedure of a glycine residue to the most stable conformers obtained for the Tyr-Gly

dipeptide. The third strategy is based on an H-bond selection method. In this method we also consider conformers that contain two and three hydrogen-bonding interactions in their structure.

4.2 Methodology

The conformational energy landscape of the tripeptide Tyr-Gly-Gly is explored based on the hierarchical selection method developed in the previous chapter. Figure 12a shows the flexible bonds of the tripeptide that were considered for the procedure. Figure 12b shows the atomic labelling of the tripeptide with the location of the Φ and Ψ angles. All the calculations were performed using the Gaussian 03 program package[30].

4.2.1 Test run

The hierarchical selection method was employed using large step sizes (90° and 120°) as a test run. The aim of the test run was to determine the best step sizes for rotating the flexible bonds of the tripeptide. The number of conformers that were created from the program was 48780. From these, 11000 conformers were selected for single point energy calculations at the HF/3-21G* level of theory. The conformers were randomly selected. From these calculations the first hundred low-lying conformers were optimized using the HF/3-21G* method. From these geometry optimizations the 75 most stable conformers were selected and optimized using the B3LYP/6-31+G* method. From these calculations 70 conformers optimized to different minimum structures. For the 70 most stable conformers obtained from the DFT calculations, the final values of the dihedral angles of the bonds rotated were collected.

4.2.2 Hierarchical selection run

From the variation of the dihedrals of the 70 most stable conformers resulting from the test run, the optimal step sizes for rotation of the internal bonds were determined. Thus, dihedrals A, E, H and K were varied with a step size of 180°, dihedrals C and G were varied in steps of 120°, dihedrals B and J were varied in steps of 60°, and dihedrals D, F and I were varied with a step size of 30°. The variation of dihedral G showed that, in principle, we should rotate this bond with a step size of 30°. However the first 25 out of 70 low-lying conformers showed that

the variation of these conformers was in steps of 120° and therefore dihedral G was varied with a step size of 120°.

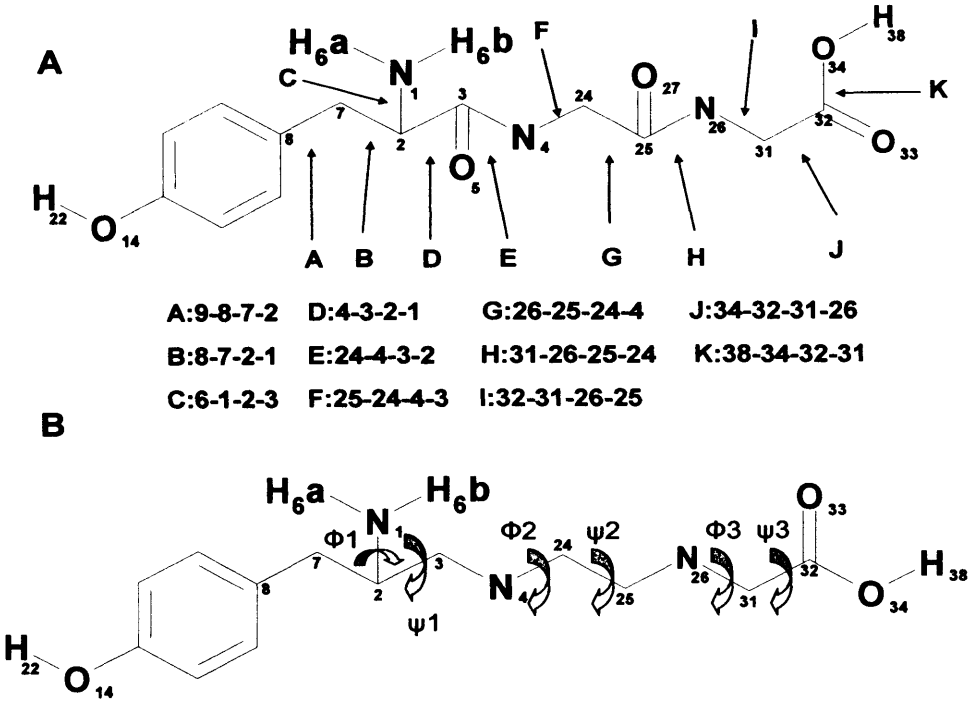


Figure 12. **A.** Atomic labelling and definition of the dihedral angles considered in the hierarchical selection method. **B.** Definition of the angles φ_1, ψ_1, φ_2, ψ_2, φ_3, ψ_3.

The total number of possible conformers generated was 1,255,112. Due to the large number of conformers generated it was too computationally expensive to perform SPE calculations for all the conformers. Instead, single point energy calculations were performed only for the conformers that form four, five, six, seven and eight hydrogen bonds. This is justified as the conformational analysis performed for the Tyr-Gly dipeptide showed that conformers that contain less than two hydrogen-bonding interactions are not among the most stable conformers. However, due to the computational constraints it was not possible to perform single point calculations for all the conformers that contain two or more hydrogen bonds. Thus conformers that contain two and three hydrogen-bonding interactions were also neglected. By neglecting these conformers the risk to miss low-lying conformers is increased. To overcome this problem, different strategies were employed to reduce the risk to miss low-lying conformers (see below). The first 200 low-lying conformers resulting from the single point calculations were

optimized at the HF/3-21G* level of theory. From these optimizations 108 conformers converged to distinct structures. The first 40 low-lying conformers were optimized using the B3LYP functional[31] and the 6-31+G* basis set. From these, 34 conformers optimized to different minimum structures.

4.2.3 Stepwise-addition method

To the 49 most stable conformers obtained from the dipeptide Tyr-Gly, a glycine residue (with Φ=-119° and Ψ=113°) was added using Molden[32] and Tyr-Gly-Gly tripeptide conformers were constructed. The resulting conformers were optimized with the B3LYP/6-31+G* method. Finally 28 conformers converged to different minimum structures. The conformers obtained from these structures were combined with the conformers obtained from the conformers that form four, five, six, seven and eight hydrogen bonds resulting in 55 distinct conformers.

4.2.4 H-bond selection method

The most stable conformers obtained from the hierarchical selection method appear to have some specific hydrogen bonding interactions. The first 30 low-lying conformers contain three specific types of hydrogen bonding interactions. Through the hierarchical selection method all the information about the conformers created, can be collected. The program has been designed to write on a file, the name of the conformer, the number of the hydrogen bonding interactions the conformers contain with the details of the explicit atoms that form these hydrogen bonding interactions. To check if there are additional low-lying conformers with these interactions, not found from the hierarchical selection method for conformers that contain four to eight hydrogen-bonding interactions, we collected all the conformers that form all these specific interactions within their structure. For these conformers single point energy calculations were performed and the most stable conformers were optimized at the HF/3-21G* and B3LYP/6-31+G* levels of theory. The 4 characteristic hydrogen bonding interactions these conformers contain are:

I. The first type of hydrogen-bonding interaction involves a combination of two hydrogen-bonding interactions. The first interaction involves a strong hydrogen bond between the hydroxyl hydrogen of tyrosine and the hydroxyl oxygen of glycine (3) ($\mathbf{O_{14}\text{-}H_{22}\cdots O_{33}}$). The second interaction involves a strong hydrogen bond between the hydroxyl hydrogen of glycine (3) and the hydroxyl oxygen of tyrosine ($\mathbf{O_{34}\text{-}H_{38}\cdots O_{14}}$).

II. The second type of hydrogen bonding interaction also contains the combination of two hydrogen-bonding interactions. The first interaction is between the amide hydrogen of glycine (2) and the nitrogen of the N-terminus ($\mathbf{N_4\text{-}H_{28}\cdots N_1}$). The second interaction is between the amide hydrogen of glycine (3) and the carboxyl oxygen of glycine (2) ($\mathbf{N_{26}\text{-}H_{35}\cdots O_5}$).

III. The third type of hydrogen-bonding interaction also contains the combination of two interactions. The first interaction is between the hydroxyl hydrogen of glycine (3) and the carboxyl oxygen of glycine (2) ($\mathbf{O_{34}\text{-}H_{38}\cdots O_{27}}$). The second interaction is between the amide hydrogen atom of glycine (2) and the nitrogen atom of the N-terminus ($\mathbf{N_4\text{-}H_{28}\cdots N_1}$).

We collected 4488 conformers with type **I** interaction, 14092 conformers with type **II**, and 3000 conformers with type **III.** Single point energy calculations were performed at the HF/3-21G* level of theory for all these conformers collected. For all the structures coming from types **I** geometry optimizations were performed using the HF/3-21G* method. For the conformers of type **II** and **III** HF/3-21G* geometry optimizations were performed for the first 500 low-lying conformers respectively, obtained from the single point energy calculations. From these calculations the 115 most stable conformers of type **I,** 49 conformers of type **II** and 41 conformers of type **III** were selected and combined. From this, the 102 most stable conformers were optimized with B3LYP/6-31+G*. From these calculations 85 conformers optimized to different minimum structures.

The 85 conformers obtained were combined with the structures obtained from the conformers that contain 4, 5, 6, 7, 8 hydrogen bonds (the hierarchical selection run) as well as with the H-bond selection method. The total number of the conformers collected was 123. Single-point MP2 calculations performed to the 45 most stable conformers. The results revealed that MP2 favours folded over extended conformers. We therefore performed further single-point energy MP2 calculations only for the remaining B3LYP optimised conformers that have a folded conformation (containing H-bonding type I, I or III). Also a few additional extended conformers were considered as well. In total the energies of 75 conformers were computed with MP2. Zero-point energies (ZPEs), scaled by 0.976[33] were calculated for the 20 most stable conformers using B3LYP/6-31+G*.

4.3 Results

The H-bond selection method located additional conformers that were not found by the main run. So complementing the main run with the H-bond selection method is essential to locate the most stable conformers of the peptide. Table 6 lists the 20 most stable conformers (based on MP2 single point energy calculations corrected by B3LYP zero point energies) resulting from the H-bond selection method. Figure 13 shows the structures of the 20 most stable conformers. Bold entries in Table 6 indicate structures not found from the main run.

Most structures adopt a folded "circle" structure (labeled "circle" in Table 6), resulting from a small (<90°) $C_8C_7C_2N_1$ torsion angle. Most of the "circle" structures enable an $NH_2 \bullet\bullet\bullet \pi$ interaction between the N-terminus and the phenyl ring of tyrosine. In addition, most conformers contain two characteristic OHO hydrogen-bonding interactions. The first interaction is between the carboxyl oxygen of glycine (3) and the hydroxyl hydrogen of tyrosine (indicated by R-OHO in Table 6). The second interaction is between the C-terminal hydroxyl group and the carbonyl oxygen of glycine (2) (indicated by OHO in Table 6).

Some conformers have only minor structural differences. For example conformers 5↔9 only differ in the orientation of the N-terminus, whereas conformers 1↔2, 17↔19 and 7↔10 differ in the orientation of the peptide backbone (differences in the value of φ_2 and ψ_2 angles by 30°). Dihedral $C_9C_8C_7C_2$ defines the position of the peptide backbone over the aromatic ring. If the absolute value of dihedral $C_9C_8C_7C_2$ is around 90°, the peptide backbone is positioned over the ring, whereas for larger values of dihedral $C_9C_8C_7C_2$ ($\geq$ 100°) the peptide backbone is positioned towards the plane of the ring. For example in conformers 17, 18 and 19 the backbone is positioned towards the plane of the ring whereas in conformers 5 and 9 it is positioned over the plane of the ring.

Conformers	Hbonds	Structure		Rel. Energies(kJ/mol)	ψ1	φ2	ψ2	φ3	ψ3
1	4	R-OHO/OHO, anti	Circle/NHN,NHO	0	-2.13	110.88	-6.26	74.24	-62.91
2	4	R-OHO/OHO, anti	Circle/NHN	6.294	11.21	82.87	-64.03	-71.92	57.22
3	**4**	**R-OHO/OHO, syn**	**Circle/NHN**	**8.479**	**-17.02**	**117.76**	**-35.15**	**-76.32**	**65.17**
4	3	R-OHO/OHO, syn	Circle/NHN	10.731	2.55	107.00	-52.71	-75.26	62.21
5	3	R-OHO/OHO, syn	—	12.820	123.69	-115.71	5.25	-75.07	64.10
6	3	anti	Circle/NHN,NHO	13.336	30.22	83.98	-65.57	-81.82	-169.30
7	3	R-OHO, anti	Circle/NHN,NHO	13.402	15.97	86.75	-55.14	-111.26	47.84
8	**4**	**R-OHO/OHO, syn**	**NHN,NHO**	**13.655**	**-37.13**	**78.72**	**62.45**	**-71.03**	**55.60**
9	**3**	**R-OHO/OHO, syn**	**—**	**13.810**	**117.82**	**-115.57**	**4.93**	**-74.75**	**63.84**
10	3	R-OHO, anti	**Circle/NHN**	14.514	3.27	114.01	-12.03	176.14	57.93
11	3	R1-OHO, syn	Circle/NHN,NHO	16.242	22.66	85.83	-61.53	-98.48	-146.03
12	3	R-OHO, anti	Circle/NHN	17.758	-0.91	118.80	-29.00	110.02	-63.39
13	**3**	**R1-OHO, syn**	**Circle/NHN**	**17.812**	**12.91**	**97.06**	**8.62**	**165.92**	**-145.39**
14	**4**	**R-OHO/OHO*, syn**	**NHN**	**18.670**	**-35.52**	**66.99**	**14.01**	**105.89**	**-25.88**
15	3	OHO, anti	Circle/NHN,NHO	19.025	-15.71	78.97	-63.70	-72.70	58.78
16	3	OHO, anti	Circle/NHN,NHO	19.076	9.89	79.74	-66.81	-72.59	60.60
17	**3**	**R-OHO,OHO, syn**	**NHO**	**19.221**	**111.62**	**-84.33**	**58.94**	**71.50**	**-59.17**
18	**3**	**R-OHO,OHO, anti**	**—**	**19.228**	**121.56**	**-121.67**	**40.97**	**76.13**	**-65.40**
19	**3**	**R-OHO,OHO, anti**	**—**	**19.414**	**137.30**	**-102.90**	**25.19**	**77.06**	**-63.99**
20	**4**	**R-OHO,OHO, anti**	**NHN**	**20.326**	**-49.64**	**80.51**	**13.99**	**74.07**	**-61.98**

Table1. Number of hydrogen bonding interactions, relative energies ΔE (in kJ mol^{-1}) based on single point MP2/6-31+G* calculations with inclusion of scaled (0.976) B3LYP/6-31+G* zero-point energies; values of the Ramachandran angles of ψ_1, φ_2, ψ_2, φ_3, ψ_3 (in degrees), and structural elements of the most stable conformers, obtained using the hierarchical conformational analysis method.
Entries in bold indicate conformers that were not found by the hierarchical selection scheme.

ψ_1=τ (N1-C2-C3-N4), φ2= τ (C25-C24-N4-C3), ψ2= τ (N26-C25-C24-N4), φ3= τ (C32-C31-N26-C25), ψ3= τ (O34-C32-C31-N26)

R-OHO: OH•••O interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of glycine (3). R1-OH•••O: interaction between the hydroxyl oxygen of tyrosine and the hydroxyl hydrogen of glycine (3). OHO: OHO•••O hydrogen bonding interaction between the C-terminal hydroxyl group and the carbonyl oxygen of glycine (2). NHN: Interaction between the amide hydrogen of glycine (2) and the nitrogen of the N-terminus. Circle: τ (C9-C8-C7-C2) < 90°. syn/anti: tyrosine OH syn or anti with respect to glycine (2) NH. NHO: Interaction between the amide hydrogen of glycine (3) and the carboxy; C=O of Tyrosine (1).

Most of the conformers have ψ_1 values close to 0° (between -17° and 30°), which allows the formation of an $N_4H_{28}\bullet\bullet\bullet N_1$ interaction. Using the ψ_1-angles we can divide the conformers into three groups (see Table 6). The first group includes conformers that have a ψ_1 angle close to 0° (between -17° and 30°) (conformers 1-4, 6, 7, 10-13, 15-16). All these contain the "circle" conformation and also contain a strong $N_4H_{28}\bullet\bullet\bullet N_1$ interaction. The second group includes conformers that have a ψ_1 angle close to -40° (between -35° and -50°) (conformers 8, 14, 20). These conformers contain a weak $N_4H_{28}\bullet\bullet\bullet N_1$ interaction and do not adopt the "circle" conformation. The third group contains conformers that have a ψ_1 angle close to 125° (between 117° and 137°) (conformers 5, 9, 17, 18, 19). These conformers do not adopt the "circle" conformation and do not adopt the $N_4H_{28}\bullet\bullet\bullet N_1$ interaction.

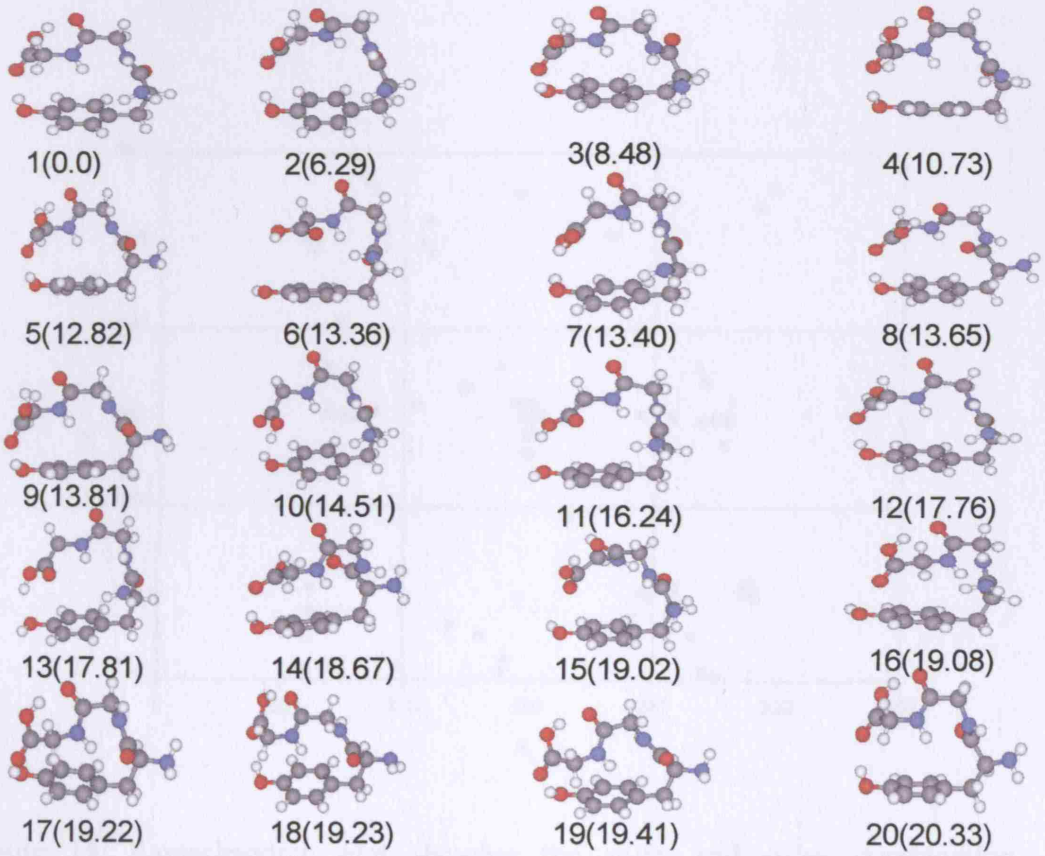


Figure 13. Twenty most stable conformers identified for Tyr-Gly-Gly. Relative energies (from single-point MP2-6-31+G* calculations with inclusion of scaled (0.976) B3LYP/6-31+G* ZPEs) are given in kJ mol^{-1}.

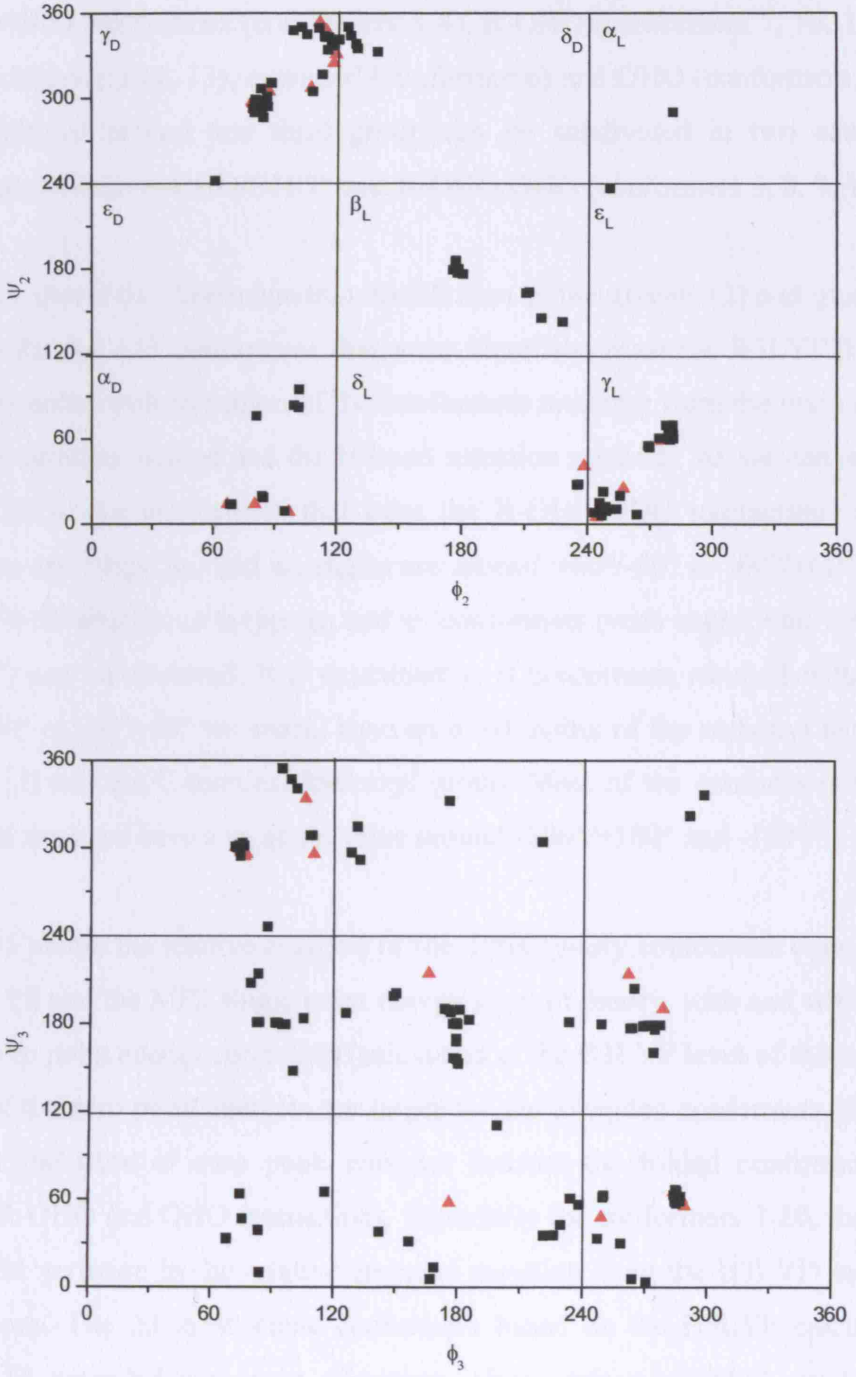


Figure 14. Ramachandran plot showing the φ_2/ψ_2 and φ_3/ψ_3 combination occurring in the 123 Tyr-Gly-Gly conformers identified and optimized with B3LYP/6-31+G* in the current work. The points belonging to the twenty most stable conformers (based on the B3LYP energies) are indicated by red colour.

The first group can be subdivided in 5 categories (see Table 6): conformers with R-OHO/OHO interactions (conformers 1-4), R-OHO (conformers 7, 10, 12), R1-OHO (conformers 11, 13), extended (conformer 6) and OHO (conformers 15-16). The combined second and third group can be subdivided in two categories: conformer 14 with R-OHO/OHO* and R-OHO/OHO (conformers 5, 8, 9, 17-20).

Figure 14 shows the Ramachandran distribution of the glycine (2) and glycine (3) residues for the 123 conformers that were identified from the B3LYP/6-31+G* geometry optimizations (union of the conformers resulting from the main run, the stepwise addition method and the H-bond selection method). As we can see from table 6 all of the conformers that have the R-OHO/OHO interactions in their structures are γ-type (φ_3 and ψ_3 angles are around +60°/-60° or -60°/+60°). From Figure 14 the absence of α-type φ_3 and ψ_3 conformers (with angles +60°/+60° or -60°/-60°) can be observed. It is explained as if conformers resulted with angles +60°/+60° or -60°/-60° we would have an overlapping of the carbonyl oxygen of glycine (2) and the C-terminal hydroxyl group. Most of the conformers with an extended structure have a ψ_3 angle value around -100°/+100° and -180°/+180°.

Figure 15 shows the relative energies of the Tyr-Gly-Gly conformers computed at the B3LYP and the MP2 single point energy level of theory, with and without the scaled zero point energy correction (calculated at the B3LYP level of theory). The values of the zero point energies are larger for the extended conformers. So MP2 with the inclusion of zero point energies favours the folded conformers that contain R-OHO and OHO interactions. Especially for conformers 7-20, there is a significant variation in the relative energies resulting from the B3LYP and MP2 calculations. The 20 most stable conformers based on the B3LYP calculations contain 11 extended structures. However, after performing MP2 single point energy calculations none of the most stable conformers have an extended structure. Additionally some of the 20 most stable conformers resulting from the MP2 single point energy calculations are not among the most stable based on B3LYP calculations (see Figure 16). For example, conformer 14 (according to MP2 single point calculations) is conformer 91 according to the order of stability that DFT predicts. This is due to the existence of a hydrogen bond between the carboxyl OH group with the carbonyl C=O of glycine (2). The reason for these

significant differences is that B3LYP fails to describe dispersion energy effects, which are important in interactions with π-electron clouds[34, 35]. The results obtained are in agreement with the results for the Tyr-Gly dipeptide[1], where the 6 most stable conformers, based on the MP2 single point energy calculations, had a folded structure ("book" conformers). DFT predicted a significantly different order of stability of the most stable conformers of the peptide than that based on the MP2 single-point calculations. Particularly, DFT underestimated the "book" conformers. Additionally geometry optimizations performed at the MP2 level increased further the stability of the folded structures of Tyr-Gly.

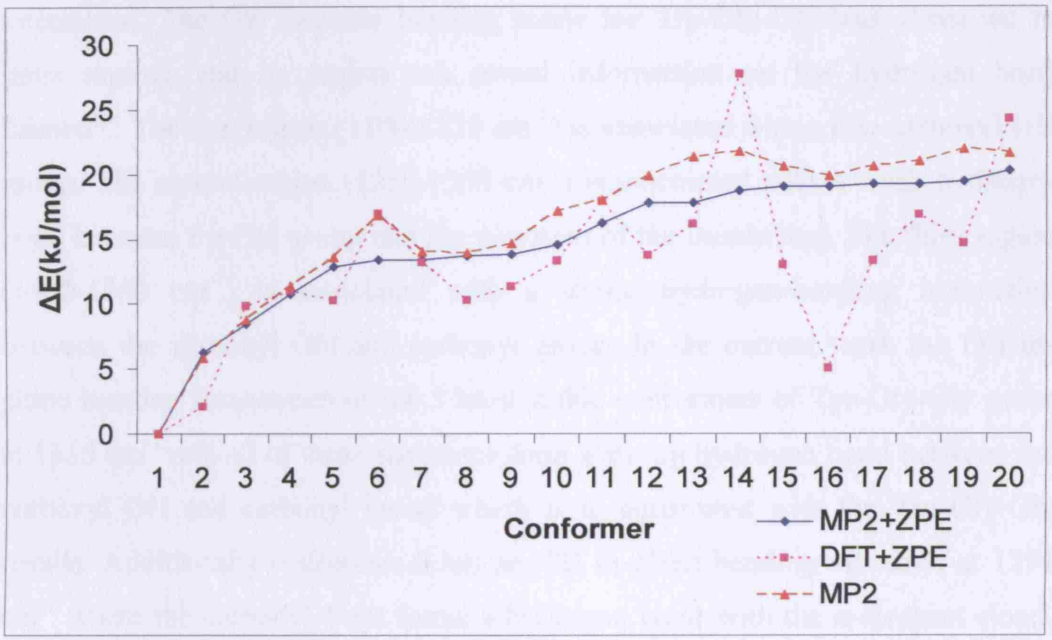


Figure 15. Comparison of the B3LYP, B3LYP+ZPE, MP2+ZPE relative energies of the twenty most stable Tyr–Gly-Gly conformers (based on the DFT geometries). The ZPEs were computed with B3LYP/6-31+G* and scaled by 0.976.

Table 7 lists the harmonic vibrational frequencies of the NH, the OH of glycine (3) as well as their in-plane bending frequencies. The OH•••O interaction induces a red shift of the OH of glycine (3) and the corresponding intensities are predicted to be very large compared with the conformers that do not contain the OH•••O interaction. The glycine (3) OH stretch frequency is of the same magnitude as those of the two most stable structures obtained for Trp-Gly-Gly by Hünig and

Kleinermanns[27]. The frequency of the C=O stretching vibrations was measured at 1671 cm^{-1} for Trp-Gly-Gly whereas in our study for Tyr-Gly-Gly it was predicted at 1658 cm^{-1}. The magnitude of the NH in-plane bending (1503 cm^{-1}) and the O-H in-plane bending vibrations (1410 cm^{-1}) are similar to those predicted for Trp-Gly-Gly (1504 cm^{-1}) the N-H in-plane bending and 1421 cm^{-1} the OH in-plane bending). Intensities of the asymmetric stretch mode of the amino group are slightly lower for the conformers that have an $N_4H_{28}\bullet\bullet\bullet N_1$ angle smaller than 100°. The NH stretch frequencies of glycine (2) are predicted to have smaller magnitude and considerably lower intensities for the circle conformers that form the $N_4H_{28}\bullet\bullet\bullet N_1$ interaction than the circle conformers that do not form the $N_4H_{28}\bullet\bullet\bullet N_1$ interaction. The OH in-plane bending mode for Trp-Gly-Gly was classified in three regions and its region can reveal information on the hydrogen bond formed[27]. The first region (1100-1175 cm^{-1}) is associated with a free carboxyl OH group. The second region (1250-1350 cm^{-1}) is associated with a weak hydrogen bond between the OH group and the π-system of the indole ring. The third region (1400-1500 cm^{-1}) is associated with a strong hydrogen-bonding interaction between the carboxyl OH and carbonyl group. In the current work the OH in-plane bending frequencies of the 5 most stable conformers of Tyr-Gly-Gly occur at 1415 cm^{-1} and all of these structures form a strong hydrogen bond between the carboxyl OH and carbonyl group which is in agreement with the Trp-Gly-Gly results. Additionally conformer 6 has an OH in-plane bending vibration at 1294 cm^{-1} where the carbonyl C=O forms a hydrogen bond with the π-electron cloud. Conformer 14 forms a hydrogen bond between the carboxyl OH group with the carbonyl C=O of glycine (2) and therefore one would expect the OH in-plane bending of this conformer at 1400-1500 cm^{-1}. However, the in-plane bending frequency of this conformer is observed at 1315 cm^{-1}. This is probably due to the existence of a weak hydrogen bond between the carboxyl OH and the π-cloud. This shows the effect of the interaction of the peptide backbone with the aromatic ring. Unfortunately experimental data for Tyr-Gly-Gly are not yet available. It would be interesting to compare the computed vibrational frequencies of the peptide with experimental data.

	Structure	NH		NH(AS)		OH(Phenyl)		NH*		NH(SS)		OH(Gly)		CO		NHB		OHB	
1	R-OHO,OHO,Circle	3436	(128)	3422	(14)	3467	(434)	3374	(92)	3331	(4)	3216	(499)	1659	(281)	1503	(173)	1410	(12)
2	R-OHO,OHO,Circle	3316	(186)	3416	(14)	3536	(293)	3383	(112)	3326	(6)	3185	(556)	1658	(152)	1526	(172)	1415	(232)
3	R-OHO,OHO,Circle	3428	(149)	3404	(3)	3466	(399)	3382	(77)	3328	(0)	3240	(460)	1658	(281)	1504	(182)	1409	(283)
4	R-OHO,OHO,Circle	3415	(246)	3425	(13)	3478	(263)	3383	(98)	3340	(2)	3226	(478)	1659	(290)	1504	(192)	1411	(263)
5	R-OHO,OHO	3427	(156)	3401	(12)	3471	(371)	3450	(21)	3319	(1)	3248	(461)	1661	(291)	1500	(139)	1406	(283)
6	Circle	3359	(127)	3413	(9)	3658	(52)	3389	(80)	3322	(5)	3591	(49)	1725	(395)	1517	(205)	1294	(9)
7	R-OHO,Circle	3362	(127)	3415	(10)	3560	(212)	3396	(95)	3328	(3)	3542	(49)	1717	(422)	1518	(173)	1371	(41)
8	R-OHO,OHO	3273	(272)	3405	(4)	3551	(277)	3411	(78)	3326	(0)	3152	(622)	1654	(99)	1538	(177)	1426	(219)
9	R-OHO,OHO	3428	(145)	3408	(5)	3468	(399)	3453	(25)	3323	(3)	3238	(471)	1661	(283)	1501	(147)	1408	(282)
10	R-OHO,Circle	3452	(79)	3419	(14)	3569	(162)	3376	(89)	3330	(4)	3464	(181)	1709	(311)	1469	(190)	1382	(57)
11	R1-OHO,Circle	3362	(123)	3414	(8)	3656	(59)	3399	(86)	3326	(2)	3496	(184)	1720	(405)	1526	(188)	1386	(23)
12	R-OHO,Circle	3428	(87)	3422	(15)	3530	(398)	3374	(94)	3331	(4)	3587	(66)	1718	(390)	1497	(222)	1370	(54)
13	R1-OHO,Circle	3449	(65)	3412	(10)	3655	(61)	3392	(84)	3323	(3)	3445	(262)	1711	(447)	1494	(235)	1403	(119)
14	R-OHO,OHO*	3439	(175)	3406	(3)	3505	(188)	3383	(68)	3328	(0)	3426	(391)	1722	(358)	1487	(107)	1315	(157)
15	OHO,Circle	3285	(246)	3412	(3)	3647	(70)	3391	(92)	3334	(0)	3213	(528)	1658	(130)	1535	(178)	1408	(204)
16	OHO,Circle	3289	(254)	3417	(16)	3651	(65)	3376	(120)	3324	(10)	3239	(496)	1657	(87)	1532	(185)	1405	(223)
17	R-OHO,OHO	3338	(169)	3406	(3)	3532	(280)	3460	(28)	3322	(1)	3192	(560)	1659	(206)	1526	(156)	1414	(233)
18	R-OHO,OHO	3424	(165)	3409	(5)	3473	(375)	3455	(22)	3324	(3)	3257	(437)	1662	(293)	1501	(167)	1405	(278)
19	R-OHO,OHO	3414	(159)	3397	(4)	3545	(318)	3450	(20)	3318	(2)	3245	(463)	1659	(304)	1504	(171)	1406	(276)
20	R-OHO,OHO	3421	(97)	3397	(3)	3475	(331)	3420	(163)	3319	(0)	3223	(492)	1657	(116)	1505	(169)	1408	(271)

Table 7. Scaled (OH: 0.976; NH: 0.956) harmonic vibrational frequencies (in cm^{-1}) of the OH and NH stretch and in-plane bending modes. Calculated intensities (km mol^{-1}) are given in round bracket.
NH*: NH stretch and in-plane bending modes amide hydrogen of glycine (3).
NH: NH stretch and in-plane bending modes amide hydrogen of glycine (2).
NH(SS): Symmetric stretching mode of the two hydrogen atoms of the N-terminus.
NH(AS): Asymmetric stretching mode of the two hydrogen atoms of the N-terminus.

4.4 Discussion

Studies based on resonance-enhanced multiphoton ionization (REMPI) spectra of jet-cooled dipeptides containing either tyrosine or tryptophan as the aromatic chromophore (C) showed that dipeptides of the form C-X (Tyr-Ala, Phe-Ala, Phe-Gly) form an interaction between the carboxyl terminus and the ring, associated with a gauche conformation[36]. In the two most stable conformers resulting from MP2 geometry optimizations of the Tyr-Gly "book" conformers, the COOH group interacts either with the carboxyl oxygen (C=O) of glycine (2) or with a tyrosine OH. Thus, the COOH group shows a tendency to bind to the carboxyl oxygen of glycine (2) or to the tyrosine OH rather than to the aromatic ring. In the most stable conformers found for Tyr-Gly-Gly, the COOH group binds to the C=O group and to the tyrosine OH at the same time expressed by the two characteristic OHO hydrogen bonding interactions found in the most stable Tyr-Gly-Gly conformers.

Experimental gas-phase data on Tyr-Gly-Gly are not yet available. However, Hünig and Kleinermanns recently published far-IR and mid-IR spectra on Trp-Gly-Gly. The IR-UV spectrum observed two conformers. The first conformer exhibits a strong hydrogen bond between the COOH group and the carboxyl C=O group (COOH•••O=C (OHO)). Also a second interaction occurs in the Trp-Gly-Gly tripeptide between the indole NH and the terminal COOH group of the peptide backbone. The second conformer adopts an unfolded structure. The DFT and MP2 results presented in the current study indicate that the most stable Tyr-Gly-Gly conformer also contains two simultaneous hydrogen bonds. The first interaction is between the COOH group and a carboxyl C=O group (COOH•••O=C (OHO)), like in Trp-Gly-Gly. In addition a second strong interaction exists between the hydroxyl OH group of the tyrosine residue and the C=O carboxyl of glycine (3).

A combined theoretical/experimental study has been recently published on Phe-Gly-Gly by Řeha et al[22]. Four conformers were identified by IR/UV double resonance spectra. The global minimum corresponds to a folded structure which is

characterized by a dispersion attraction between the phenyl and the COOH carboxylic groups. The global minimum obtained for Phe-Gly-Gly has a structure similar to that of conformer 6 for Tyr-Gly-Gly predicted from our study. In contrast to Phe-Gly-Gly the global minimum of Trp-Gly-Gly investigated by Hünig and Kleinermanns[26] involves a hydrogen bond between the indole NH and the terminal COOH group of the peptide backbone while in our study for Tyr-Gly-Gly the global minimum involves an interaction between the hydroxyl OH group of the tyrosine and the C=O carboxyl of glycine (3).

Most of the fifteen low-lying conformers of Phe-Gly-Gly investigated by Řeha are stabilized by three hydrogen-bonding interactions. The first one is between the amide hydrogen of glycine (2) and the nitrogen of the N-terminus (N_4-H_{28}•••N_1). The second is between the amide hydrogen of glycine (3) and the carboxyl oxygen of glycine (2) (N_{26}-H_{35}•••O_5) and the third interaction is between the hydroxyl hydrogen of glycine (3) and the carboxylic oxygen of glycine (2) (O_{34}-H_{38}•••O_{27}). These results are comparable with the most stable conformers, including the global minimum, resulting from our study for Tyr-Gly-Gly where the same interactions stabilize these conformers. However, in Tyr-Gly-Gly a strong hydrogen bond exists between the hydroxyl hydrogen of tyrosine and the hydroxyl oxygen of glycine (3) (O_{14}-H_{22}•••O_{33}), while in Phe-Gly-Gly there is a long distance between the peptide backbone and the phenyl ring. So the hydroxyl group attached to the phenyl group of the aromatic residue plays an important role in the stability of the conformers.

The two most stable conformers found with B3LYP remain the two most stable conformers with MP2 (see Figure 16). These conformers form a folded structure. This is an evidence for the interaction of the peptide backbone with the aromatic ring. Additionally the 20 most stable conformers according to the single point energies at the MP2 level have a folded conformation. The large structural changes in the stability of the conformers upon MP2 single point calculations indicates that care should be taken in interpreting the DFT-computed conformations as well as the DFT-computed IR frequencies. It would be better to perform geometry optimizations at the MP2 level for the most stable conformers resulting from the MP2 single point calculations and to calculate the IR spectra

for MP2 geometry optimized structures. However, due to computational restraints we have not attempted this.

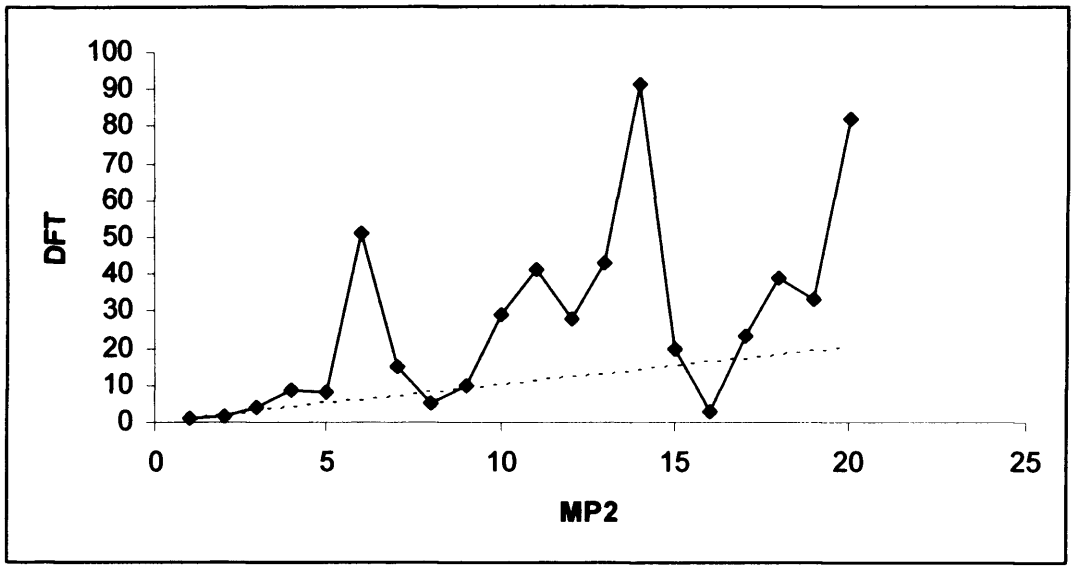


Figure 16. Relative stability of the 20 most stable conformers found for Tyr-Gly-Gly with B3LYP+ZPE and MP2+ZPE. The dotted line indicates ideal agreement.

In the current study it was impossible to perform single point calculations for all possible conformations of the peptide. The structural characterisation of small peptides is a difficult task to perform. The large number of possible conformers and the need to perform high-level quantum chemical methods that can account for dispersion make it nearly impossible to find all the most stable conformers of the peptide without experimental guidance. A combination of different theoretical approaches as done in this work may reduce the risk to miss low-lying conformers. In the next chapter we explore the current method for a larger (tetra) peptide, for which the number of possible conformers is much larger and thus, even more restrictions need to be applied in order to perform single point HF calculations.

for MP2 geometry optimized structures. However, due to computational restraints we have not attempted this.

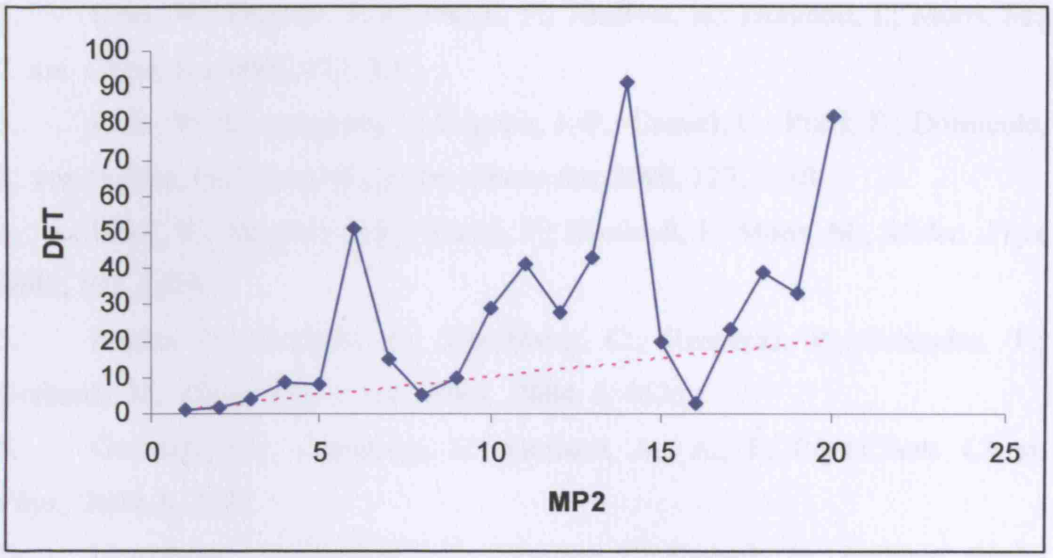


Figure 16. Relative stability of the 20 most stable conformers found for Tyr-Gly-Gly with B3LYP+ZPE and MP2+ZPE. The dotted line indicates ideal agreement.

In the current study it was impossible to perform single point calculations for all possible conformations of the peptide. The structural characterisation of small peptides is a difficult task to perform. The large number of possible conformers and the need to perform high-level quantum chemical methods that can account for dispersion make it nearly impossible to find all the most stable conformers of the peptide without experimental guidance. A combination of different theoretical approaches as done in this work may reduce the risk to miss low-lying conformers. In the next chapter we explore the current method for a larger (tetra) peptide, for which the number of possible conformers is much larger and thus, even more restrictions need to be applied in order to perform single point HF calculations.

References

1. Toroz, D.; Van Mourik, T., *Molec. Phys* **2006,** 104, (4), 559.
2. Chin, W.; Dognon, J.-P.; Piuzzi, F.; Tardivel, B.; Dimicoli, I.; Mons, M., *J. Am. Chem. Soc* **2005,** 127, 707.
3. Chin, W.; Compagnon, I.; Dognon, J.-P.; Canuel, C.; Pizzi, F.; Domicole, I.; von Helden, G.; Mons, M., *J. Am. Chem. Soc* **2005,** 127, 1388.
4. Chin, W.; Dognon, J.-P.; Piuzzi, F.; Dimicoli, I.; Mons, M., *Molec. Phys* **2005,** 103, 1579.
5. Fricke, H.; Gerlach, A.; Untenberg, C.; Rzepecki, P.; Schrader, T.; Gerhards, M., *Phys.Chem. Chem. Phys.*, **2004,** 6, 4636.
6. Gerhards, M.; Untenberg, C.; Gerlach, A.; A., J., *Phys.Chem. Chem. Phys.*, **2004,** 6, 2682.
7. Untenberg, C.; Gerhards, M.; Schrader, T.; Gerlach, A., *J. Chem. Phys.*, **2003,** 118, 8296.
8. Gerhards, M.; Untenberg, C., *Phys.Chem. Chem. Phys.*, **2002,** 4, 1760.
9. Dian, B. C.; Longarte, A.; S., Z. T., *Science* **2002,** 296, 2369.
10. Dian, B. C.; Longarte, A.; Mercier, S.; Evans, D. A.; Wales, D. J.; S., Z. T., *J. Chem. Phys.*, **2002,** 117, 10688.
11. Antohi, O.; Naider, F.; Sapse, A., *J. Molec. Struct (Theochem)* **1996,** 360, 99.
12. Antohi, O.; Sapse, A., *J. Molec. Struct (Theochem)* **1998,** 430, 247.
13. Padiyar, G. S.; Seshadri, T. P., *Acta Cryst.* **1996,** C52, 1693.
14. Pavone, V.; Lombardi, A.; Saviano, M.; Nastri, F.; Zaccaro, L.; Maglio, O.; Pedone, C.; Omote, Y.; Yamanaka, Y.; Yamada, T., *J. Pep. Sci.* **1998,** 4, 21.
15. Ivanova, B. B.; Arnaudov, M. G., *Spectrochimica Acta A* **2006,** 65, 56.
16. Torii, H.; Tasumi, M., *Journ. of Ram. Spec.* **1998,** 29, 81.
17. Salpieto, S.; Viskolcz, J.; Csizmadia, I. G., *J. Molec. Struct (Theochem)* **2003,** 666-667, 89.
18. Mu, Y.; Stock, G., *J. Phys. Chem B.* **2002,** 106, 5294.
19. Nguyen, P.; Stock, G., *J. Chem. Phys.* **2003,** 119, 11350.
20. Mu, Y.; Kosov, D.; Stock, G., *J. Chem. Phys. B.* **2003,** 106, 5294.
21. Duan, G.; Smith, V. H.; Weaver, J. D. F., *Int. J. Quantum Chem.* **2002,** 99, 669.

22. Reha, D., Valdés H.,Vondráek, J.,Hobza, P.,Abu-Riziq A., Bridgit Crews, B., De Vries, M., S., *Chem. Eur. J.* **2005,** 11, 6803.

23. Brooks, B. R.; R.E., B.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., *J. Comp. Chem* **1983,** 4, 187.

24. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A., *J. Comp. Chem* **1986,** 7, 230.

25. Elstner, M.; Hobza, B.; Frauenheim, T.; Suhai, S.; Kaxiras, E., *J. Chem. Phys.* **2001,** 114, (22), 5149.

26. Hunig, I.; Kleinermanns, K., *Phys.Chem. Chem. Phys.,* **2004,** 6, 2650.

27. Bakker, J. M.; Plützer, C.; Hünig, I.; Häber, T.; Compagnon, I.; Helden, G., V.; Meijer, G.; Kleinermanns, K., *ChemPhysChem* **2005,** 6, 120.

28. Chin, W.; Mons, M., *Phys.Chem. Chem. Phys.,* **2004,** 6, 2700.

29. Williams, D. E., *J. Comput. chem.* **2001,** 22, (11), 1154.

30. Gaussian 03, Revision B.04, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, J., T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, Q. C., J. V.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A., Gaussian, Inc., Pittsburgh PA, 2003.

31. Becke, A. D., *J. Chem. Phys.,* **1993,** 98, 5648-5662.

32. Schaftenaar, G.; Noordik J.H., *J. Comp-Aided Mol Design,* **2000,** 14, (2),123-134.

33. Caminati, W.; Di Bernardo, S., *J. Mol. Struct.* **1990,** 240, 253.

34. Hobza, P.; Sponer, J., *Chem. Rev.* **1999,** 99, 2471.

35. Cybulski, S. M.; Severseen, C. E., *J. Chem.Phys.* **2005,** 122, 14177.

36. Cohen, R.; Brauer, B.; Nir, E.; Grace, L.; de Vries, M. S., *J.phys. Chem. A,* **2000,** 104, 6351-6355.

5

THE STRUCTURE OF THE TYR-GLY-GLY-PHE TETRAPEPTIDE

5.1 Introduction

In the previous chapter we applied the hierarchical selection scheme to explore the conformational features of the tripeptide Tyr-Gly-Gly. In the current study we apply the hierarchical selection method to explore the conformational energy landscape of the tetrapeptide Tyr-Gly-Gly-Phe, which contains the first four amino acids residues of the pentapeptide Leu-enkephalin Tyr-Gly-Gly-Phe-Leu. To explore the conformational energy landscape of a peptide that contains four residues, considering the high flexibility of such a peptide, the creation of all possible conformers would result in a very large number of possible conformations. In addition the absence of suitable experimental data on this peptide chain makes it even more difficult to locate the preferred conformers of this peptide since all possible conformations should be taken into account. From the previous chapters the following conclusions can be made:

- Conformers of the Tyr-Gly dipeptide (chapter 3) that contain less than two hydrogen-bonding interactions were not among the most stable ones.
- The most stable conformers of the Tyr-Gly dipeptide resulting from MP2 single-point energy calculations (based on B3LYP geometries), contain characteristic folded "book" conformers. MP2 geometry optimizations give more folded structures.
- Conformers of the Tyr-Gly-Gly tripeptide (chapter 4) that contain less than three hydrogen-bonding interactions were not among the most stable ones.
- The most stable conformers of the Tyr-Gly-Gly tripeptide resulting from MP2 single-point energy calculations (based on B3LYP geometries), contain characteristic folded "circle" conformers.
- The most stable conformations of the Tyr-Gly-Gly tripeptide contain a characteristic interaction between the hydroxyl OH of tyrosine and the carboxyl oxygen of glycine (3).

The conclusions extracted from chapters 3 and 4 justify that we can neglect Tyr-Gly-Gly-Phe conformers with less than three hydrogen-bonding interactions in their structure. However, due to the large number of conformers created, it was not possible to perform calculations for all conformers that contain three or more hydrogen-bonding interactions. The structure motif of the conformers obtained for the Tyr-Gly-Gly tripeptide led us to employ the following strategy: the hierarchical selection method is applied to explore the conformational features of Tyr-Gly-Gly-Phe, collecting only conformers that contain the specific hydrogen-bonding interactions that describe the folded conformers of Tyr-Gly-Gly (chapter 4) and the folded structure resulting from the MP2 geometry optimization of Tyr-Gly conformer 6 (chapter 3).

A number of studies have been published on the conformational analysis of tetrapeptides using experimental methods[1-6]. Garbay-Jaureguiberry et al. published a study on the conformational analysis by ^{15}N NMR spectroscopy using Tyr-Gly-Gly-Phe as a model compound[7]. The tetrapeptide adopted a folded conformation both in DMSO solution and in solid state. This structure was characterized by a β-turn structure with an interaction between the carboxyl of phenylalanine (4) and the amide hydrogen of tyrosine (1). Zaluski et al. performed an X-ray and ^{1}H NMR spectroscopy study on Tyr-Gly-Gly-Phe and Gly-Gly-Phe-X (X=Leu, Met). From this study a bent conformation was found with an intramolecular bond around the phenylalanine (4) residue[8]. Due to the high flexibility of the tetrapeptide Tyr-Gly-Gly-Phe it is impossible to perform a complete scan of the potential energy surface of this peptide. However studies have investigated the conformational features of smaller alanine and glycine tetrapeptides[9-11]. Gresh et al.[12] performed a theoretical study to examine the intramolecular interaction energies in alanine and glycine tetrapeptides using ab initio (HF/MP2), DFT, and polarizable molecular mechanics methods[13]. Beachy et al. computed the energies of 10 conformers of the alanine tetrapeptide by several molecular mechanics force fields[14]. The results of these calculations were compared with ab initio LMP2 results. It was concluded that ab initio and DFT methods are essential to validate the results from a molecular mechanics conformational analysis.

5.2 Methodology

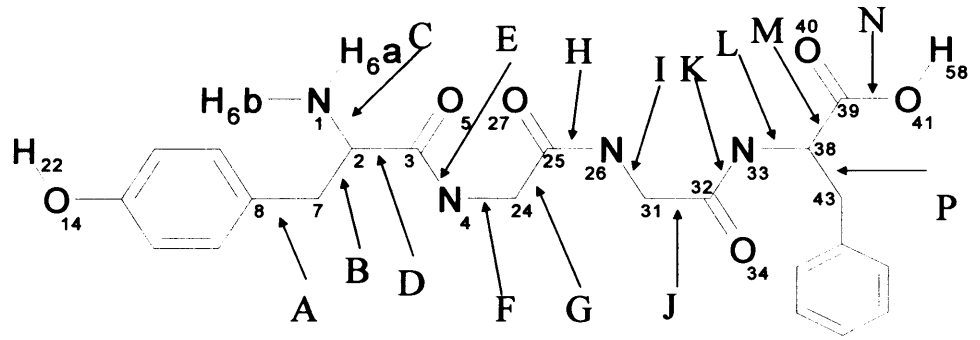
The hierarchical selection scheme developed in chapter 3 was applied to explore the conformational energy landscape of the tetrapeptide Tyr-Gly-Gly-Phe. Figure 17a shows the flexible bonds of the tetrapeptide that were considered for the procedure. Figure 17b shows the atomic labeling of the tetrapeptide with the location of the φ and ψ angles. All the electronic structure calculations were performed with the software package Gaussian (versions 98 and 03)[15, 16] and Molpro on clusters of 2.8 GHz Xeon PCs running Linux at the Chemistry Department at University College London (UCL), as well as on 1.7 GHz Power5 IBM processors included in p690+ Regatta nodes, provided by the High Performance Computing Service HPCx .

5.2.1 Test run

As in chapter 4 (for Tyr-Gly-Gly) the hierarchical selection method was first employed using large step sizes (120° and 180°) as a test run, to determine the best sizes for rotating the flexible bonds of the peptide. On the conformational energy landscape of Tyr-Gly-Gly, conformers that contain less than three hydrogen bonds were not among the most stable ones. This led us to perform single point energy calculations only for the Tyr-Gly-Gly-Phe conformers that contain three to eight hydrogen bonds. The total number of conformers considered for the single point calculations was 63360. The first 100 low-lying conformers obtained from the single point energy calculations were optimized using the HF/3-21G* method. For these, the final values of the dihedral angles of the bonds rotated were collected. The most stable conformers obtained from the test run have folded structures which contain two characteristic H-bond interactions: the first interaction is observed between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of phenylalanine whereas the second interaction is observed between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of glycine (3).

5.2.2 Hierarchical selection method

From the variation of the dihedrals of the most stable conformers resulting from the test run, the optimal step sizes for rotation of the internal bonds were determined. Thus, dihedrals D, F, I and L should be varied in steps of 30°, dihedrals G, J and M should be varied in steps of 60°, dihedral C and B should be varied in steps of 120°, and dihedrals A, E, H, K, P and N should be varied in steps of 180°. However, the number of conformers that would be created employing the hierarchical selection scheme using these step sizes was still too large. It was observed that some of the 20 most stable conformers obtained for Tyr-Gly-Gly (chapter 4) had only minor structural differences (difference in 30° for a specific angle). These conformers essentially belong to the same structure motif of the peptide and therefore they do not to need be considered in the search of the conformational energy landscape of the tetrapeptide.



A:C9C8C7C2 D:N4C3C2N1 G:N4C24C25N26 J:N26C31C32N33 M:N33C38C39O41
B:C8C7C2C3 E:C2C3N4C24 H:C24C25N26C31 K:C31C32N33C38 N:C38C39O41H58
C:H6N1C2C3 F:C3N4C24C25 I:C25N26C31C32 L:C32N33C38C39 P:C44C43C38C39

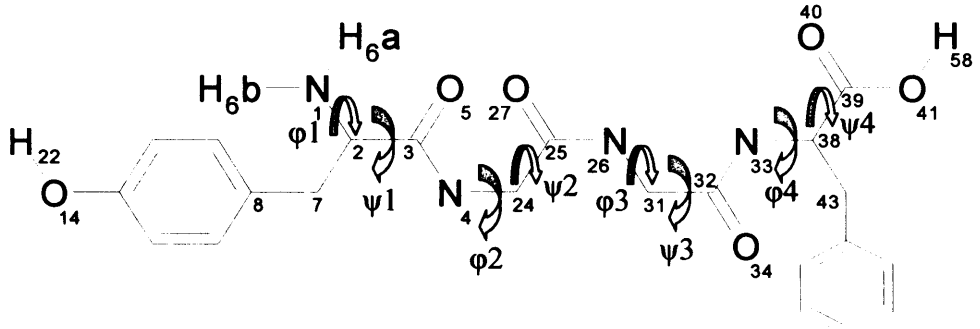


Figure 17. A. Atom labeling and definition of the dihedral angles considered in the hierarchical selection procedure.

B. Definition of the dihedral angles ω, φ, ψ.

All of the 30 most stable conformers obtained for Tyr-Gly-Gly are characterized by folded structures which contain a specific hydrogen-bonding interaction between the hydroxyl group of tyrosine and the carboxyl oxygen of glycine (3). Thus, considering the conclusions resulted from above for Tyr-Gly-Gly as well as from the conformers resulting from the test run we decided to vary dihedrals D, F, I and L in steps of 60°, dihedrals C, B, G, J, and M were varied by 120° and dihedrals A, E, H, K, P and N were varied in steps of 180°. In addition the hierarchical selection method employed considered only conformers that contain three types of specific hydrogen-bonding interactions. The first type includes an interaction between the hydroxyl hydrogen of tyrosine and the hydroxyl oxygen of phenylalanine **(O_{14}-H_{22}-O_{41})** as observed in the most stable conformers obtained from the test run. The second type includes an interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of phenylalanine **(O_{14}-H_{22}-O_{40})** as also observed in the most stable conformers obtained from the test run. The third type of interaction considered is formed between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of glycine (3) **(O_{14}-H_{22}-O_{34})** as found in the most stable Tyr-Gly-Gly conformers (chapter 4).

The total number of conformers created was 9460. Single point energy calculations were performed using the HF/3-21G* method for all the conformers. For the first 500 most stable conformers resulting from the single point energy calculations, geometry optimizations were performed using the HF/3-21G* method. The 50 most stable conformers resulting from the HF optimizations were optimized at the B3LYP[17] level of theory and the 6-31+G* basis set. The relative energies of the 20 most stable conformers according to the B3LYP/6-31+G* geometry optimizations were evaluated by single point calculations at the MP2/6-31+G* level of theory. The energies of the MP2/6-31+G* single point calculations were compared to results obtained by performing single-point energy calculations using the df-LMP2[18-21] level of theory. Zero point energies (ZPEs), scaled by 0.976[22] were calculated using B3LYP/6-31+G*.

5.3 Results

From the test run performed for the tetrapeptide Tyr-Gly-Gly-Phe it was found that folded conformers are favoured over extended structures. Thus for the hierarchical selection run only conformers with folded structures with selected H-bonding interactions were taking into consideration using the H-bond selection method. Table 8 lists the 20 most stable conformers (based on MP2 single point energy calculations corrected by B3LYP zero point energies) resulting from the H-bond selection method. Figure 18 shows the structures of the 20 most stable conformers.

All of the conformers obtained are characterized by a specific interaction between the hydroxyl group of tyrosine with either the carboxyl oxygen of glycine (3) or the carboxyl group of phenylalanine (4). These are the conformers we selected for the H-bond selection method. Most of the conformers contain an NH_2---π interaction between the N-terminus and the phenyl ring of tyrosine. Some of the conformers have only minor structural differences. For example, the only difference between the most stable and the second most stable conformer is observed at the dihedral angle ($C_{43}C_{38}N_{33}C_{32}$; the ψ_4 Ramachandran angle), which differs by about 65°. This difference changes the structural motif of the two conformers. The most stable conformer found is characterized by a strong hydrogen-bonding interaction between the hydroxyl group of tyrosine and the carboxyl group of phenylalanine (4), whereas the second most stable conformer is characterized by a strong hydrogen-bonding interaction between the hydroxyl group of tyrosine and the carboxyl group of glycine (3) (difference in energy 2.32 kJ/mol). Additionally the structures of the most stable conformer and conformer 10 are very similar. The only difference observed between these conformers concerns the dihedral angle $O_{41}C_{39}C_{38}C_{33}$ (φ_4) where the variation is about 75°. Similar differences have been observed between conformers 2-9, 11-12 and 16-19. From the observations listed above we can conclude that the phenyl group of the phenylalanine plays an important role in the conformational preferences of the tetrapeptide Tyr-Gly-Gly-Phe as the dihedral angles that determine the orientation

of this group (φ_4 and ψ_4) cause significant differences in the order of stability of the conformers (variation between conformers 1 and 2).

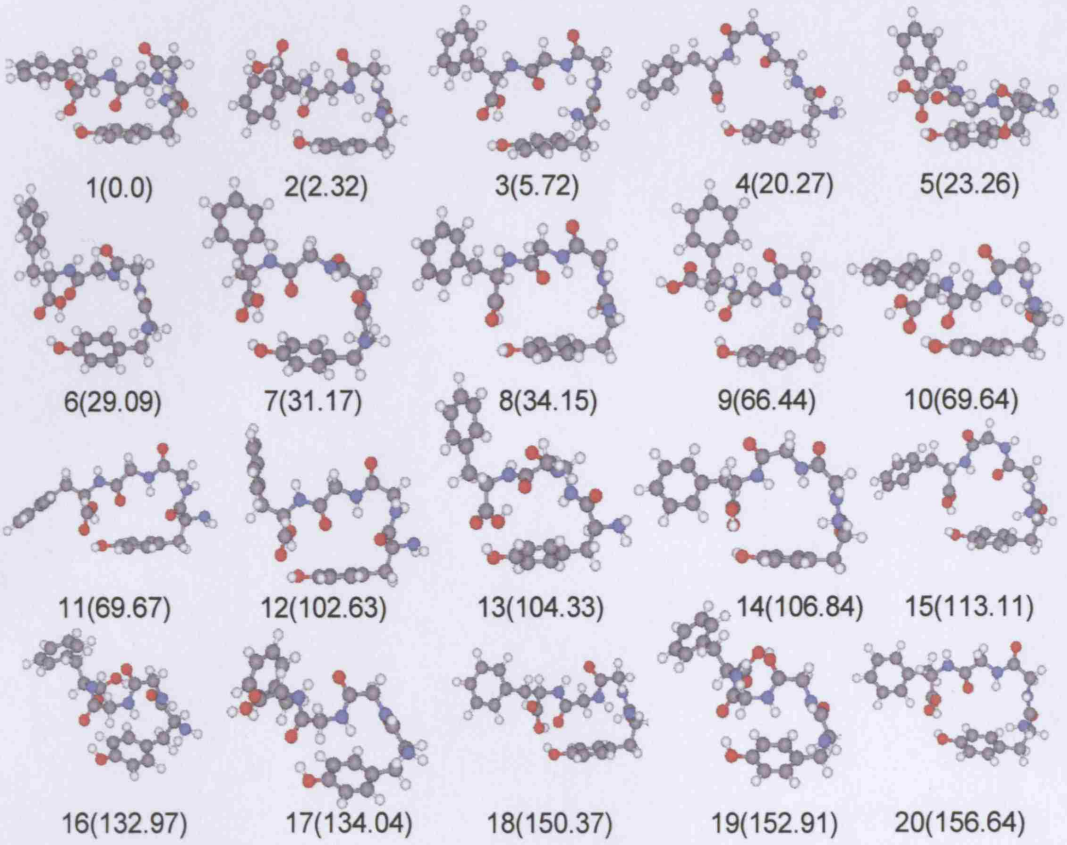


Figure 18. Relative energies of the 20 most stable conformers (from single-point MP2-6-31+G* calculations with inclusion of scaled (0.976) B3LYP/6-31+G* ZPEs) are given in kJ mol^{-1}.

As was also observed for the tripeptide Tyr-Gly-Gly, most of the conformers of Tyr-Gly-Gly-Phe have ψ_1 values close to 0° (between -2° and 30°), which allows the formation of an $N_4H_{28}\bullet\bullet\bullet N_1$ interaction. Using the ψ_1-angles we can divide the conformers into three groups (see Table 8). The first group includes conformers that have a ψ_1 angle close to 0° (between -2° and 30°) (conformers 1-3, 6, 8, 9, 10, 14-20). All these contain a strong $N_4H_{28}\bullet\bullet\bullet N_1$ interaction and also form an $NH_2\bullet\bullet\bullet\pi$ interaction between the N-terminus and the phenyl ring of tyrosine. The second group includes conformers that have a ψ_1 angle close to 45° (conformers 5, 11, 12). All these contain a strong $N_4H_{28}\bullet\bullet\bullet N_1$ interaction but they do not have the $NH_2\bullet\bullet\bullet\pi$ interaction between the N-terminus and the phenyl ring of tyrosine. The

third group includes conformers that have a ψ_1 angle close to 130° (conformers 4, 7, 13, 14). All these do not contain an $N_4H_{28}\bullet\bullet\bullet N_1$ interaction and they do not contain an $NH_2\bullet\bullet\bullet\pi$ interaction between the N-terminus and the phenyl ring of tyrosine. Instead these conformers contain an $NH\bullet\bullet\bullet\pi$ interaction between the nitrogen of glycine (2) and the phenyl ring of tyrosine.

Conformers	Structure	Hbonds	Rel. Energy (kJ/mol)	ψ1	φ2	ψ2	φ3	ψ3	φ4	ψ4	ω1	ω2	ω3
1	RP-anti	5	0.00	10.58	81.44	-70.29	-79.84	70.65	-51.38	129.08	169.83	181.84	181.84
2	RG-anti	4	2.32	11.62	85.34	-58.76	-81.43	64.83	-130.78	124.80	173.46	181.84	181.84
3	RPD-syn	4	5.72	36.74	89.21	-65.17	-84.81	175.05	-73.68	-14.53	-178.42	164.06	164.76
4	RPD-syn	4	20.27	132.23	-172.42	185.46	85.15	-71.87	-143.65	-64.63	175.51	180.06	177.89
5	RP-syn	4	23.26	-29.71	59.63	-131.11	-111.98	28.98	-57.02	136.50	-176.31	176.06	176.06
6	RPD-anti	5	29.09	6.08	79.05	-60.37	-80.89	72.50	64.89	68.04	169.19	176.06	173.67
7	RPD-anti	4	31.17	179.21	-81.92	58.54	91.37	163.27	52.40	46.39	-164.28	178.36	174.67
8	RPD-syn	4	34.15	16.57	117.16	-9.33	-172.19	-167.91	-65.59	-20.77	175.44	178.94	169.22
9	RG-anti	4	66.44	11.86	86.32	-60.10	-81.31	63.62	-77.87	168.75	173.60	186.27	176.24
10	RP-anti	5	69.64	10.03	81.25	-70.45	-79.94	71.02	-43.88	-54.56	169.60	186.27	186.27
11	RPD-syn	4	69.67	45.26	-83.72	71.46	-179.67	-165.87	-69.98	-43.81	177.25	186.27	176.23
12	RPD-syn	4	102.63	46.59	-85.14	66.84	177.02	-171.89	-61.31	-39.11	178.82	186.27	186.27
13	RP-OHO-syn	5	104.33	125.76	-80.53	58.80	79.16	-59.66	-73.64	67.42	-166.62	166.79	186.27
14	RPD-anti	5	106.84	12.83	138.12	-16.43	130.50	-16.95	-124.38	79.26	169.91	165.44	186.27
15	RPD-syn	4	113.11	4.58	120.91	111.49	82.27	-78.17	-157.06	-47.10	175.10	173.51	186.27
16	RG-anti	5	132.97	-0.32	104.02	0.63	84.44	-61.56	-79.23	-50.63	171.12	186.58	152.75
17	RG-anti	4	134.04	-1.83	114.63	-12.25	80.73	-67.90	58.83	-142.26	179.74	177.55	181.92
18	RP-OHO-anti	5	150.37	9.82	80.43	-58.49	-78.41	60.65	66.68	-59.52	169.27	189.00	169.90
19	RG-anti	4	152.91	-2.23	113.68	-9.98	82.62	-64.71	-109.98	-61.94	178.99	178.05	178.08
20	RPD-anti	4	156.64	11.60	124.46	-40.14	164.98	161.87	48.27	51.82	168.66	185.37	179.13

Table 8. Number of hydrogen-bonding interactions, relative energies ΔE (in kJ mol^{-1}) based on single point MP2/6-31+G* calculations with inclusion of scaled (0.976) B3LYP/6-31+G* zero-point energies; values of the Ramachandran angles ψ_1, φ_2, ψ_2, φ_3, ψ_3, φ_4, ψ_4, ω_1, ω_2, ω_3 (in degrees), and structural elements of the most stable conformers, obtained using the hierarchical conformational analysis method.

ψ_1=τ (N1-C2-C3-N4), φ2= τ (C25-C24-N4-C3), $\psi 2$= τ (N26-C25-C24-N4), φ3= τ (C32-C31-N26-C25), $\psi 3$= τ (N33-C32-C31-N26), φ4= τ (C39-C38-N33-C32), $\psi 4$= τ (O41-C39-C38-N33), ω1= τ (C2-C3-N4-C24), ω2= τ (C25-C24-N26-C31), ω3= τ (C31-C32-N3-C38),RP: OH•••O interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of phenylalanine (4). RG: OH•••O: interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of glycine (3). RPD: Double OH•••O hydrogen-bonding interaction between the hydroxyl hydrogen of tyrosine (1) and the carbonyl oxygen of phenylalanine (4) and between the hydroxyl hydrogen of tyrosine (1) and the hydroxyl hydrogen of phenylalanine (4). RP-OHO: RP structure with an additional O-H•••O interaction between the hydroxyl group of phenylalanine (4) and the carboxyl group of glycine (3).

These groups can be alternatively grouped into four different groups according to the hydrogen-bonding interactions observed in the conformers. The first group contains conformers that are characterized by a hydrogen-bonding interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of phenylalanine (4) (conformers 5 and 10) or the hydroxyl oxygen of phenylalanine (4) (conformer 1) (RP as indicated in table 8). The second group contains conformers that contain a strong interaction between the hydroxyl hydrogen of tyrosine and the carbonyl oxygen of glycine (3) (conformers 2, 9, 16, 17 and 19) (RG as indicated in table 8). The third group contains conformers that consist of two OH•••O hydrogen-bonding interactions. The first interaction is between the hydroxyl hydrogen of tyrosine and either the carboxyl oxygen of phenylalanine (4) (conformer 13) or the hydroxyl oxygen of phenylalanine (4) (conformer 18) (RP-OHO as indicated in table 8). Finally the fourth group consist of conformers that contain a double OH•••O hydrogen-bonding interactions between the hydroxyl hydrogen of tyrosine (1) and the carboxyl oxygen of phenylalanine (4) and between the hydroxyl hydrogen of phenylalanine (4) and the hydroxyl oxygen of tyrosine (1) (RPD as indicated in table 8) (conformers 3-4, 6-8, 11-12, 14-15, 20).

Figure 19 shows the Ramachandran distribution of the glycine (2), glycine (3) and phenylalanine (4) residues for the 48 conformers that were identified from the B3LYP/6-31+G* geometry optimizations. As we can see from table 8 all of the 20 most stable conformers are γ-type (φ_2 and ψ_2 angles are around 60°/+60° or -60°/+60°). From figure 19a the absence of α-type conformers for the most stable conformers (with angles +60°/+60° or -60°/-60°) can be observed. It is explained as if conformers adopted angles +60°/+60° or -60°/-60° this would result in an extended structure (we only collected folded structures thus extended conformers were not obtained). Similar patterns are observed from the Ramachandran plot of φ_3 against ψ_3 angle for the tetrapeptide in figure 19b. Figure 20c shows the distribution of the φ_4 and ψ_4 angles for the tetrapeptide. As can be observed for the 48 conformers obtained these are conformers that have φ_4/ψ_4 combinations of +60°/+60° and -60°/-60°. The conformers that have these combinations form the RPD conformers where the COOH group of phenylalanine (4) interacts with the hydroxyl group of tyrosine (1) forming two H-bonding interactions.

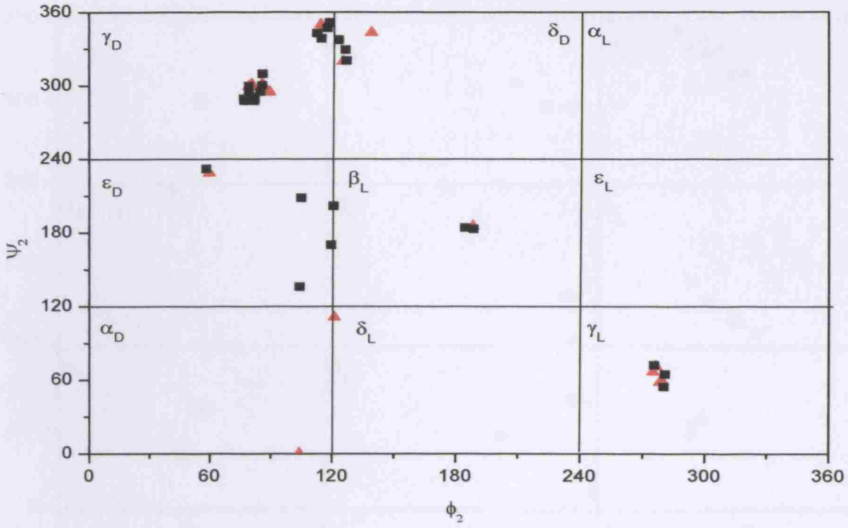


Figure 19a. Ramachandran plot showing the φ_2/ψ_2 combinations occurring in the 48 Tyr-Gly-Gly-Phe conformers identified and optimized with B3LYP/6-31+G* in the current work. The points belonging to the twenty most stable conformers are indicated by red colour.

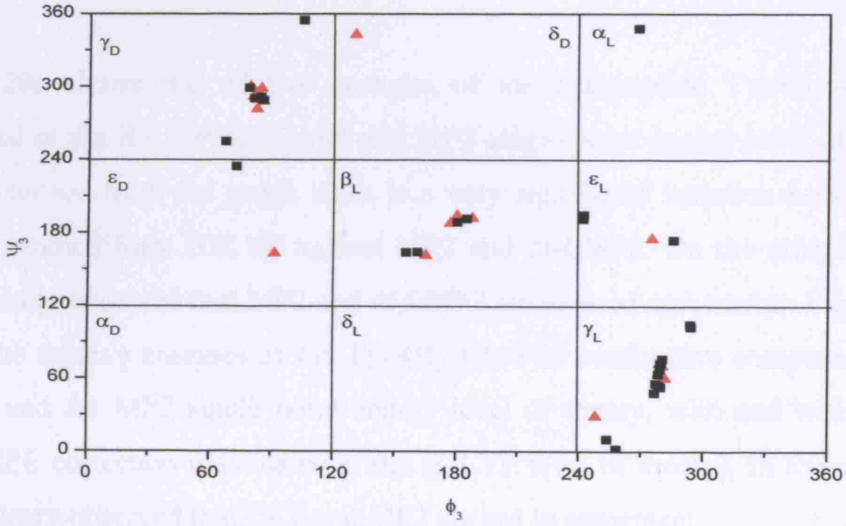


Figure 19b. Ramachandran plot showing the φ_3/ψ_3 combinations occurring in the 48 Tyr-Gly-Gly-Phe conformers identified and optimized with B3LYP/6-31+G* in the current work. The points belonging to the twenty most stable conformers are indicated by red colour.

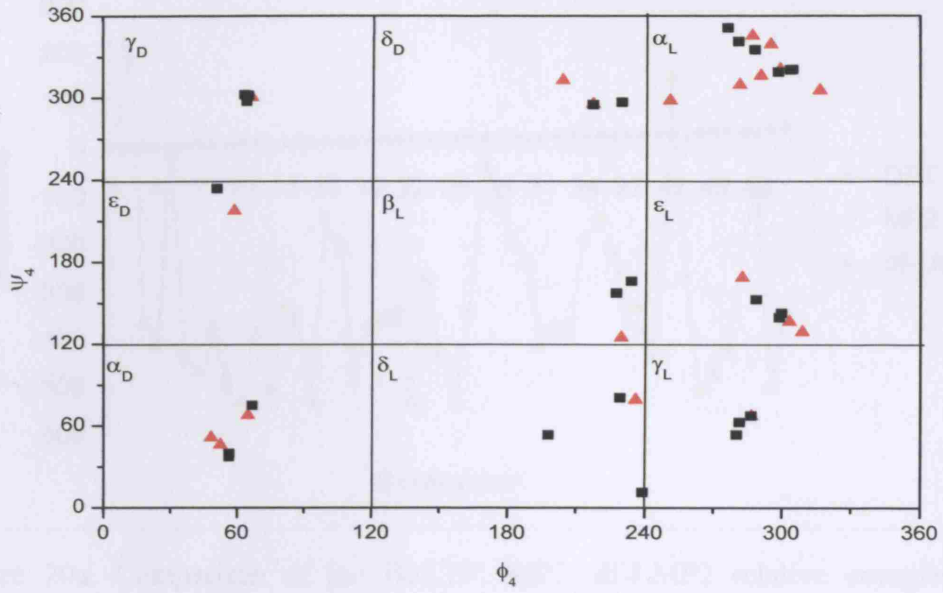


Figure 19c. Ramachandran plot showing the φ_4/ψ_4 combination occurring in the 48 Tyr-Gly-Gly-Phe conformers identified and optimized with B3LYP/6-31+G* in the current work. The points belonging to the twenty most stable conformers are indicated by red colour.

Figure 20a shows the relative energies of the tetrapeptide Tyr-Gly-Gly-Phe computed at the B3LYP, MP2 and df-LMP2 single point energy level of theory. As we can see from the graph there is a very significant variation between the results obtained from B3LYP against MP2 and df-LMP2. On the other hand, it can be easily observed that MP2 and df-LMP2 are in good agreement. Figure 20b shows the relative energies of the Tyr-Gly-Gly-Phe conformers computed at the B3LYP and the MP2 single point energy level of theory, with and without the scaled ZPE correction (calculated at the B3LYP level of theory). In this graph it can be cleary observed that DFT and MP2 are not in agreement.

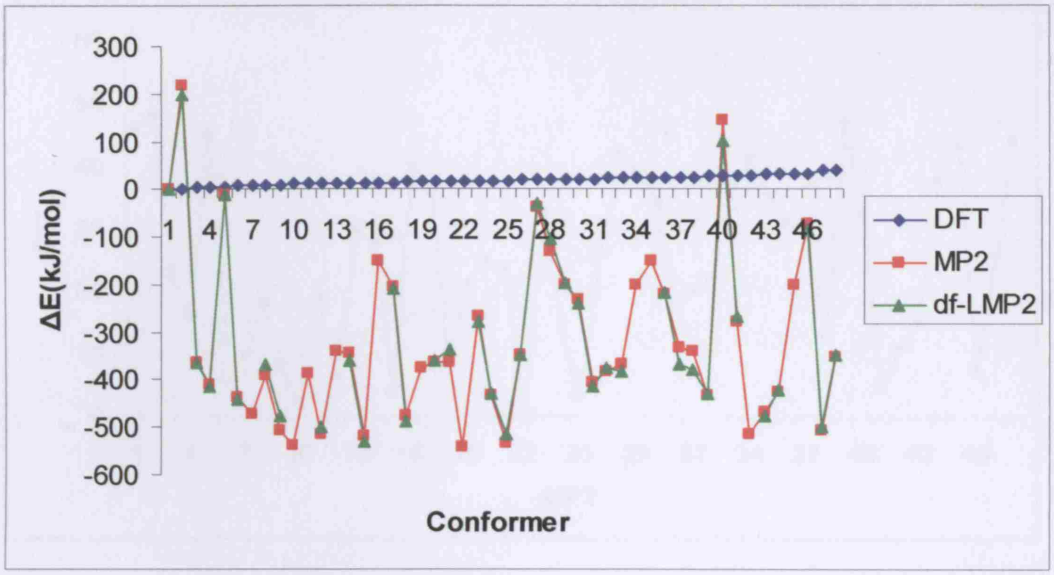


Figure 20a. Comparison of the B3LYP, MP2, df-LMP2 relative energies in (kJ/mol) of the 48 most stable Tyr-Gly-Gly-Phe conformers (based on the DFT geometries).

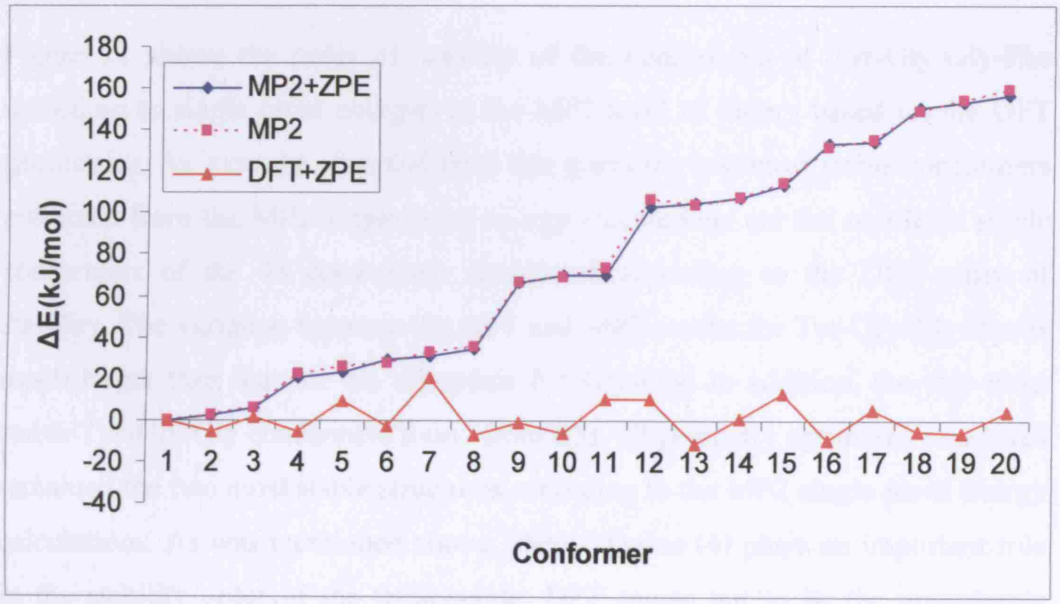


Figure 20b. Comparison of the B3LYP, B3LYP+ZPE, MP2+ZPE relative energies in (kJ/mol) of the twenty most stable Tyr-Gly-Gly-Phe conformers (based on the DFT geometries). The ZPEs were computed with B3LYP/6-31+G* and scaled by 0.976.

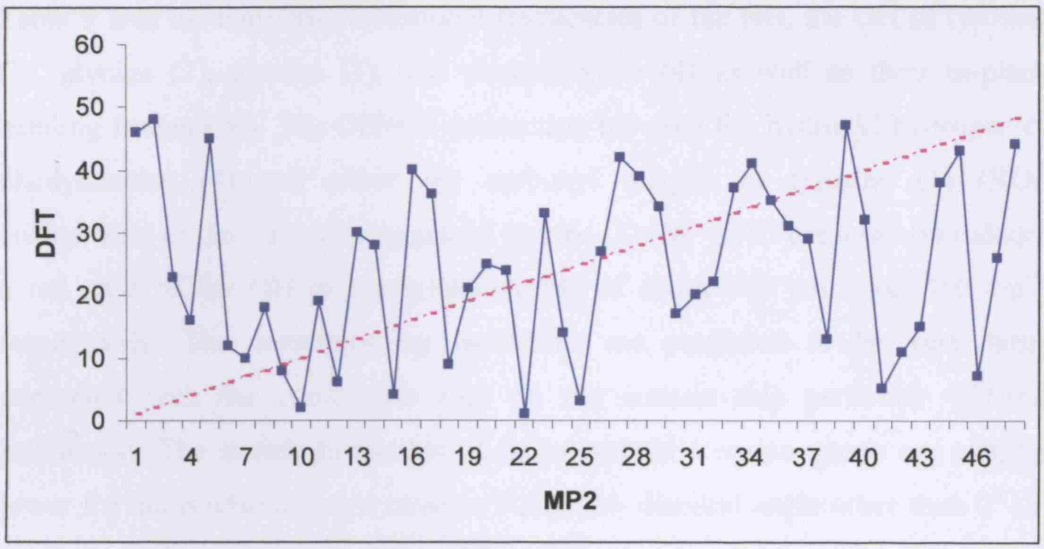


Figure 21. Relative stability of the 20 most stable conformers found for Tyr-Gly-Gly-Phe with B3LYP and MP2 level of theory. The dotted line indicates ideal agreement.

Figure 21 shows the order of stability of the conformers of Tyr-Gly-Gly-Phe according to single point energies at the MP2 level of theory based on the DFT geometries. As it can be observed from this graph the two most stable conformers predicted from the MP2 single point energy calculations are the two least stable conformers of the 48 conformers considered according to the DFT order of stability. The variation between the DFT and MP2 results for Tyr-Gly-Gly-Phe is much larger than that for the tripeptide Tyr-Gly-Gly. In addition, the two most stable Tyr-Gly-Gly conformers found from B3LYP geometry optimized structures remained the two most stable structures according to the MP2 single point energy calculations. As was mentioned above, phenylalanine (4) plays an important role in the stability order of the tetrapeptide. DFT seems not to be the appropriate method to describe the tetrapeptide Tyr-Gly-Gly-Phe, as the dispersion effects arising from the two phenyl groups may be important and this probably causes this large variation between the DFT and MP2 results obtained for Tyr-Gly-Gly-Phe.

Table 9 lists the harmonic vibrational frequencies of the NH, the OH of tyrosine (1), glycine (2), glycine (3), and phenylalanine (4) as well as their in-plane bending frequencies. The OH•••O interaction between the hydroxyl hydrogen of phenylalanine (4) and either the carboxyl oxygen of tyrosine (1) (RDP conformers) or the carboxyl oxygen of glycine (3) (RP-OHO conformers) induces a red shift of the OH of phenylalanine (4) of about 200 cm^{-1} and 450 cm^{-1}, respectively. The corresponding intensities are predicted to be very large compared with the conformers that do not contain this particular OH•••O interaction. The stretch intensities of the asymmetric amino group are slightly lower for the conformers that have an $N_4C_3C_2N_1$ dihedral angle other than 0° i.e. conformers that do not contain an NH_2•••π interaction between the N-terminus and the phenyl ring of tyrosine.

	Structure	Gly(2)NHB	Phe(4)NHB	Gly(3)NHB	Tyr(1)C=OB	Gly(2)C=OB	NH2(SS)	NH2(AS)
1	RP-anti	1556 (173)	1584 (250)	1593 (174)	1707 (68)	1725 (222)	3476 (7)	3571 (16)
2	RG-anti	1559 (205)	1593 (290)	1586 (135)	1714 (65)	1732 (553)	3479 (5)	3571 (12)
3	RPD-syn	1559 (147)	1541 (323)	1580 (208)	1724 (155)	1754 (403)	3472 (10)	3568 (12)
4	RPD-syn	1531 (393)	1565 (243)	1571 (169)	1722 (300)	1740 (19)	3465 (1)	3549 (3)
5	RP-syn	1560 (301)	1554 (70)	1576 (232)	1722 (366)	1739 (295)	3481 (0)	3563 (3)
6	RPD-anti	1565 (208)	1589 (190)	1606 (168)	1707 (77)	1726 (196)	3482 (7)	3577 (13)
7	RPD-anti	1550 (106)	1551 (192)	1600 (206)	1715 (143)	1748 (342)	3481 (2)	3556 (6)
8	RPD-syn	1533 (632)	1551 (253)	1566 (132)	1747 (214)	1739 (471)	3475 (3)	3567 (10)
9	RG-anti	1554 (140)	1585 (311)	1588 (206)	1711 (76)	1731 (716)	3476 (4)	3570 (30)
10	RP-anti	1559 (106)	1588 (288)	1594 (184)	1706 (66)	1724 (251)	3475 (16)	3570 (10)
11	RPD-syn	1554 (263)	1538 (423)	1562 (80)	1728 (71)	1741 (562)	3471 (0)	3559 (2)
12	RPD-syn	1554 (243)	1538 (457)	1561 (28)	1728 (130)	1740 (523)	3470 (0)	3558 (2)
13	RP-OHO-syn	1565 (168)	1598 (309)	1603 (146)	1732 (246)	1716 (253)	3478 (3)	3567 (5)
14	RPD-anti	1539 (43)	1545 (332)	1558 (103)	1752 (17)	1764 (321)	3475 (3)	3566 (17)
15	RPD-syn	1559 (269)	1569 (375)	1551 (148)	1753 (502)	1735 (134)	3480 (6)	3574 (14)
16	RG-anti	1568 (233)	1589 (288)	1548 (124)	1725 (216)	1720 (138)	3489 (9)	3581 (17)
17	RG-anti	1544 (188)	1607 (261)	1569 (149)	1750 (201)	1710 (55)	3483 (6)	3578 (14)
18	RP-OHO-anti	1559 (132)	1614 (238)	1602 (162)	1704 (85)	1725 (451)	3476 (10)	3573 (17)
19	RG-anti	1543 (200)	1595 (362)	1567 (142)	1751 (191)	1727 (587)	3482 (6)	3577 (14)
20	RPD-anti	1565 (77)	1561 (232)	1538 (457)	1743 (277)	1738 (369)	3477 (4)	3570 (9)

Table 9a. Scaled (OH: 0.976; NH: 0.956) harmonic vibrational frequencies (in cm^{-1}) of the OH and NH in-plane bending modes and stretching modes of the two hydrogen atoms of the N-terminus (NH2(SS) and NH2(AS)). Calculated intensities (km mol^{-1}) are given in round brackets. SS: Symmetric stretching mode, AS: Assymetric stretching mode, C=OB: In plane-bending mode of carboxyl oxygen of a peptide bond. NHB: In plane-bending mode of the amide hydrogen of a peptide bond.

	Structure	Gly(3)C=O	Phe(4)C=O	Phe(4)OH	Gly(3)NH	Tyr(1)OH	Phe(4)NH	Gly(2)NH
1	RP-anti	1748 (432)	1840 (273)	3680 (56)	3466 (208)	3667 (332)	3485 (152)	3538 (105)
2	RG-anti	1709 (177)	1819 (215)	3668 (59)	3508 (137)	3554 (391)	3449 (185)	3545 (86)
3	RPD-syn	1768 (210)	1779 (236)	3430 (550)	3548 (93)	3674 (129)	3574 (50)	3550 (69)
4	RPD-syn	1752 (402)	1788 (240)	3375 (738)	3609 (38)	3568 (489)	3539 (98)	3577 (75)
5	RP-syn	1766 (173)	1796 (360)	3681 (63)	3612 (21)	3653 (290)	3525 (178)	3580 (93)
6	RPD-anti	1743 (232)	1792 (368)	3522 (74)	3448 (291)	3567 (660)	3496 (165)	3536 (109)
7	RPD-anti	1767 (91)	1781 (399)	3455 (336)	3490 (247)	3561 (549)	3624 (24)	3628 (22)
8	RPD-syn	1755 (150)	1778 (242)	3392 (722)	3588 (85)	3640 (211)	3573 (670)	3536 (77)
9	RG-anti	1719 (96)	1834 (215)	3684 (68)	3513 (151)	3572 (395)	3503 (118)	3545 (93)
10	RP-anti	1752 (338)	1792 (376)	3685 (80)	3457 (213)	3566 (585)	3471 (149)	3535 (105)
11	RPD-syn	1766 (91)	1793 (257)	3411 (529)	3514 (116)	3579 (427)	3596 (21)	3602 (65)
12	RPD-syn	1758 (91)	1781 (325)	3413 (600)	3501 (126)	3592 (381)	3600 (63)	3610 (43)
13	RP-OHO-syn	1691 (211)	1806 (366)	3195 (550)	3446 (285)	3586 (569)	3424 (225)	3626 (29)
14	RPD-anti	1752 (608)	1772 (109)	3444 (423)	3588 (76)	3571 (461)	3589 (27)	3525 (75)
15	RPD-syn	1746 (105)	1794 (122)	3383 (783)	3628 (32)	3645 (218)	3526 (128)	3536 (87)
16	RG-anti	1735 (474)	1813 (266)	3492 (127)	3587 (82)	3534 (512)	3463 (510)	3502 (97)
17	RG-anti	1726 (573)	1820 (323)	3681 (71)	3594 (117)	3467 (717)	3489 (275)	3528 (107)
18	RP-OHO-anti	1689 (157)	1842 (272)	3145 (625)	3432 (268)	3625 (534)	3413 (371)	3529 (117)
19	RG-anti	1711 (73)	1831 (272)	3681 (63)	3590 (121)	3497 (739)	3458 (156)	3531 (111)
20	RPD-anti	1755 (128)	1790(242)	3421 (466)	3579 (88)	3557 (598)	3618 (31)	3534 (85)

Table 9b. Scaled (OH: 0.976; NH: 0.956) harmonic vibrational frequencies (in cm^{-1}) of the OH and NH stretching modes. Calculated intensities (km mol^{-1}) are given in round bracket.

C=O: Stretching mode of carboxyl oxygen of a peptide bond. NH: Stretching mode of the amide hydrogen of a peptide bond.

5.4 Discussion

In this chapter we applied the hierarchical selection method to explore the conformational features of the enkephalin-related tetrapeptide Tyr-Gly-Gly-Phe. In the two previous chapters we explored the conformation features of the dipeptide Tyr-Gly (chapter 3) and the tripeptide Tyr-Gly-Gly (chapter 4). In the second most stable Tyr-Gly conformer based on MP2 geometry optimizations was observed a tendency for the peptide backbone to interact with the hydroxyl group of tyrosine ("book" conformers)[23]. Additionally the most stable structures found for the tripeptide Tyr-Gly-Gly are characterized by an OH•••O interaction between the hydroxyl group of tyrosine and the carboxyl oxygen of glycine (3). From the test run performed for the tetrapeptide it was found that folded conformers are preferred over extended conformers. The most stable conformers are characterized by a strong interaction between the hydroxyl group of tyrosine and the carboxylic group of phenylalanine (4) (conformer 1) and between the hydroxyl group of tyrosine and the carboxylic group of glycine (3) (conformer 2).

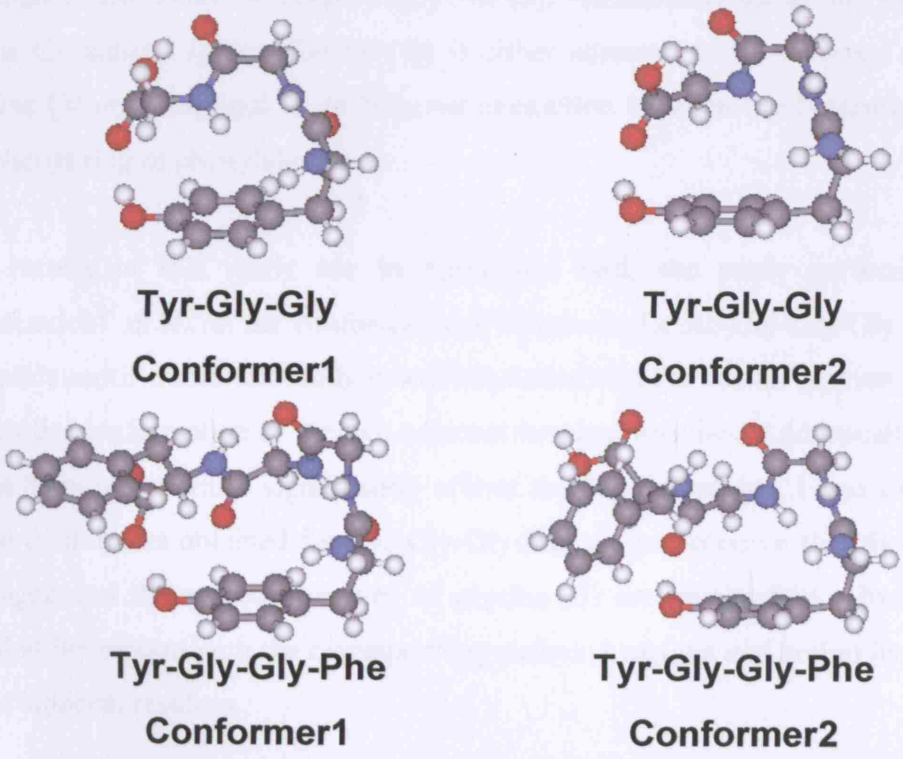


Figure 22. The first and second most stable conformer found for the Tyr-Gly-Gly tripeptide and the Tyr-Gly-Gly-Phe tetrapeptide.

Figure 22 shows the first and the second most stable conformers found for Tyr-Gly-Gly and Tyr-Gly-Gly-Phe. It was found that the dihedral angles of the flexible bonds of the Tyr-Gly-Gly part of the second most stable Tyr-Gly-Gly-Phe conformer are similar to those of the second most stable Tyr-Gly-Gly conformer. The major difference between the most stable and the second most stable Tyr-Gly-Gly conformer is observed in the orientation of the peptide backbone (differences in the value of $\varphi(gly)_2$ of 30° and $\psi(gly)_2$ angles of 60°) (see chapter 4), whereas the only difference between the most stable and the second most stable Tyr-Gly-Gly-Phe conformer is a change of about 65° degrees in the ψ_4 angle ($C_{43}C_{38}N_{33}C_{32}$).

Zaluski et al performed an X-ray and ^1H NMR spectroscopy study for the tetrapeptide Tyr-Gly-Gly-Phe[8]. In this spectroscopic study a highly stable folded β-bend conformation with a strong hydrogen-bonding interaction between the carboxyl oxygen of the tyrosine (1) residue and the amino hydrogen of phenylalanine (4) was observed. The most stable conformers found from our computational study for Tyr-Gly-Gly-Phe show that the carboxyl oxygen is attracted to the amino hydrogen of glycine (3). Additionally the amino hydrogen of the C-terminus (phenylalanine (4)) is either attracted to the carboxyl oxygen glycine (3) or is engaged in an $NH_2 \bullet\bullet\bullet \pi$ interaction between the N-terminus and the phenyl ring of phenylalanine.

The results of this study are in agreement with the study performed by Mehdizadeh[24] et al. on the conformational effects in the Ac-Gly-Gly-Gly-NHMe tripeptide motif. From this study it was concluded that the central glycine residue affects the conformation of the two adjacent terminal residues. Additionally, each of the terminal residues significantly affects the central residue. From the most stable conformers obtained for Tyr-Gly-Gly-Phe we can observe that the amino hydrogen and the carboxyl oxygen of glycine (3) are involved in a hydrogen-bonding interaction with the corresponding carboxyl oxygen and amino hydrogen of the adjacent residues.

The results obtained from this study can be compared to the results that have been published for enkephalins. Previous conformational studies on enkephalins predicted stable conformations that are described by characteristic hydrogen-bonding interactions between the C=O carboxyl of Phe (4) and the hydrogen amide of Gly (2) using NMR spectroscopy and Molecular Dynamics calculations both in neutral and zwitterionic form[25] and calculations with SCC-DFTB[26] method in the gas phase[27]. However, other studies have predicted stable structures that contain a strong interaction between the hydroxyl group of tyrosine and either the carboxyl oxygen of glycine (3) and phenylalanine (4) indicating that the tyrosine and phenylalanine residues are the key elements essential for the particular bioactivity of the enkephalin molecule[28-30]. The results of these studies are in agreement with our study as we found that the tyrosine hydroxyl group plays an important role in the conformational preferences of the peptide.

For Tyr-Gly-Gly we observed a disagreement between the order of stability predicted by DFT and MP2 for the most stable conformers. However the two most stable conformers predicted by DFT remain the two most stable conformers according to the MP2 single point energy calculations. Large differences were observed for the extended conformers that were among the first 20 most stable conformers predicted by DFT, which were replaced by folded structures after performing MP2 single point calculations (see chapter 4). In addition for the tetrapeptide large variations exist between the relative energies predicted by DFT and MP2 (see figure 20). The large variation in the relative energies between these two methods implies that geometry optimizations at the MP2 level of theory would change significantly the geometry of the conformers.

From the current study we can conclude that DFT is a not sufficiently accurate method to predict the conformational features of the tetrapeptide. Geometry calculations at the MP2 level of theory may be needed to characterize the conformational features of these peptides, whereas large basis sets will be needed in order to reduce the basis set superposition error. However, the flexibility of the peptides results in a large number of possible conformers and the absence of experimental guidance make it nearly impossible to find all the most stable conformers.

It is clear that in order to fully explore the conformational preferences of such peptides a combination of approaches should be applied that include high-level quantum chemical methods that can calculate the dispersion arising from interactions with the aromatic residues of the peptide and that do not produce much basis set superposition error. In the next chapter we explore the current method for a larger (penta) peptide, for which the number of possible conformers is much larger and thus, the conclusions derived from this study will be used in order to describe the conformational features of this peptide.

References

1. Hayashi, T.; Asai, T.; Ogoshi, H., *Tetrahedron, Letters* **1997,** 17, 3039.

2. Perczel, A.; Hollosi, M.; Foxman, B. M.; Fasman, G. D., *J. Am. Chem. Soc.* **1991,** 113, 9772.

3. Kessler, H.; Anders, U.; Schudok, M., *J. Am. Chem. Soc.* **1990,** 112, 5908.

4. Sibanda, B.; Thornton, J. M., *Nature* **1985,** 316, 170.

5. Wuthrich, K.; Billeter, M.; Braun, W. J., *J. Mol. Biol.* **1984,** 180, 715.

6. Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E., *J. Mol. Biol.* **1988,** 201, 201.

7. Garbay-Jaureguiberry, C.; Baudet, J.; Florentin, D.; Roques, B. P., *FEBS Letters* **1980,** 115, (2), 315.

8. Fournie-Zaluski, M. C.; Prange, T.; Pascard, C.; Roques, B. P., *Biochem. Biophys. Res. Commun.* **1977,** 79, 1199.

9. Prabhu, N. V.; Perkyns, J. S.; Blatt, H. D.; Smith, P. E.; Petit, B. M., *Biophys. Chem.* **1999,** 5, 113.

10. Daniele, P. G.; Prenesti, E.; Zerbinati, O.; Aigotti, R.; Ostacoli, G., *Spectrochimica Acta Part A: Molec. Spectr.*, **1993,** 49, (9), 1373.

11. Becke, O. M., *J. Mol. Struct.* **1997,** 398, 507.

12. Gresh, N.; Kafafi, S. A.; Trunchon, J. F.; Salahub, R. D., *J. Comp. Chem.* **2003,** 25, (6), 823.

13. Garmer, D. R.; Stevens, W. J., *J. Phys. Chem.* **1989,** 93, (8263).

14. Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, A., *J. Am. Chem. Soc.* **1997,** 119, 5908.

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

Gordon, M.; Replogle, E. S.; and Pople, J. A., Gaussian, Inc., Pittsburgh, PA, 1998.

16. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, J., T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, Q. C., J. V.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A., Gaussian, Inc., Pittsburgh PA, 2003.

17. Becke, A. D., *J. Chem. Phys.*, **1993,** 98, 5648-5662.

18. Hetzer, G.; Pulay, P.; Werner, H.-J., *Chem. Phys. Lett.* **1998,** 290, 143.

19. Hetzer, G.; Schutz, M.; Stoll, H.; Werner, H.-J., *J. Chem. Phys.*, **2000,** 113, 9443.

20. Schutz, M.; Hetzer, G.; Werner, H.-J., *J. Chem. Phys.*, **1999,** 111, 5691.

21. Werner, H.-J.; Manby, F. R.; Knowles, P. J., *J. Chem. Phys.*, **2003,** 118, 8149.

22. Caminati, W.; Di Bernardo, S., *J. Mol. Struct.* **1990,** 240, 253.

23. Toroz, D.; van Mourik, T., *Molec. Phys.* **2006,** 104, 559-570.

24. Mehdizadeh A., C., G.A, Farkas, O., Perczel A., Torday, L.L., Varro A., Papp, G.J., *J. Mol. Struct.* **2002,** 588, 187-200.

25. Abdali, S.; Jensen, M.; Bohr, H., *J.Phys. Condens. Matter* **2003,** 15, 1853-1860.

26. Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E., *J. Chem. Phys.* **2001,** 114, (12), 5149-5155.

27. Abdali, S.; Niehus, T. A.; Jalkanen, K. J.; Cao, X.; Nafie, L. A.; Frauenheim, T.; Suhai, S.; Bohr, H., *Phys. Chem. Chem. Phys.* **2003,** 5, 1295-1300.

28. De Coen, J.; Humblet, C.; Koch, M., *FEBS Lett.* **1977,** 73, 38-42.

29. Perez, J.; Villar, H.; Loew, G. H., *J. Comput. Aided Mol. Des.* **1992,** 6, (2), 175-190.

30. Kriz, Z.; Koca, J.; Carlsen, P. H. J., *J. Mol. Struct. (Theochem)* **2001,** 540, 231-250.

6

THE STRUCTURE OF THE TYR-GLY-GLY-PHE-LEU (LEU-ENKEPHALIN) PENTAPEPTIDE

6.1 Introduction

In the previous chapter the hierarchical selection method was applied to explore the conformational features of the enkephalin fragment Tyr-Gly-Gly-Phe. In the current chapter we apply the hierarchical selection method to the gas-phase pentapeptide Tyr-Gly-Gly-Phe-Leu (Leu-enkephalin).

Several studies have been published presenting algorithmic approaches in order to perform a conformational search for biological molecules[1-12]. Hansmann et al. has published a study based on a parallel tempering algorithm that explores the conformational features of biological molecules[13]. The method uses both Monte Carlo and molecular mechanics approaches. Meirovitch et al. published a method for the conformational search of macromolecules[14]. With this method an initial set of relatively low-energy conformers is generated, and their energies are further minimized with a procedure that enables escaping from local energy minima[15]. The method was applied to the Met- and Leu-enkephalin pentapeptides. A systematic stepwise variation method to overcome the multiple minima problem was published by Klein et al[16]. In this method large step sizes were applied for the rotation of the flexible bonds of the peptides. Only chosen dihedral angles were varied with small step sizes. For the generated conformers the GROMOS-87 force field[17] with a distance-dependent dielectricity constant[18] and the ECEPP/3 force field[19] were used.

Rayan et al. published a stochastic search method to explore the conformational space of cyclic peptides[20]. The method is based on generation of random conformations which are evaluated by a function for ring closure ability. The structures obtained were optimized with the Kollman force field[21]. Santagata et al. published a geometrical algorithm to search the conformational space (GASCOS)

and create all possible conformations of a flexible molecule[22]. This approach is primarily based on mathematical approaches. The program generates inputs for force field, semi-empirical or ab initio molecular computations at the Hartree-Fock level of theory. Most of the methods discussed above include calculations that use force field and molecular mechanics calculations. In the hierarchical selection method developed in this thesis only electronic structure theory methods are used, including correlated methods, such as B3LYP geometry optimizations and single-point energy calculations at the MP2 level of theory.

Varying all the flexible bonds of the peptide with small step sizes would result in such a large number of conformers that is impossible to be created for Leu-enkephalin. From the literature review discussed in chapter 2 it was observed that tyrosine (1) plays an important role in the stability of the conformers[23]. The most stable conformers found for Leu-enkephalin involved characteristic strong hydrogen-bonding interactions between the hydroxyl hydrogen of tyrosine (1) and either the carboxyl oxygen of glycine (3) or the phenylalanine (4) residue[24, 25]. Additionally, Snoek provided us with experimental data of Leu-enkephalin. The IR spectra show the presence of an OH•••O interaction between the hydrogen of the hydroxyl group of tyrosine (1) and a carboxyl group of the peptide backbone[26]. From the calculations performed after the test run for the tetrapeptide it was observed that the 20 most stable conformers of this peptide are all folded structures.

The conclusion made from the study on the most stable conformers of the tetrapeptide and the experimental data provided lead us to the conclusion that the most stable conformers of the pentapeptide Tyr-Gly-Gly-Phe-Leu will likely be folded structures that are characterized by an interaction between tyrosine (1) and a carboxyl oxygen of the peptide backbone. Thus, in the scan of the potential energy surface for the pentapeptide, the hierarchical selection method was applied collecting only conformers that form the characteristic OHO interaction between the hydroxyl hydrogen of tyrosine (1) and the carboxyl oxygen of glycine (3), or the carboxyl oxygen of phenylalanine (4), or the carboxyl oxygen of leucine (5).

6.2 Methodology

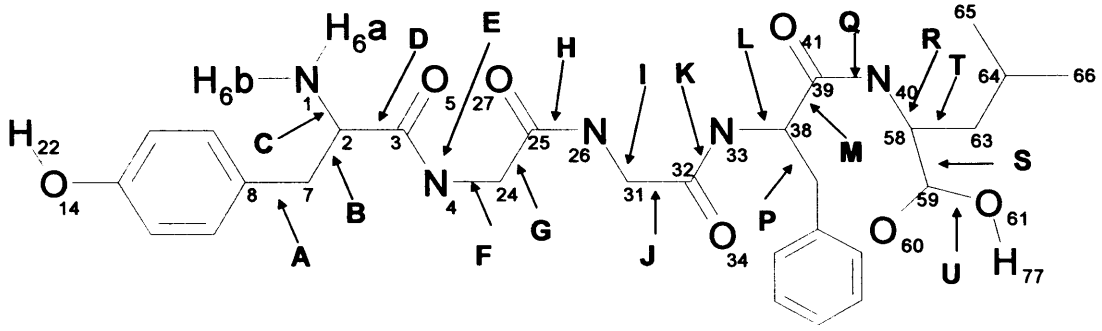
The hierarchical selection scheme developed in chapter 3 was applied to explore the conformational energy landscape of the pentapeptide Tyr-Gly-Gly-Phe-Leu (Leu-enkephalin). The flexible bonds of the pentapeptide that were considered for the procedure are presented in figure 23a. Figure 23b shows the atomic labeling of the pentapeptide with the location of the φ and ψ angles. All the electronic structure calculations were performed with the software package Gaussian[27] (version 03) on clusters of 2.8 GHz Xeons running Linux at the Chemistry Department at University College London (UCL), as well as on 1.7 GHz Power5 IBM processors included in p690+ Regatta nodes, provided by the High Performance Computing Service HPCx .

6.2.1 Hierarchical selection run

In chapter 5 the test run of the hierarchical selection method was applied using large step sizes (120° and 180°) in order to obtain the best step sizes for rotating the flexible bonds of Tyr-Gly-Gly-Phe. The optimal step sizes were determined from the dihedral angle variation of the most stable conformers obtained from the test run. From the test run performed for the tetrapeptide it was observed that folded conformers are preferred over extended conformers. Thus, for the tetrapeptide, in the hierarchical selection scheme only conformers that form the specific hydrogen-bonding interactions (two OH•••O interactions between the hydroxyl hydrogen of tyrosine (1) and the carboxyl oxygen of phenylalanine (4) and glycine (3)) of these folded conformers were collected.

The finding that the most stable conformers of the tetrapeptide have folded structures led us not to perform a test run for the pentapeptide Tyr-Gly-Gly-Phe-Leu, but just employ the hierarchical selection scheme collecting only conformers that contain the characteristic hydrogen-bonding interactions that the most stable folded tetrapeptide conformers form as well as conformers of Tyr-Gly-Gly-Phe-Leu that form an OH•••O interaction between the hydroxyl group of tyrosine (1) and the carboxyl COOH of leucine (5). Considering the variation of the dihedral angles of the flexible bonds of the tetrapeptide, for the pentapeptide dihedrals D,

F, I, L should be varied in steps of 60°, dihedrals A, B, E, H in steps of 180°, and dihedrals C, G, J, P and Q in steps of 120°. For the dihedrals that define the leucine residue the step sizes were determined with the following criteria: dihedrals S, U, R and Q where determined according to the variation of corresponding dihedral angles of the flexible bonds of the C-terminus of the tetrapeptide. Dihedral T, which defines the position of two methyl groups of leucine (5), was set to be varied in steps of 180°. However, this would result in a too large number of possible conformations to be created. Thus dihedrals A, B, E, H, I, K, M, R, T, U were varied in steps of 180°, dihedrals D, C, G, P, Q, S were varied in steps of 120° and dihedrals F and L were varied in steps of 60°. This variation of the dihedral angles resulted in 53 million possible structures. To further reduce the number of conformers to be created the hierarchical selection method was employed considering only conformers that contain three types of specific hydrogen-bonding interactions. The first type includes an interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of glycine (3) **(O_{14}-H_{22}-O_{34})** as found in the most stable Tyr-Gly-Gly conformers (chapter 4). The second type includes an interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of phenylalanine (4) **(O_{14}-H_{22}-O_{41})**. The third type of interaction considered is formed between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of leucine (5) **(O_{14}-H_{22}-O_{60})**.



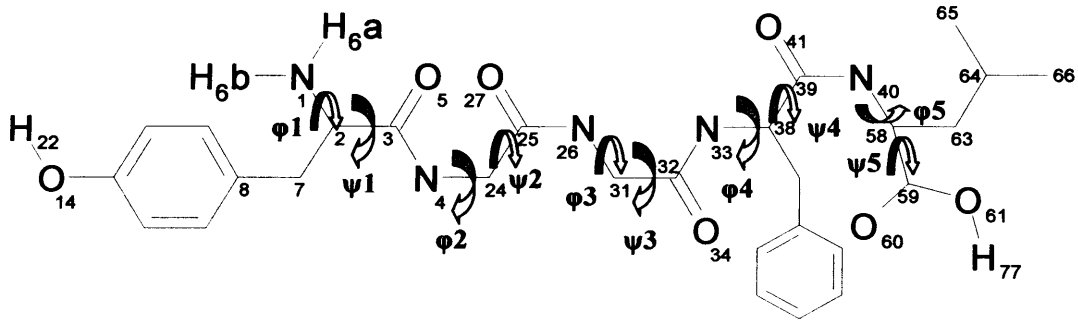


Figure 23 A. Atom labeling and definition of the dihedral angles considered in the hierarchical selection procedure.

B. Definition of the dihedral angles ω, φ, ψ.

The total number of conformers created was 9480. Single-point energy calculations were performed using the HF/3-21G* method for all the conformers. For the 500 most stable conformers resulting from the single-point energy calculations, geometry optimizations were performed using the HF/3-21G* method. The 20 most stable conformers resulting from the HF optimizations were optimized at the B3LYP/6-31+G* level of theory. From these calculations 19 conformers converged to distinct minimum structures. The relative energies of these conformers were evaluated by single-point calculations at the MP2/6-31+G* level of theory.

6.3 Results

From the studies performed in chapters 4 and 5 for the tripeptide Tyr-Gly-Gly and the tetrapeptide Tyr-Gly-Gly-Phe it was found that folded conformers are favoured over extended conformers. Thus only conformers with selected H-bonding interactions were taking into consideration using the H-bond selection method for Tyr-Gly-Gly-Phe-Leu. Table 10 lists the 19 most stable conformers (based on MP2 single-point energy calculations) resulting from the H-bond selection method. Figure 24 shows the structures of the 16 most stable conformers.

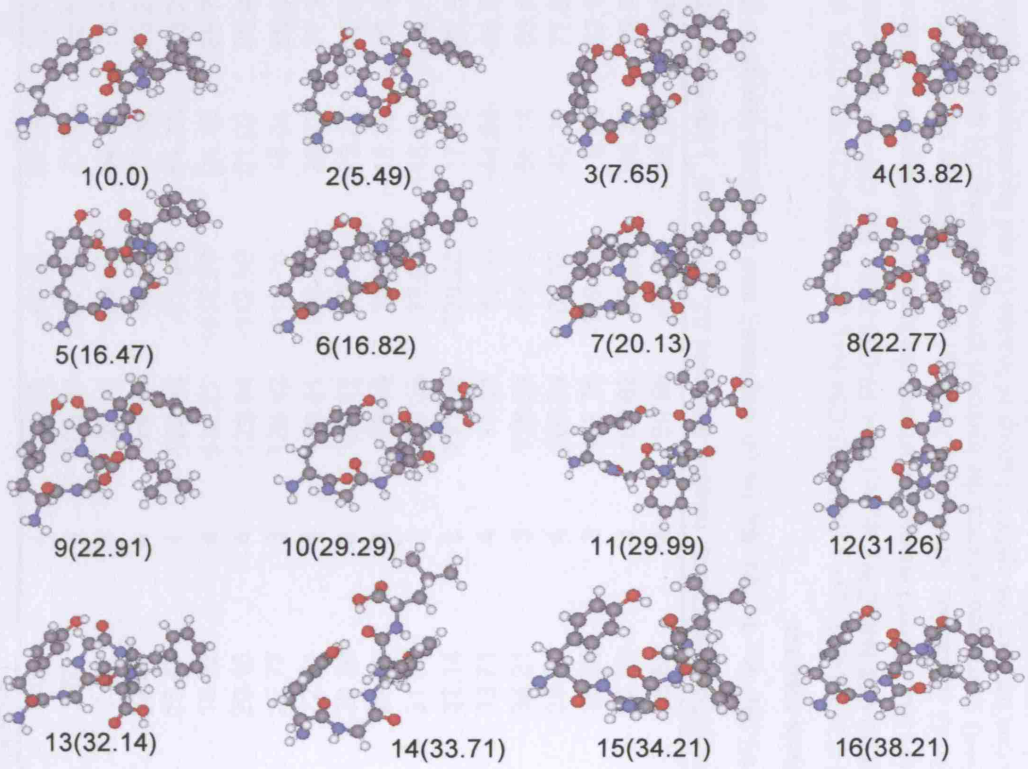


Figure 24. The 16 most stable Leu-enkephalin conformers identified in the current work. Their relative energies (from single point MP2/6-31+G* calculations optimized at the B3LYP/6-31+G* level of theory) are given in kJ mol^{-1}.

Conformers	Structure	Rel. Energy (kJ/mol)	Hbonds	ψ1	φ2	ψ2	φ3	ψ3	φ4	ψ4	φ5	ψ5
1	RP-anti	0.00	4	131.93	-80.03	62.05	73.16	-85.13	-103.77	-23.65	-119.09	156.59
2	RG-OHO-anti	5.50	5	103.48	-87.81	43.60	78.53	-69.87	-119.00	-19.38	-79.16	75.93
3	RG-OHO-anti	7.65	5	99.62	-89.37	44.19	78.43	-69.90	-118.98	-17.99	-79.16	75.73
4	RP-anti	13.83	4	125.34	-80.48	61.32	73.28	-89.01	-99.78	-25.34	-96.75	112.85
5	RP-anti	16.47	4	122.66	-81.29	61.25	73.22	-89.07	-99.76	-25.36	-96.97	112.82
6	RG-anti	18.82	4	132.27	-110.96	21.15	85.72	-65.58	-111.49	92.43	53.11	-135.34
7	RG-anti	20.13	4	129.14	-112.59	21.02	85.91	-65.61	-111.47	93.10	54.00	-137.35
8	RG-anti	22.77	3	126.12	-113.78	15.04	86.91	-80.16	-92.87	-8.46	-125.56	96.93
9	RG-anti	22.91	4	108.41	-90.63	39.23	77.38	-73.85	-118.10	0.54	-123.71	103.07
10	RP-NHN-anti	29.30	3	-30.12	62.18	-123.22	-115.06	33.99	-65.76	124.88	-135.17	153.22
11	RP-NHN-anti	29.99	3	-28.98	78.67	-118.30	-115.49	24.41	-83.75	71.94	-112.01	164.28
12	RP-NHN-anti	31.67	3	-28.36	77.97	-118.84	-114.77	23.23	-83.95	68.30	-78.03	161.91
13	RG-anti	32.14	3	131.95	-120.28	17.87	82.57	-63.20	-92.61	174.71	-63.15	-30.36
14	RP-OHO(a)-syn	33.71	4	11.21	82.06	-44.64	-88.85	-15.19	71.73	-64.75	-77.63	68.42
15	RP-OHO(b)-anti	34.21	3	118.99	-84.70	56.14	88.86	-122.04	-78.66	-40.76	-127.70	-54.71
16	RG-anti	38.21	4	100.34	-86.07	40.29	77.62	-70.18	-123.62	2.16	-122.32	102.83
17	RG-anti	39.29	3	122.76	-109.78	6.99	87.10	-74.06	-99.53	115.34	82.19	78.09
18	RG-anti	42.99	4	105.62	-93.27	36.78	79.70	-59.00	-150.48	40.36	-160.24	135.59
19	RG-anti	53.51	4	99.89	-86.03	38.86	79.27	-58.00	-151.47	40.03	-159.83	133.53

Table 10. Number of hydrogen-bonding interactions; relative energies ΔE (in kJ mol^{-1}) based on single point MP2/6-31+G* calculations; values of the Ramachandran angles ψ_1, φ_2, ψ_2, φ_3, ψ_3, φ_4, ψ_4, φ_5, ψ_5 (in degrees); and structural elements of the most stable conformers, obtained using the hierarchical conformational analysis method.

ψ_1=τ (N1-C2-C3-N4), φ_2= τ (C25-C24-N4-C3), ψ_2= τ (N26-C25-C24-N4), φ_3= τ (C32-C31-N26-C25), ψ_3= τ (N33-C32-C31-N26), φ_4= τ (C39-C38-N33-C32), ψ_4= τ (O41-C39-C38-N33), φ_5=τ(C59-C58-N40-C39), ψ_5=τ(O60-C59-C58-N40), RP: OH•••O interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of phenylalanine (4). RG: OH•••O interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of glycine (3). RP-OHO^1: RP structure with an additional O-H•••O interaction between the hydroxyl group of leucine (5) and the carboxyl group of phenylalanine (4). RP-OHO^2: RP structure with an additional O-H•••O interaction between the hydroxyl group of leucine (5) and the carboxyl oxygen of glycine (2). RG-OHO: RG structure with an additional O-H•••O interaction between the hydroxyl group of leucine (5) and the carboxyl oxygen of phenylalanine (4). NH•••N: Interaction between the amide hydrogen of glycine (2) and the nitrogen of the N-terminus of the peptide.

As was found for the tetrapeptide Tyr-Gly-Gly-Phe, most of the Leu-enkephalin conformers obtained are characterised by specific interactions between the hydroxyl group of tyrosine and either the carboxyl oxygen of glycine (3) or phenylalanine (4). Only conformer 14 contains the third interaction as the carboxyl oxygen of phenylalanine (4) forms a bifurcated bond with the hydroxyl oxygen of leucine (5) and the hydroxyl oxygen of tyrosine (1). These interactions observed for the conformers were the conformers with the specific interactions that were selected for the H-bond selection method. As was also observed for the smaller peptides, some of the conformers have only minor structural differences. For example the only difference between conformers 2↔3, 4↔5, 6↔7, 8↔9 and 18↔19 is observed at the position of the nitrogen atom of the N-terminus of the peptide. Conformers 10 and 11 only differ by about 110° in dihedral $C_{64}C_{63}C_{58}N_{40}$, which defines the position of the two methyl groups of leucine (5) and by about 60° in dihedral $N_{40}C_{39}C_{38}N_{33}$ (the ψ_4 Ramachandran angle). The most stable conformer found is characterised by a strong hydrogen-bonding interaction between the hydroxyl group of tyrosine (1) and the carboxyl oxygen of phenylalanine (4), whereas the second most stable conformer is characterised by a strong hydrogen-bonding interaction between the hydroxyl group of tyrosine (1) and the carboxyl group of glycine (3) (difference in energy 5.5 kJ/mol).

For the most stable conformer found three additional NHO interactions can be observed. The first interaction is between the amide hydrogen of glycine (2) and the carboxyl oxygen of leucine (5). The second interaction is between the amide hydrogen of glycine (3) and the carboxyl oxygen of tyrosine (1). Finally the third interaction is between the carboxyl oxygen of glycine (3) and both the hydrogen amide of phenylalanine (4) and the hydrogen amide of leucine (5), forming a bifurcated bond. The three NHO interactions observed for the most stable conformer are also present in the second most stable conformer. The second most stable conformer contains an additional OHO interaction between the hydroxyl oxygen of leucine (5) and the carboxyl oxygen of phenylalanine (4). Conformers 1 and 4 are very similar in structure. They only differ in the dihedral angles that define the position of the C-terminus of the peptide (dihedrals C_{59}-C_{58}-N_{40}-C_{39} (φ_5) and O_{60}-C_{59}-C_{58}-N_{40} (ψ_5) both differ by about 30°, whereas C_{64}-C_{63}-C_{58}-N_{40} differs by about 100°).

The majority of the most stable conformers found for the tripeptide Tyr-Gly-Gly and for the tetrapeptide Tyr-Gly-Gly-Phe have ψ_1 values close to 0° as this allows the formation of an $N_4H_{28}\bullet\bullet\bullet N_1$ interaction. Additionally the majority of the most stable conformers found for Tyr-Gly-Gly and Tyr-Gly-Gly-Phe form an $NH_2\bullet\bullet\bullet\pi$ interaction between the N-terminus and the phenyl ring of tyrosine. However, from the 19 most stable Leu-Enkephalin conformers only conformer 14 contains a strong $N_4H_{28}\bullet\bullet\bullet N_1$ interaction and an $NH_2\bullet\bullet\bullet\pi$ interaction between the N-terminus and the phenyl ring of tyrosine. The majority of the most stable conformers have ψ_1-angle values around 100°, which enables an interaction between the amide hydrogen of glycine (2) and the carboxyl oxygen of leucine (5) (conformers 1-9, 13,15-19). Finally conformers 10-12 contain a strong $N_4H_{28}\bullet\bullet\bullet N_1$ interaction but do not form an $NH_2\bullet\bullet\bullet\pi$ interaction between the N-terminus and the phenyl ring of tyrosine.

The most stable conformers found for the Tyr-Gly-Gly-Phe-Leu pentapeptide can be grouped in four different groups according to the number of the hydrogen-bonding interactions observed in the conformers. The first group contains conformers that are characterised by a hydrogen-bonding interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of phenylalanine (4) (conformers 1, 4, 5) (labelled RP in table 10). The second group contains conformers that have a strong interaction between the hydroxyl hydrogen of tyrosine and the carboxyl oxygen of glycine (3) (conformers 6, 7, 8, 9, 13, 16, 17, 18, 19) (labelled RG in table 10). The third group contains conformers that have the RP interaction also observed in the first group of conformers as well as a strong NH•••N interaction between the amide hydrogen of glycine (2) and the nitrogen atom of the N-terminus (labelled NHN in table 10) (conformers 10, 11, 12). The fourth group contains conformers that are characterised by two OHO interactions (conformers 2, 3, 14, 15). Conformers 2 and 3 contain the RG interaction as well as a strong OHO interaction between the hydroxyl oxygen of leucine (5) and the carboxyl oxygen of phenylalanine (4). Conformer 14 is characterised by a bifurcated OHO interaction between the carboxyl oxygen of phenylalanine (4) (RP in table 10) and both the hydroxyl hydrogen of tyrosine and the hydroxyl hydrogen of leucine (5). Finally conformer 15 contains the RP

interaction as well as an additional OHO interaction between the hydroxyl hydrogen of leucine (5) and the carboxyl oxygen of glycine (3).

Figure 25 shows the Ramachandran distribution of glycine (2) (3a), glycine (3) (3b), phenylalanine (4) (3c) and leucine (5) (3d) for the 19 conformers that were identified from the B3LYP/6-31+G* geometry optimizations. As we can observe from figure 25a all of the 19 most stable conformers are γ-type (φ_2 and ψ_2 angles are around -60°/+60° and +60°/-60°). This type of conformers enables the formation of a folded structure. The same type (γ-type) of conformers is observed to dominate the φ_3 and ψ_3 angle combinations (figure 25b). From this plot we can also observe that only conformer 14 has α-type motif. This conformer is the only one that contains an $NH_2 \cdots \pi$ interaction between the N-terminus and the phenyl ring of tyrosine. From the plot of φ_4 against ψ_4 (figure 25c) we can observe that the five most stable conformers are of α-type. The plot of φ_5 against ψ_5 (figure 25d) shows that the five most stable conformers have a γ-type motif which enables the formation of the bifurcated NHO interaction (N_{33}-$H_{42} \cdots O_{27}$, N_{40}-$H_{62} \cdots O_{27}$).

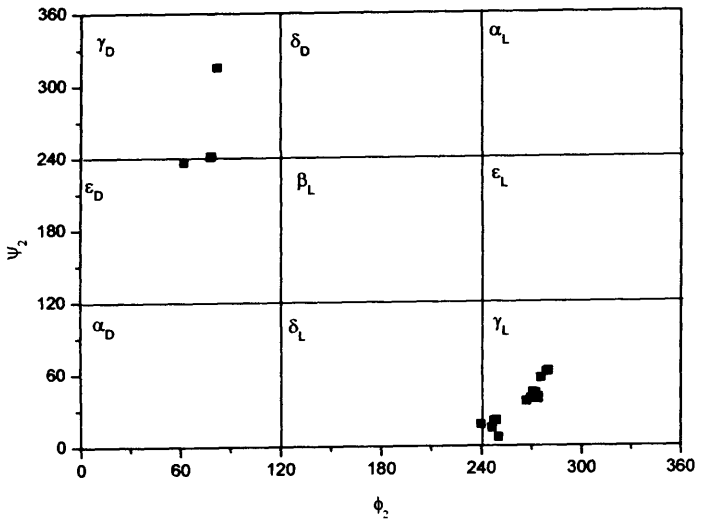


Figure 25a. Ramachandran plot showing the φ_2/ψ_2 combinations occurring in the 19 Tyr-Gly-Gly-Phe-Leu conformers identified and optimized with B3LYP/6-31+G* in the current work.

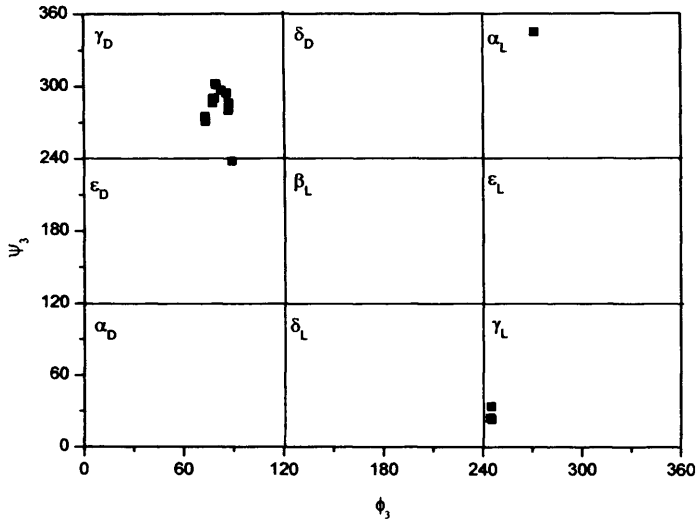


Figure 25b. Ramachandran plot showing the φ_3/ψ_3 combinations occurring in the 19 Tyr-Gly-Gly-Phe-Leu conformers identified and optimized with B3LYP/6-31+G* in the current work.

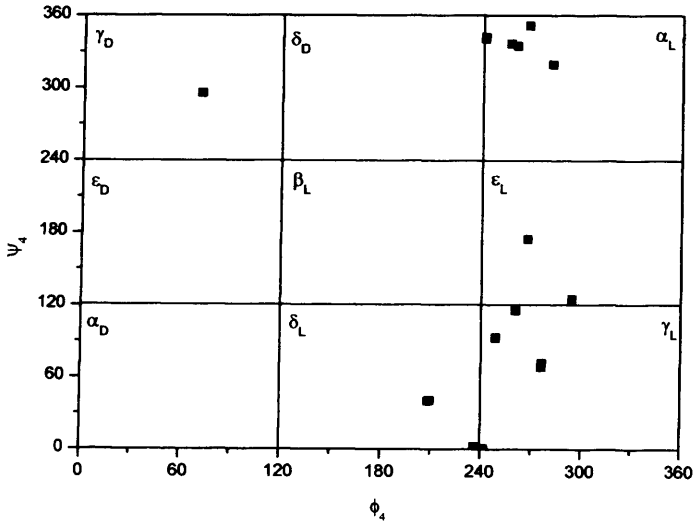


Figure 25c. Ramachandran plot showing the φ_4/ψ_4 combination occurring in the 19 Tyr-Gly-Gly-Phe-Leu conformers identified and optimized with B3LYP/6-31+G* in the current work.

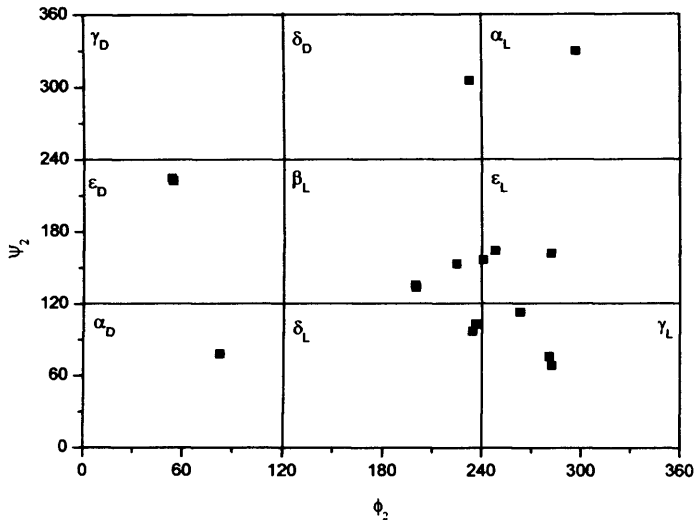


Figure 25d. Ramachandran plot showing the φ_5/ψ_5 combination occurring in the 19 Tyr-Gly-Gly-Phe-Leu conformers identified and optimized with B3LYP/6-31+G* in the current work.

Figure 26 shows the relative energies of the pentapeptide Tyr-Gly-Gly-Phe-Leu computed at the B3LYP and MP2 single-point energy level of theory. As we can observe from the graph significant variations between the results obtained with the two methods exist as also has been found in the previous chapters for Tyr-Gly, Tyr-Gly-Gly and Tyr-Gly-Gly-Phe. It was too computationally expensive to perform zero-point energy calculations for the most stable conformers. However from the previous chapters it was observed that zero-point energy corrections did not change significantly the order of stability of the most stable conformers. Thus we did not include zero-point energy calculations in this chapter.

Figure 27 shows the order of stability of the Tyr-Gly-Gly-Phe-Leu pentapeptide conformers according to single-point energies at the MP2 level of theory based on the DFT geometries. As can be observed from the graph the most stable conformer found from DFT remains the most stable conformer according to the MP2 single-point energies. Large variations are observed for conformer 14, 12, and 11 according to the B3LYP geometry optimizations. These conformers are observed to be the second, third and fourth most stable conformers respectively according to MP2 single-point energy calculations. Conformers 11, 12, and 14

contain the formation of a strong $N_4H_{28}\bullet\bullet\bullet N_1$ hydrogen bond. As was found in the previous chapters a large variation is observed between the DFT and MP2 order of stabilities. However the variation is not as large as was observed for the tetrapeptide Tyr-Gly-Gly-Phe. This is probably due to the fact that for the pentapeptide only the first 20 most stable conformers resulted from HF geometry optimizations were optimized at the B3LYP level of theory.

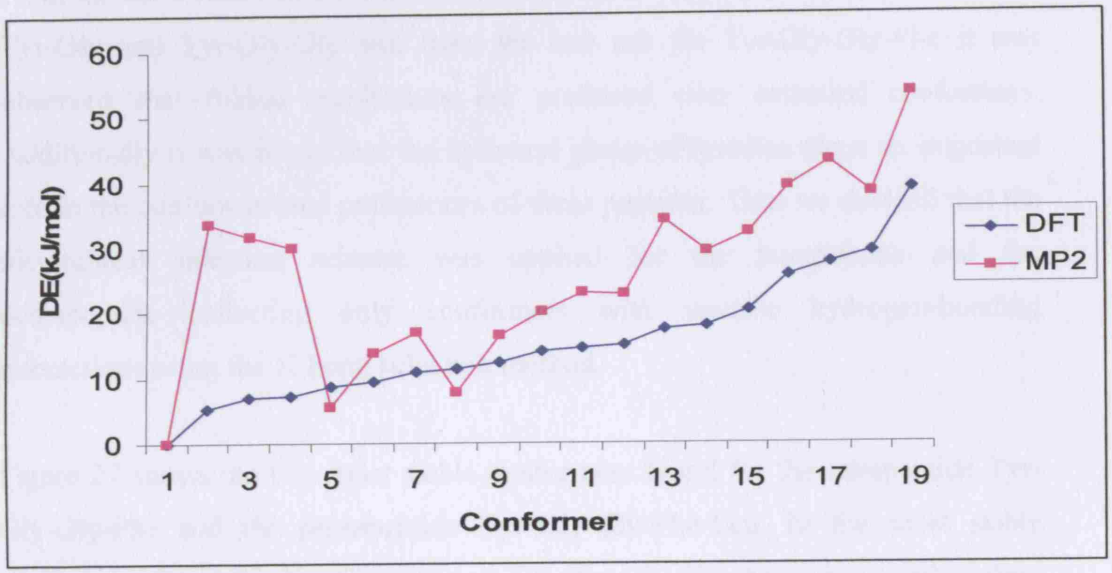


Figure 26. Comparison of the B3LYP and MP2 relative energies of the 19 most stable Tyr-Gly-Gly-Phe-Leu conformers (based on the DFT geometries).

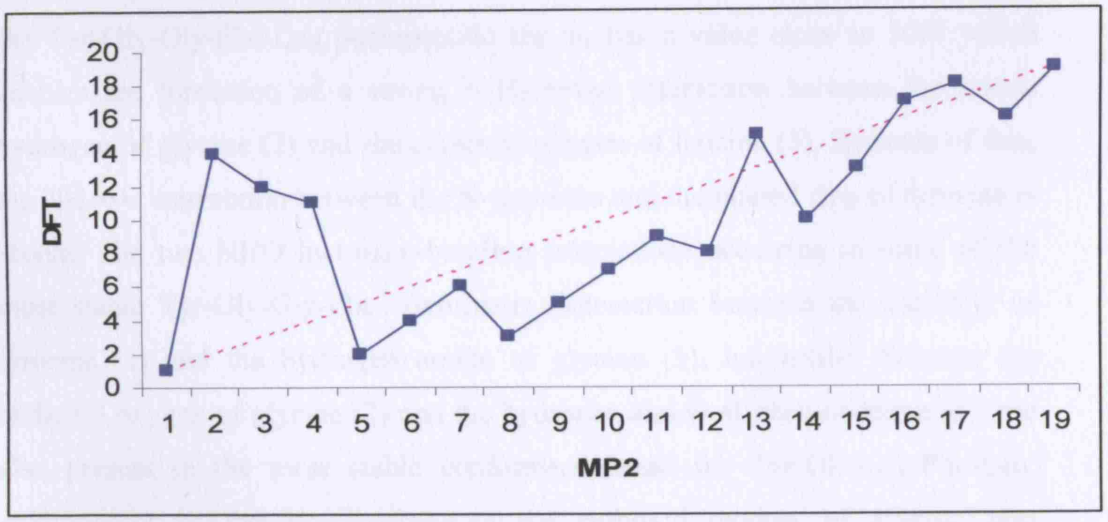


Figure 27. Comparison of the DFT and MP2 orders of stability of the 19 most stable conformers of Leu-enkephalin. The dotted line indicates ideal agreement.

6.4 Discussion

In this chapter we applied the hierarchical selection scheme to explore the conformation features of the pentapeptide Tyr-Gly-Gly-Phe-Leu. In the previous chapters we performed studies to explore the conformation features of the Tyr-Gly dipeptide, Tyr-Gly-Gly tripeptide and the Tyr-Gly-Gly-Phe tetrapeptide. From the most stable conformers found from the hierarchical selection scheme for Tyr-Gly and Tyr-Gly-Gly and from the test run for Tyr-Gly-Gly-Phe it was observed that folded conformers are preferred over extended conformers. Additionally it was found that the hydroxyl group of tyrosine plays an important role in the conformational preferences of these peptides. Thus we decided that the hierarchical selection scheme was applied for the tetrapeptide and the pentapeptide collecting only conformers with specific hydrogen-bonding interactions using the H-bond selection method.

Figure 27 shows the two most stable conformers found for the tetrapeptide Tyr-Gly-Gly-Phe and the pentapeptide Tyr-Gly-Gly-Phe-Leu. In the most stable conformers found for Tyr-Gly-Gly and Tyr-Gly-Gly-Phe the ψ_1 has a value close to 0°, which enables the formation of a strong $N_4H_{28}\bullet\bullet\bullet N_1$ interaction. These conformers also form an $NH_2\bullet\bullet\bullet\pi$ interaction between the N-terminus and the phenyl ring of tyrosine. In the majority of the most stable conformers found for the Tyr-Gly-Gly-Phe-Leu pentapeptide the ψ_1 has a value close to 100° which enables the formation of a strong $N_4H_{28}\bullet\bullet\bullet O_{60}$ interaction between the amide hydrogen of glycine (2) and the carboxyl oxygen of leucine (5). Because of this, the $NH_2\bullet\bullet\bullet\pi$ interaction between the N-terminus and the phenyl ring of tyrosine is absent. The two NHO hydrogen-bonding interactions occurring in some of the most stable Tyr-Gly-Gly-Phe conformers (interaction between the carboxyl of tyrosine (1) and the hydrogen amide of glycine (3), interaction between the carboxyl oxygen of glycine (2) and the hydrogen amide of phenylalanine (4)) are also present in the most stable conformers found for Tyr-Gly-Gly-Phe-Leu. Additionally in Tyr-Gly-Gly-Phe-Leu the carboxyl oxygen of glycine also interacts with the amide hydrogen of leucine (5), which results in a bifurcated hydrogen bond. Finally an important difference between the most stable

conformers of the tetrapeptide and pentapeptide is the position of the hydroxyl group of tyrosine, which differs by about 180°.

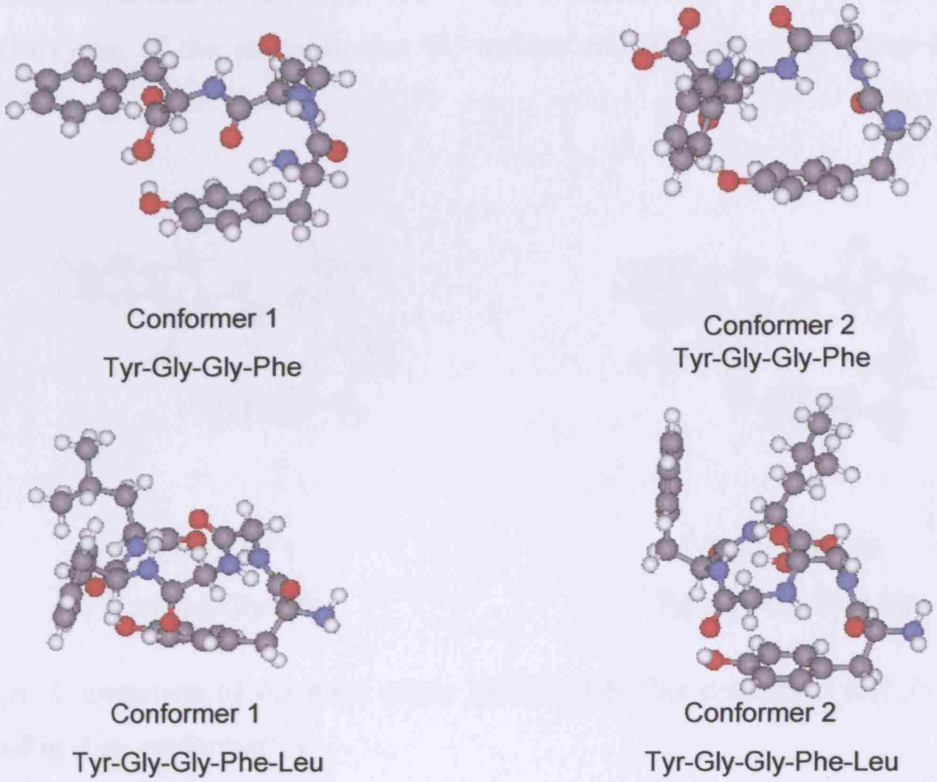
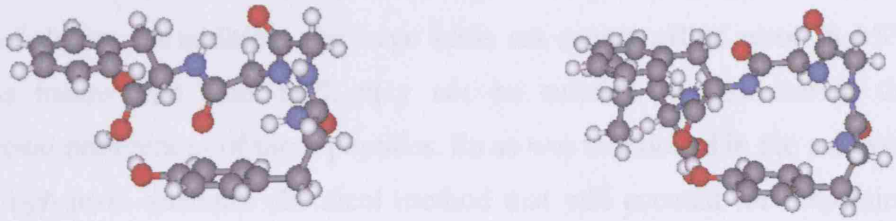


Figure 28. The two most stable conformers found for Tyr-Gly-Gly-Phe and Tyr-Gly-Gly-Phe-Leu.

From previous conformational studies on enkephalins discussed in chapter 2 it was found that the low-lying conformers obtained contain a characteristic hydrogen-bonding interaction between the C=O of phenylalanine and the amide hydrogen of glycine (2)[28, 29]. From our study it was found that only conformers 13 and 14 according to the MP2 single-point energy calculations exhibit this interaction. However, conformer 14 is the second most stable conformer according to the order of stability that the B3LYP geometry optimizations gives (the only conformer that contains a strong $N_4H_{28}\bullet\bullet\bullet N_1$ interaction and an $NH_2\bullet\bullet\bullet\pi$ interaction between the N-terminus and the phenyl ring of tyrosine). Additionally the dihedral angles of the flexible bonds of the most stable Tyr-Gly-Gly-Phe conformers are similar to Tyr-Gly-Gly-Phe part of the Tyr-Gly-Gly-Phe-Leu conformer 14. Figure 29 shows the most stable conformer found for Tyr-Gly-Gly-

Phe and conformer 14 of Tyr-Gly-Gly-Phe-Leu. Conformer 14 is the only conformer found from our study where the hydroxyl oxygen of leucine (5) is hydrogen bonded to the carboxyl oxygen of phenylalanine (4). None of the previous studies discussed in chapter two[30-33] has found an interaction between the carboxyl oxygen of the phenylalanine (4) residue and the carboxylic group of leucine (5).



Conformer 1
Tyr-GLy-Gly-Phe

Conformer 14
Tyr-GLy-Gly-Phe-Leu

Figure 29. Comparison of the most stable Tyr-Gly-Gly-Phe conformer and Tyr-Gly-Gly-Phe-Leu conformer 14.

From the current study only the 20 most stable conformers obtained from Hartree-Fock geometry optimizations were selected and optimized at the B3LYP level of theory. It is clear that the risk of missing low-lying conformers has increased. In the study performed in chapter 5 it was observed that, from the 48 most stable conformers of Tyr-Gly-Gly-Phe, the second most stable conformer according to B3LYP geometry optimizations has become conformer 48 according to MP2 single-point energy calculations. Thus, it is desirable to select at least the first 100 most stable conformers resulting from the Hartree-Fock geometry optimizations in order to reduce the risk to miss low-lying conformers. However it was computationally too expensive to perform geometry optimizations for this range of conformers for the pentapeptide Tyr-Gly-Gly-Phe-Leu.

The large variation between DFT and MP2 orders of stability that has been reported in the previous chapter for the Tyr-Gly-Gly tripeptide and the tetrapeptide Tyr-Gly-Gly-Phe has been also observed for Tyr-Gly-Gly-Phe-Leu. From the previous chapters we found that DFT is not the appropriate method to characterize the energetics of these peptides as it fails to describe the dispersion effects arising from interactions involving the aromatic residues. Although MP2 probably would describe more accurately the conformational preferences of these peptides, it is too computationally expensive to perform geometry optimizations at this level of theory. In addition, the large basis set superposition error in MP2 calculations means that also MP2 may not be suitable to characterize the conformational preferences of these peptides. So as was mentioned in the previous chapter, a high-level quantum chemical method that will account for dispersion effects arising from interactions with the aromatic residues, and that would not produce large the basis set superposition errors, is needed in order to perform a complete scan of the conformation pool of these peptides. However from our calculations performed for Tyr-Gly, Tyr-Gly-Gly, Tyr-Gly-Gly-Phe and Tyr-Gly-Gly-Phe-Leu some interesting conclusions were derived concerning the structural preferences of these peptides and these will be discussed in the next concluding chapter.

References

1. Saundwers, M. J., *J. Am. Chem. Soc.* **1987,** 108, 3150-3152.
2. Jacchieri, S. G., *Computers and Chemistry* **2001,** 25, 145-159.
3. Paine, G. H.; Scheraga, H. A., *Biopolymers* **1985,** 24, 1391-1436.
4. Vasquez, M.; Scheraga, H. A., *Biopolymers* **1985,** 24, 1437-1447.
5. Ripoll, D. R.; Scheraga, H. A., *J. Protein Chem.* **1989,** 8, 263-287.
6. Okamoto, Y.; Kikuchi, T.; Kawai, H., *Chem. Lett.* **1992,** 3, 1275-1278.
7. Nayeem, A.; Vila, J.; Scheraga, H. A., *J. Comput. Chem.* **1991,** 12, 594.
8. Hansmann, U. H. E.; Okamoto, Y., *J. Comput. Chem.* **1993,** 14, 1333.
9. Abagyan, R. A.; Argos, P. J., *J. Mol. Biol.* **1992,** 225, 519.
10. Shaumann, T. S.; Braun, W.; Wuthrich, K., *Biopolymers* **1990,** 29, 679.
11. Vajda, S.; Delisi, C., *Biopolymers* **1990,** 29, 1755.
12. Premilat, S.; Maigret, B. J., *J. Phys. Chem.* **1980,** 84, 293-299.
13. Hansmann, U. H. E., *Chem. Phys. Lett.* **1997,** 281, 140.
14. Meirovitch, H.; Meirovitch, E.; Michel, A. G.; Vasquez, M., *J. Phys. Chem.* **1994,** 98, 6241.
15. Meirovitch, H.; Scheraga, H. A., *Macromolecules* **1981,** 14, 1250.
16. Klein, C. T.; Mayer, B.; Kohler, P.; Wolschann, P., *J. Comput. Chem.* **1988,** 19, 1470.
17. van Gunsteren, W. F.; Berendsen, H. J. C., Groningen Molecular Simulation (GROMOS) Library Manual, Biomos,Groningen. **1987.**
18. Solmajer, T.; Mehler, E. L., *Int. J. Quant. Chem.* **1992,** 44, 291.
19. Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, A.; Zagari, A.; Rumsey, S.; Scheraga, H. A., *J. Chem. Phys.* **1992,** 96, 6472.
20. Rayan, A.; H., S.; Goldblum, A., *J. Mol. Graphics Modell.* **2004,** 22, 319-333.
21. Weiner, S. J.; Kollmann, P. A., *J. Comput. Chem.* **1986,** 7, 230.
22. Santagata, L. N.; Suvire, F. D.; Enriz, R. D.; Torday, L. L.; Csizmadia, I. G., *J. Mol. Struct.(Theochem)* **1999,** 465, 33-67.
23. Kriz, Z.; Koca, J.; Carlsen, P. H. J., *J. Mol. Struct. (Theochem)* **2001,** 540, 231-250.
24. Perez, J.; Villar, H.; Loew, G. H., *J. Comput. Aided Mol. Des.* **1992,** 6, (2), 175-190.

25. Hirst, J. D.; Watson, T. M., *Phys.Chem.Chem.Phys.* **2004,** 6, 2580-2587.

26. Snoek, L. C., Private Communication.

27. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, J., T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, Q. C., J. V.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A., Gaussian, Inc., Pittsburgh PA, 2003.

28. Jalkanen, K. J., *J. Phys. Condens. Matter* **2003,** 15, 1823-1851.

29. Abdali, S.; Jensen, M.; Bohr, H., *J. Phys. Condens. Matter* **2003,** 15, 1853-1860.

30. De Coen, J.; Humblet, C.; Koch, M., *FEBS Lett.* **1977,** 73, 38-42.

31. Abdali, S.; Jalkanen, K. J.; Cao, X.; Nafie, L. A.; Bohr, H., *Phys. Chem. Chem. Phys.* **2004,** 5, 1295-1300.

32. Abdali, S.; Niehus, T. A.; Jalkanen, K. J.; Cao, X.; Nafie, L. A.; Frauenheim, T.; Suhai, S.; Bohr, H., *Phys. Chem. Chem. Phys.* **2003,** 6, 2580-2587.

33. Chandrasekhar, I.; van Gunsteren, W. F.; Zandomeneghi, G.; Williamson, P.; Meier, T. F., *J. Am. Chem. Soc.* **2006,** 128, 159-170.

7

THE STRUCTURE OF THE GLY-GLY-GLY TRIPEPTIDE

7.1. Introduction

The hierarchical selection method developed in chapter 3 has been applied to study the conformational features of the aromatic peptides Tyr-Gly (chapter 3), Tyr-Gly-Gly (chapter 4), Tyr-Gly-Gly-Phe (chapter 5) and Tyr-Gly-Gly-Phe-Leu (Leu-enkephalin) (chapter 6). It was observed that B3LYP predicts a different order of stability of the most stable conformers than that based on MP2 single-point calculations. This is likely due to the fact that B3LYP (and most others commonly-used DFT functionals) fails to describe the long-range dispersion effects[1-4]. In the current study we employ the hierarchical selection method to study the simplest tripeptide, Gly-Gly-Gly. The purpose of this study is to employ the hierarchical selection method to a non-aromatic peptide, for which dispersion is expected to be not so important. From the order of stability of the most stable conformers for Gly-Gly-Gly resulting from B3LYP and MP2 calculations we can identify the differences of the two methods for the conformational preferences of a peptide that does not contain an aromatic residue.

Due to the high flexibility of peptides the majority of the studies performed are focused on simple peptides consisting of alanine and glycine residues and their derivatives using experimental and theoretical approaches[5-9]. Electronic structure theory methods have been used in order to explore the potential energy surfaces of alanine and glycine dipeptides in the gas phase and in solvent environment[10, 11]. In these studies Hartree-Fock theory was applied to investigate the potential energy surface of these peptides. The results obtained showed that Hartree-Fock level calculations provide a useful database against which the performance of various empirical and semi-empirical methods can be tested. In addition the calculations on alanine and glycine dipeptides in the gas phase and in a solvent environment showed that the solvent environment is critical to the stabilization of helical minima for small peptides.

Wang and Duan[10] studied the solvation effects of alanine dipeptide studying the (φ, ψ) energy maps and conformers in the gas phase, ether and aqueous solution using the polarizable continuum model (PCM) [12-16] at the MP2 level of theory[17]. It was found that the PCM calculations give good descriptions of the solvent effects and non-bonded interactions. Other studies have used density functional theory to determine the conformations and energies of these peptides[18-20].

A number of papers have been published exploring the conformational features of tripeptides. Due to the computational cost the majority of the studies used classical molecular dynamics methods and semi-empirical potentials[6-9] (such as CHARMM[21] and AMBER[22]). Torii et al. published a study to explore the conformational energy landscape of glycine dipeptide and tripeptide[6]. In this study the coupling constants between the amide I vibrations of neighbouring peptide groups were calculated (at the HF/6-31G** level of theory). It was suggested that the conformation with (φ, ψ) = (-100°, 170°) is most likely for the extended helix.

Mehdizadeh published a study on the conformational effect of one glycine on the other glycine residues using the tripeptide Ac-Gly-Gly-Gly-NHMe[23]. It was found that the central glycine residue affects the conformation of the two adjacent terminal residues. Additionally, each of the terminal residues can affect significantly the central residue. However, the effect of one terminal residue on the other terminal residue is minimal. Salpietro et al. performed Hartree-Fock calculations to study the entropy of various backbone conformers of HCO-Gly-Gly-Gly-NH_2[5]. From this study ΔE, ΔH and ΔG functions showed that a global minimum is observed at 180° whereas the entropy curves showed local maxima at 180°. The conformational features of alanine and glycine derivatives were studied by Perczel et al. using molecular mechanics and ab initio SCF methods[24]. Both methods predicted a γ-type conformation to be the global minimum on the Ramachandran map[25]. Although molecular mechanics calculations were able to predict the global minimum on the potential energy surface, it was found that only ab initio SCF was capable of describing properly the repulsive and attractive interactions.

In chapter 3 it was shown that Tyr-Gly conformers with less than two hydrogen-bonding interactions in their structure are not among the most stable conformers. This indicates that Gly-Gly-Gly conformers with less than two hydrogen-bonding interactions can be neglected. To explore the most stable conformers for Gly-Gly-Gly we employ two strategies similar to those employed for the tripeptide Tyr-Gly-Gly in chapter 4. The first strategy is based on the hierarchical selection method. The second strategy is based on the H-bond selection method.

7.2. Methodology

The hierarchical selection method developed in chapter 3 was applied to explore the conformational features of the tripeptide Gly-Gly-Gly. Figure 30a shows the flexible bonds of the tripeptide that were considered for the procedure. Figure 30b shows the atomic labelling of the tripeptide with the location of the Φ and Ψ angles.

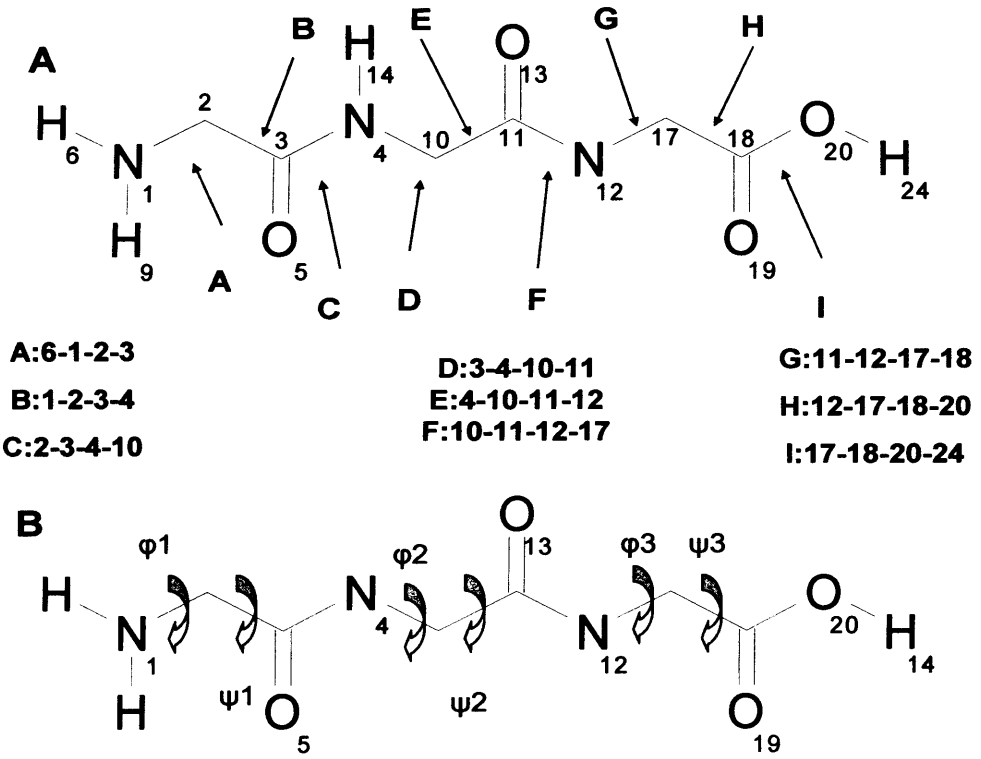


Figure 30. **A.** Atomic labelling and definition of the dihedral angles considered in the hierarchical selection method.

B. Definition of the angles ψ_1, φ_2, ψ_2, φ_3, ψ_3.

7.2.1 Test run

As in chapter 4 (for Tyr-Gly-Gly) the hierarchical selection method was first employed using large step sizes as a test run, to determine the best step sizes for rotating the flexible bonds of the tripeptide. The total number of conformers created from the program was 29395. Single point energy calculations were performed at the HF/3-21G* level of theory for all the conformers. The first 1000

low-lying conformers obtained from the SPE calculations were optimized using the HF/3-21G* method. From these, geometry optimizations were performed for the 55 most stable conformers using the B3LYP/6-31+G* method. 53 conformers optimized to different minimum structures. For these, the final values of the dihedral angles of the bonds rotated were collected.

7.2.2. Hierarchical selection method

From the variation of the dihedrals of the 53 most stable conformers resulting from the test run, the optimal step sizes for rotation of the internal bonds were determined. Thus, dihedrals C, F and I were varied with a step size of 180°, dihedrals B, E, G and H were varied in steps of 60°, and dihedrals A and D were varied with a step size of 30°.

The total number of conformers generated was 266,116. Due to the large number of conformers generated it was too computationally expensive to perform single point calculations for all conformers. Instead, HF/3-21G* SPE calculations were performed for the conformers that contain one to six hydrogen bonds in their structure and the conformers that did not have any hydrogen bonds were neglected. This is justified, as the results obtained from the test run showed that conformers that do not contain any hydrogen-bonding interactions are not among the most stable conformers. In addition it was found that conformers with less than two hydrogen-bonding interactions were not among the most stable for the Tyr-Gly dipeptide and Tyr-Gly-Gly tripeptide. The first 2000 low-lying conformers obtained from the SPE calculations were optimized using the HF/3-21G* method. From these optimizations only 86 conformers optimized to different minimum structures. These 86 conformers resulting from the HF/3-21+G* optimizations were optimized using the B3LYP functional and the 6-31+G* basis set. From these, 63 conformers optimized to different minimum structures.

7.2.3. Comparison using semi-empirical methods

For the 30 most stable conformers obtained from the single point energy calculations, geometry optimizations were performed using HF/3-21G* and the semi-empirical methods AM1[26], PM3[27], and PM3MM[28] to check if the results of these methods, which are much less computationally expensive, correlate with those obtained at the HF level of theory. PM3MM specifies the PM3 model including the optional molecular mechanics correction for the peptide bond (HCON) linkages[28]. In addition geometry optimizations using the AM1 semi-empirical method were performed for the first 2000 conformers obtained from the HF/3-21G* single point calculations.

7.2.4. H-bond selection method

The 50 most stable conformers contain five different types of hydrogen-bonding interactions. As for Tyr-Gly-Gly, to check if there are additional low-lying conformers with these interactions, not found from the hierarchical selection method for conformers that contain one to six hydrogen-bonding interactions, we collected the conformers that form all these specific interactions within their structure using the H-bond interaction method introduced in chapter 4. For these conformers SPE calculations at the HF/3-21G* level were performed and the most stable conformers were optimized at the HF/3-21G* and B3LYP/6-31+G* levels of theory.

The five types of the specific hydrogen-bonding interactions, type **I, II, III, IV** and **V** contain two hydrogen-bonding interactions. For types **I, II** and **IV** the first interaction occurs between the amide hydrogen of glycine (2) and the nitrogen of the N-terminus ($\mathbf{N_4\text{-}H_{14}\cdots N_1}$). The second interaction occurs between the amide hydrogen of glycine (3) and the carboxyl oxygen of glycine (1) ($\mathbf{N_{12}\text{-}H_{21}\cdots O_5}$) in type **I**, between the amide hydrogen of glycine (3) and the amide nitrogen of glycine (2) ($\mathbf{N_{12}\text{-}H_{21}\cdots N_4}$) in type **II** and between the hydroxyl hydrogen of glycine (3) and the carboxylic oxygen of glycine (1) ($\mathbf{O_{20}\text{-}H_{24}\cdots O_5}$) in type **IV**. For types **III** and **V** the first interaction occurs between the amide hydrogen of

glycine (3) and the carboxyl oxygen of glycine (1) **(N_{12}-H_{21}•••O_5)**. The second interaction in type **III** occurs between the amide hydrogen of glycine (3) and the carboxyl oxygen of glycine (1) **(N_1-H_{24}•••O_{20})**, whereas in type **V** it occurs between the hydroxyl hydrogen of glycine (3) and the carboxylic oxygen of glycine (2) **(O_{20}-H_{24}•••O_{13})**.

We collected 912 conformers of type **I**, 1536 conformers of type **II**, 117 conformers of type **III**, 168 conformers of type **IV** and 101 conformers of type **V**. Geometry optimizations were performed for all these conformers at the HF/3-21G* level of theory. From these optimizations 18 conformers of type **I**, 40 conformers of type **II**, 25 conformers of type **III**, 10 conformers of type **IV** and 6 conformers of type **V** optimized to different minimum structures. All these optimized conformers were combined yielding 81 different minimum structures. These 81 conformers were optimized at the B3LYP/6-31+G* level of theory. From these calculations 57 conformers optimized to different minimum structures. The 57 conformers obtained were combined with the structures obtained from the conformers that contain 1, 2, 3, 4, 5 and 6 hydrogen bonds resulting in 97 unique conformers. The relative energies of the 20 most stable conformers according to the B3LYP/6-31+G* geometry optimizations were evaluated by single-point calculations at the MP2 level of theory. Zero-point energies (ZPEs), scaled by 0.976[29] were calculated using B3LYP/6-31+G*.

7.3. Results

As for Tyr-Gly-Gly, the H-bond selection method located additional conformers that were not found by the main run. So complementing the main run with the H-bond selection method is essential to locate the most stable conformers of the Gly-Gly-Gly tripeptide. Table 11 lists the 20 most stable conformers (based on MP2 single-point calculations corrected with B3LYP zero-point energies) resulting from the selection method and the main run. Figure 31 shows the 20 most stable conformers. Bold entries in Table 11 indicate conformers that were not found by the main run.

Conformer 1 can be described as a folded conformer exhibiting a weak interaction between the COOH carboxyl oxygen of glycine (3) and the hydrogen of the NH_2 terminus. Most of the conformers adopt an extended structure and contain at least two hydrogen-bonding interactions in their structure. This confirms that conformers with a small number of hydrogen-bonding interactions can be discarded. In the most stable conformers three main interactions occur. The first interaction is between the nitrogen of the N-terminus of the peptide and the nitrogen of glycine (2) (N_4-H_{14}•••N_1). The second interaction is between the hydrogen attached to the nitrogen of glycine (3) and the carboxyl oxygen of glycine (1) (N_{12}-H_{21}•••O_5). The third interaction is between the hydroxyl hydrogen of glycine (3) and the carboxyl oxygen of glycine (2) (O_{20}-H_{24}•••O_{13}). According to the types of specific H-bonds considered in the H-bond selection method for Gly-Gly-Gly, conformers 1-12, 16, 17 and 20 correspond to the type I motif, whereas conformers 11, 13, 14, 18, 19 correspond to type II and conformer 15 corresponds to type IV. Finally conformers 3, 5, 9 and 10, which exhibit three H-bonds in their formation, all contain the type V combination. None of the twenty most stable Gly-Gly-Gly conformers contain structures with the type III combination. Many of the conformers are very similar in their structure. For example conformers 1-2, 3-5, 4-6, 9-10, 11-12, 16-17, 18-19 only differ in the orientation of the N-terminus. Additionally conformers 7-8, 13-14, and 16-20 are only different in the position of the C-terminus of the peptide. All the conformers have a ψ_1 angle close to 0° (-16° to 22°) which allows the formation of the strong N_4-H_{14}•••N_1 interaction.

Conformers	*Structure*	*TYPE*	*HBonds*	*Rel. Energy(kJ/mol)*	ψ1	φ2	ψ2	φ3	ψ3	ω1	ω2
1	**anti**	**I**	**2**	**0**	**21.73**	**79.69**	**-69.51**	**-83.45**	**178.83**	**-174.47**	**164.80**
2	anti	I	2	4.281	-12.09	79.68	-67.91	-89.89	179.35	-177.14	170.25
3	syn-OHO	I/V	3	10.132	-12.51	-79.54	65.05	73.04	-58.60	-174.90	176.16
4	**anti**	**I**	**2**	**10.201**	**-12.50**	**80.41**	**-59.80**	**90.94**	**-179.36**	**177.49**	**-174.34**
5	syn-OHO	I/V	3	10.377	10.61	-79.24	64.87	73.24	-58.82	-177.92	175.87
6	anti	I	2	10.619	12.73	80.79	-59.95	90.04	-179.04	173.94	-174.04
7	**syn**	**I**	**2**	**11.751**	**11.93**	**-80.25**	**65.07**	**95.57**	**-6.29**	**-178.94**	**-176.34**
8	**anti**	**I**	**2**	**11.987**	**13.62**	**80.72**	**-64.99**	**-96.06**	**6.53**	**175.53**	**176.87**
9	**anti-OHO**	**I/V**	**3**	**12.167**	**-11.67**	**79.00**	**-60.77**	**74.80**	**-57.21**	**175.87**	**173.24**
10	anti-OHO	I/V	3	12.731	11.37	79.36	-60.88	74.73	-57.22	172.50	173.33
11	syn	I	2	13.830	12.33	-80.41	60.14	-96.74	6.28	-177.49	176.64
12	**syn**	**I**	**2**	**14.117**	**-12.54**	**-80.81**	**60.36**	**-95.89**	**5.84**	**-174.01**	**176.41**
13	syn	II	2	15.621	-11.64	-110.40	11.51	174.86	-179.52	-173.63	178.01
14	anti	II	2	16.418	-15.93	108.77	-10.81	-178.69	179.49	176.30	-177.22
15	syn-OHO*	IV	3	18.699	-15.00	-57.17	121.79	98.85	-14.78	177.22	-171.60
16	anti	I	2	21.510	-10.96	81.01	-52.02	-112.86	10.71	174.34	-171.36
17	anti	I	2	21.684	10.53	81.51	-52.37	-112.68	10.48	171.18	-171.68
18	Syn-OHO	II	3	22.070	15.80	-107.32	8.97	75.35	-58.56	-172.06	171.57
19	Syn-OHO	II	3	22.283	-7.27	-107.75	10.77	75.53	-58.80	-168.34	171.32
20	anti	I	2	23.256	-10.81	79.85	-70.03	116.10	-10.02	176.74	169.24

Table 11. Number of hydrogen-bonding interactions; relative energies ΔE (in kJ mol^{-1}) based on single-point MP2/6-31+G* calculations with inclusion of scaled (0.976) B3LYP/6-31+G* zero-point energies; values of the Ramachandran angles of ψ_1, φ_2, ψ_2, φ_3, ψ_3 (in degrees); and structural elements of the most stable conformers, obtained using the hierarchical conformational analysis method.

ψ2= τ (N4-C3-C2-N1), φ2= τ (C3-N4-C10-C11), ψ2= τ (N4-C10-C11-N12), φ3= τ (C11-N12-C17-C18), ψ3= τ (N12-C17-C18-O20), ω1= τ (C2-C3-N4-C10), ω2= τ (C10-C11-N12-C17).

OHO: OHO•••O hydrogen bonding interaction between the C-terminal hydroxyl group and the carbonyl oxygen of glycine (2).

Syn/anti: N-terminus NH_2 *syn* or *anti* with respect to glycine OH.

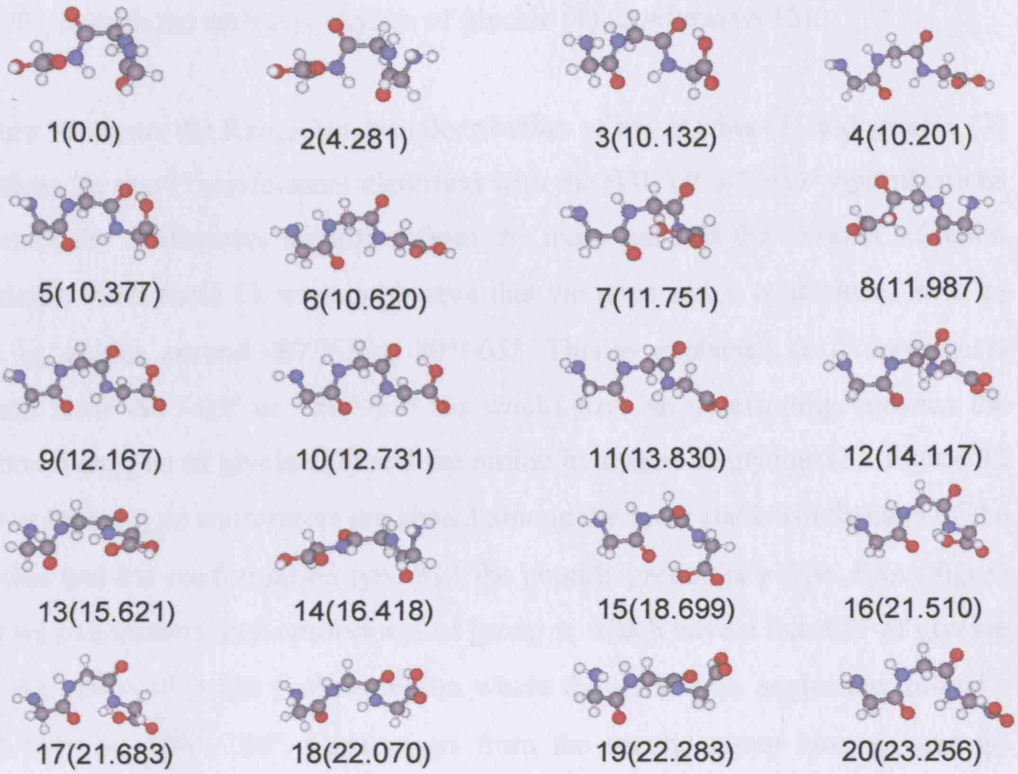


Figure 31. Structures of the 20 most stable Gly-Gly-Gly conformers. Relative energies (from single-point MP2/6-31+G* calculations with inclusion of scaled (0.976) B3LYP/6-31+G* ZPEs) are given in kJ mol^{-1}.

According to the structural elements of the peptide we can divide the conformers into 3 categories. The first group contains conformers that have the ψ_3 angle around 180° and have a non-H-bonded OH hydroxyl group of glycine (3) (conformers 1, 2, 4, 6, 13, 14). The second group contains conformers that have the ψ_3 angle around 60°. These conformers contain a characteristic OHO (O_{20}-H_{24}•••O_{13}) hydrogen-bonding interaction between the hydroxyl hydrogen of glycine (3) and the carboxyl oxygen of glycine (2) (conformers 3, 5, 9, 10, 18, 19). The third group contains conformations that have the ψ_3 angle around 0° (conformers 7, 8, 11, 12, 15, 16, 17, 20). This group can be subdivided in two groups: A) Conformers that have a free OH hydroxyl group of glycine (3) (conformers 7, 8 11, 12). B) Conformers that form a weak interaction between the

OH hydroxyl group of glycine (3) and the nitrogen of glycine (3) (conformers 16, 17, 20), or with the carboxyl oxygen of glycine (1) (conformers 15).

Figure 32 shows the Ramachandran distribution of the glycine (2) and glycine (3) residues for the 97 conformers identified with the B3LYP/6-31+G* optimizations (sum of the conformers identified from the main run and the H-bond selection method). From table 11 we can observe that the most stable conformers have φ_2 and ψ_2 angles around -80°/65°or 80°/-65°. This is explained as if conformers would have -80°/-65° or +80°+65° we would have an overlapping between the carboxyl oxygen of glycine (3) and the amide hydrogen of glycine (3). Figure 32 shows that α-type conformers are absent among the most stable conformers of the peptide and the conformation type that the peptide prefers is γ-type. From figure 32b we can observe that conformers of group 1, which have a free OH of glycine (3), are observed in the β-sheet region where the φ_3 and ψ_3 angles are around -90°/-180° or -180°/-180°. Conformers from the second group have φ_3 and ψ_3 angles around 75°/-60° as this is required to form the OHO interaction between the hydroxyl hydrogen of glycine (3) and the carboxyl oxygen of glycine (2). Finally conformers from the third group have the φ_3 and ψ_3 angles around -90/6°or 90/-6°. The results obtained are in agreement with our results for the Tyr-Gly-Gly tripeptide. For Tyr-Gly-Gly conformers with an extended structure the ψ_3 angle value is around -100°/+100° and -180°/+180° which correspond to the Gly-Gly-Gly conformers of group 1.

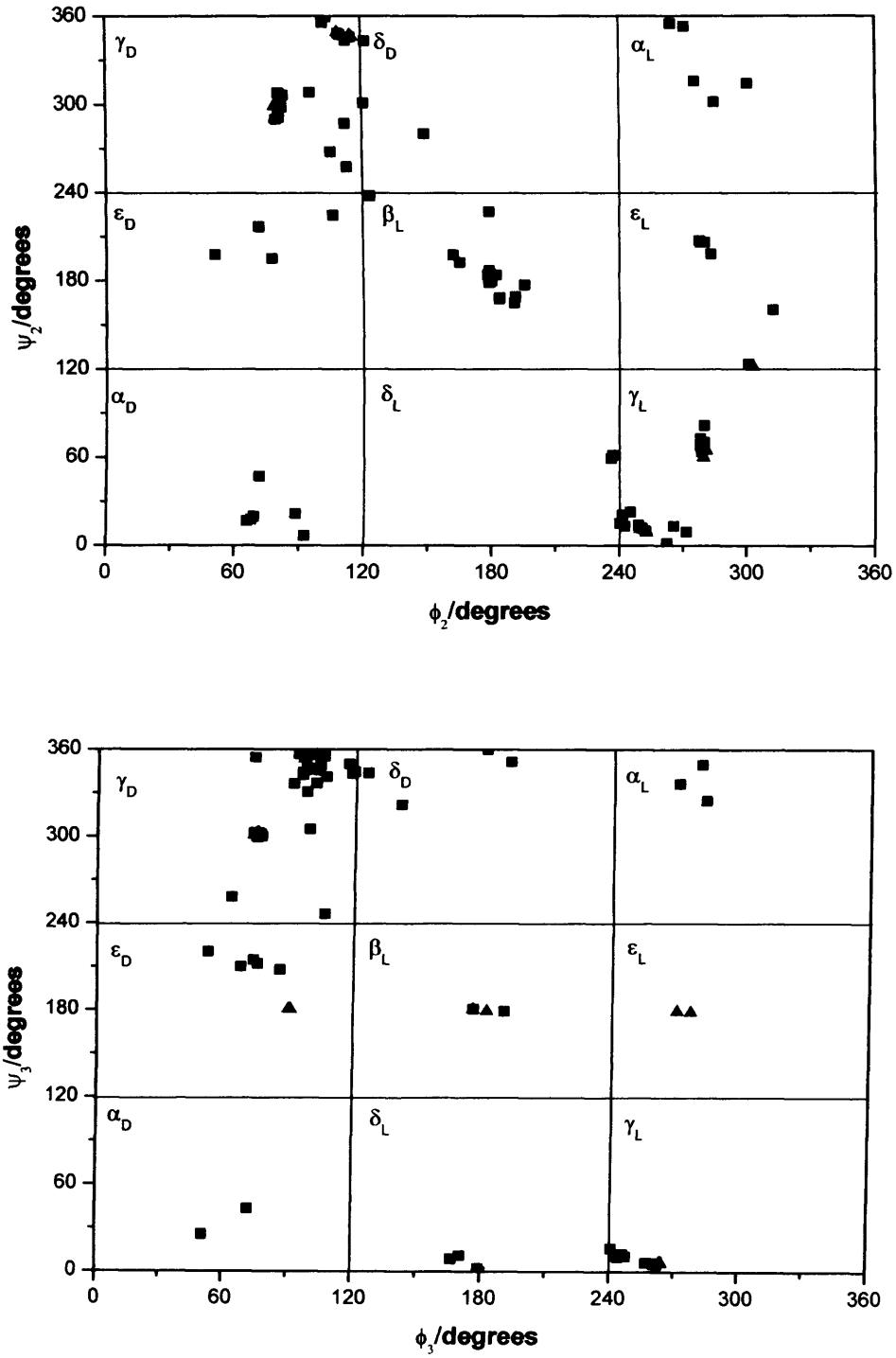


Figure 32. Ramachandran plot showing the φ_2/ψ_2 and φ_3/ψ_3 combinations occurring for the 97 Gly-Gly-Gly conformers identified and optimized with B3LYP/6-31+G* in the current work. In red are presented the angles of the 20 most stable conformers whereas in black are the rest of the conformer.

Figure 33 shows the relative energies of the Gly–Gly–Gly tripeptide computed at the B3LYP and the MP2 single-point energy level of theory, with and without the ZPE correction (scaled at the B3LYP level of theory). The ZPE values are larger for the conformers that form the characteristic OHO interaction (conformers 3, 5, 9, 10, 15, 18, 19). From figure 33 it can be observed that with inclusion of ZPE the relative energy is getting higher for the conformers that form the characteristic OHO interaction (conformers 3, 5, 9, 10, 15, 18, 19). So MP2 with the inclusion of ZPE favours the conformers that have a free hydroxyl group of the C-terminus. Also a significant difference is observed for conformers 13 and 14 where the peptide backbone is perpendicular to the N-terminus of the peptide. Figure 34 presents the order of stability of the conformers using the B3LYP and MP2 level of theory. The plot shows that there is no correlation for conformers 13 and 14 and for conformers that contain the OHO interaction. Additionally there is no correlation for conformers 16, 17 and 20. The weak interaction formed between the OH hydroxyl group of glycine (3) and the nitrogen of glycine (3) may explain this effect. This observation is in agreement with the study performed for Tyr-Gly where the MP2 single-point calculation inclusive the ZPE showed that conformers that had an OHO interaction and did not form a "book" structure were less stable than conformers that formed an extended structure and do not exhibit an OHO interaction.

Table 12 lists the harmonic vibrational frequencies of the NH stretches of glycine (2) and glycine (3) as well as their in-plane bending vibrations. The OH stretch and the in-plane bending vibration of glycine (3) are also listed. Conformers that contain an OHO interaction exhibit a red shift of the OH frequency ($\approx$ 320-350 cm^{-1} compared to the OH mode of the non-OHO conformers), and the corresponding intensities are predicted to be very large. The glycine (3) OH frequency is of similar magnitude as that of the aromatic tripeptides Tyr-Gly-Gly, Trp-Gly-Gly[30] and Phe-Gly-Gly[31] that contain an OHO interaction. The NH stretch frequencies of Gly-Gly-Gly are of the same magnitude as those calculated for the Tyr-Gly-Gly conformers that form an $N_4H_{14}\bullet\bullet\bullet N_1$ interaction. However the stretches for the NH_2 group have very small intensities and it remains to be seen if they can be observed experimentally.

	Structure	OH		NH(AS)		NH-Gly1		NH(SS)		NH-Gly2		CO(Gly2)B[a]		CO(Gly1)B		NH(Gly2)B		NH(Gly1)B	
1	anti	3594	(59)	3427	(8)	3388	(99)	3342	(3)	3329	(183)	1722	(402)	1680	(185)	1524	(213)	1497	(177)
2	anti	3596	(59)	3438	(8)	3394	(104)	3356	(2)	3319	(219)	1719	(420)	1678	(175)	1529	(225)	1499	(187)
3	syn-OHO	3273	(552)	3437	(10)	3387	(120)	3355	(5)	3299	(265)	1663	(177)	1623	(360)	1531	(209)	1497	(199)
4	anti	3596	(62)	3434	(8)	3396	(109)	3353	(2)	3336	(215)	1717	(426)	1676	(173)	1530	(244)	1501	(198)
5	syn-OHO	3279	(547)	3440	(11)	3386	(118)	3359	(5)	3300	(264)	1663	(164)	1624	(368)	1531	(209)	1497	(197)
6	anti	3596	(62)	3436	(8)	3396	(113)	3355	(3)	3335	(216)	1717	(425)	1676	(176)	1530	(243)	1501	(203)
7	syn	3598	(59)	3435	(8)	3396	(105)	3354	(2)	3325	(225)	1718	(418)	1677	(167)	1531	(239)	1501	(194)
8	anti	3597	(58)	3436	(8)	3396	(109)	3354	(2)	3324	(225)	1718	(418)	1676	(171)	1530	(239)	1501	(198)
9	anti-OHO	3269	(640)	3437	(10)	3389	(119)	3356	(5)	3302	(276)	1664	(215)	1624	(333)	1534	(192)	1499	(196)
10	anti-OHO	3268	(640)	3438	(10)	3388	(122)	3358	(5)	3300	(279)	1664	(224)	1624	(330)	1535	(191)	1499	(199)
11	syn	3598	(66)	3435	(9)	3394	(111)	3353	(3)	3331	(224)	1717	(422)	1676	(173)	1531	(246)	1501	(198)
12	syn	3598	(66)	3437	(9)	3394	(115)	3355	(3)	3331	(225)	1717	(422)	1676	(176)	1531	(245)	1502	(203)
13	syn	3595	(77)	3431	(8)	3384	(96)	3349	(2)	3443	(67)	1706	(416)	1665	(285)	1495	(234)	1476	(297)
14	anti	3596	(76)	3434	(8)	3384	(100)	3353	(3)	3438	(67)	1719	(285)	1678	(412)	1496	(229)	1477	(305)
15	syn-OHO*	3426	(694)	3435	(9)	3404	(106)	3353	(10)	3453	(27)	1727	(250)	1686	(371)	1502	(226)	1498	(205)
16	anti	3541	(148)	3439	(11)	3384	(121)	3358	(6)	3286	(263)	1731	(378)	1689	(196)	1516	(258)	1499	(203)
17	anti	3539	(151)	3439	(10)	3382	(122)	3358	(7)	3285	(263)	1731	(378)	1689	(197)	1516	(257)	1497	(206)
18	OHO-Syn	3263	(624)	3434	(10)	3375	(109)	3354	(6)	3454	(64)	1724	(280)	1683	(291)	1505	(180)	1481	(296)
19	OHO-Syn	3263	(617)	3440	(10)	3373	(106)	3359	(10)	3456	(63)	1724	(279)	1682	(285)	1504	(181)	1479	(293)
20	anti	3541	(144)	3439	(11)	3388	(124)	3358	(5)	3271	(299)	1733	(375)	1691	(188)	1516	(249)	1499	(201)

Table 12. Scaled (OH: 0.976; NH: 0.956) harmonic vibrational frequencies (in cm^{-1}) of the OH and NH stretch and in-plane bending modes. Calculated intensities (km mol^{-1}) are given in round brackets.

[a] **B**=bending, AS:Assymetric stretching modes of the two hydrogen atoms of the N-terminus.

SS: Symmetric stretching modes of the two hydrogen atoms of the N-terminus.

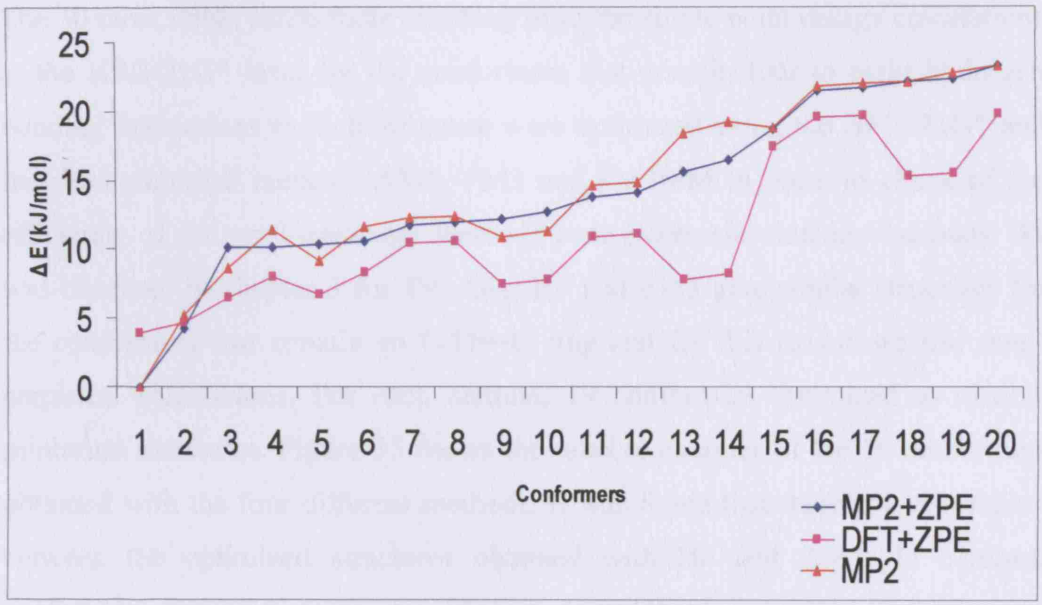


Figure 33. Comparison of the B3LYP+ZPE, MP2 and MP2+ZPE relative energies of the twenty most stable Gly-Gly-Gly conformers (based on the DFT geometries). The ZPEs were computed with B3LYP/6-31+G* and scaled by 0.976.

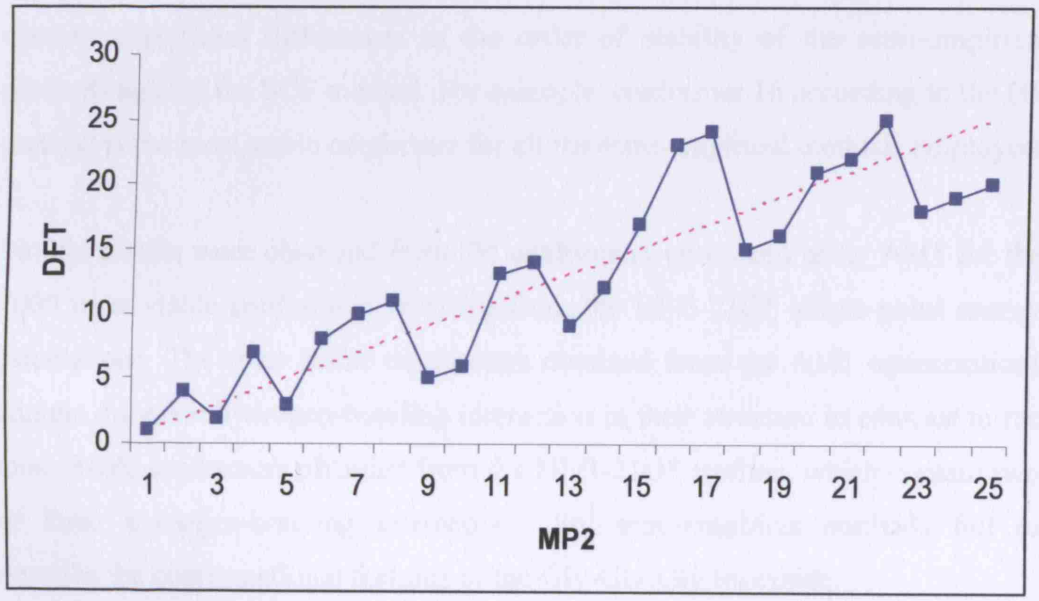


Figure 34. Relative order of stability of the 20 most stable conformers found for Gly-Gly-Gly at the B3LYP and MP2 level of theory.

The 30 most stable conformers resulting from the single point energy calculations at the HF/3-21G* level for the conformers that contain four to eight hydrogen bonding interactions in their structure were optimized using the HF/3-21G* and the semi-empirical methods AM1, PM3 and PM3MM in order to check of the efficiency of the semi-empirical methods over electronic structure methods. As was observed in chapter 3 for Tyr–Gly, HF and PM3 give similar structures for the conformers that contain an O-H•••O ring and for this reason we use semi-empirical calculations. For each method, 19 conformers optimized to distinct minimum structures. Figure 35 shows the relative energies of the 19 conformers obtained with the four different methods. It was found that there is a correlation between the optimized structures obtained with HF and AM1. In contrast, conformers resulting from the PM3 and PM3MM geometry optimizations converged to different minimum structures than those resulting from the HF optimizations. Although PM3MM and PM3 fail to correlate with the HF method, PM3MM gives slightly better structures than PM3. Table 13 shows the order of stability of the conformers obtained with AM1, PM3 and PM3MM in comparison with the order of stability resulting from the SCF method. From table 13 we can observe significant differences in the order of stability of the semi-empirical methods against the SCF method. For example, conformer 16 according to the HF method is the most stable conformer for all the semi-empirical methods employed.

Similar results were observed from the conformers optimized using AM1 for the 2000 most stable conformers resulting from the HF/3-21G* single-point energy calculations. The most stable conformers obtained from the AM1 optimizations contain only one hydrogen-bonding interaction in their structure in contrast to the most stable conformers obtained from the HF/3-21G* method, which contain two or three hydrogen-bonding interactions. So, semi-empirical methods fail to describe the conformational features of the Gly-Gly-Gly tripeptide.

CONFORMER	METHOD			
	HF/3-21G*	AM1	PM3	PM3MM
1	1	16	16	16
2	2	8	8	15
3	3	7	7	6
4	4	15	15	1
5	5	11	11	14
6	6	6	6	5
7	7	19	19	4
8	8	9	9	3
9	9	10	10	2
10	10	12	12	17
11	11	1	1	7
12	12	14	14	11
13	13	18	18	8
14	14	4	4	19
15	15	5	5	12
16	16	3	3	9
17	17	13	13	10
18	18	17	17	13
19	19	2	2	18

Table 13. Order of stability of the conformers obtained with AM1, PM3 and PM3MM in comparison to the order of stability resulting from the HF/3-21G* method.

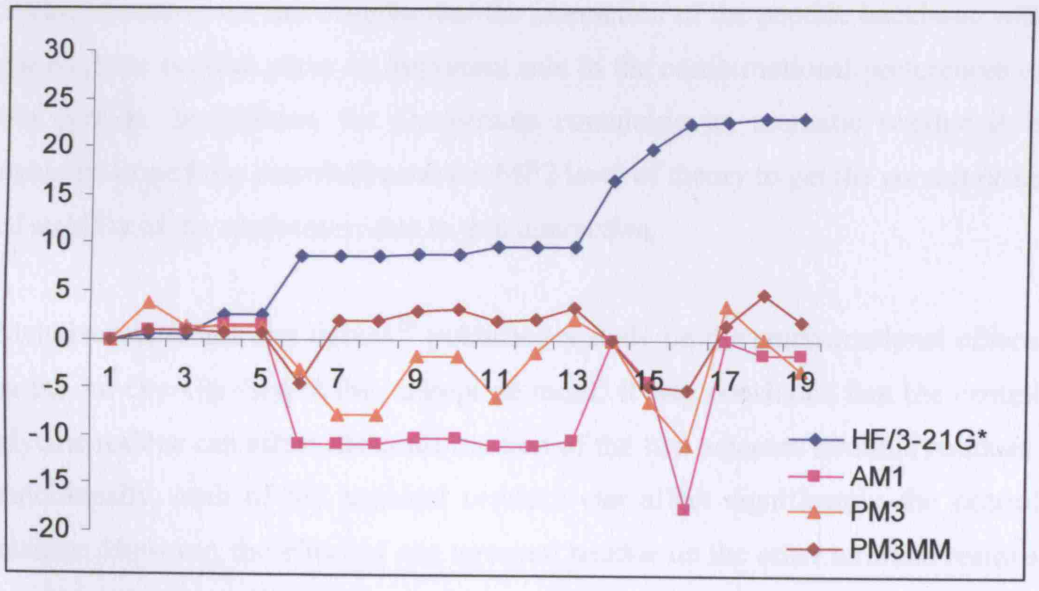


Figure 35. Relative stability (kJ mol^{-1}) of the 19 most stable conformers of Gly-Gly-Gly resulting from the 30 most stable conformers obtained from single point calculations employing the main run calculated with HF/3-21G*, AM1, PM3 and PM3MM.

7.4. Discussion

The results obtained for Gly-Gly-Gly can be compared to those of the peptides that contain an aromatic residue in their structure. In the Tyr-Gly dipeptide according to geometry optimizations at the MP2/6-31+G* level, the most stable conformer contains an OHO interaction between the hydroxyl OH of glycine (3) and the carboxyl of glycine (2). The second most stable conformer contains an OHO interaction between the carboxyl hydrogen of glycine (3) and the hydroxyl oxygen of tyrosine. In addition these structures form a "book" structure where the dihedral angle $C_8C_7C_3C_2$ is less than 90°. In the Tyr-Gly-Gly tripeptide the most stable conformers form a folded structure with two OHO interactions: The first interaction is between the carboxyl oxygen of glycine (3) and the hydroxyl hydrogen of tyrosine as in the most stable Tyr-Gly conformer. The second interaction is between the C-terminal hydroxyl group and the carbonyl oxygen as in the second most stable Tyr-Gly conformer according to the MP2 optimizations. In contrast, in glycine tripeptide the most stable conformers have extended (not folded) structures; some of these contain an OHO interaction. From the structures obtained for Tyr-Gly and Tyr-Gly-Gly where the most stable conformers form a folded structure, we can observe that the interaction of the peptide backbone with the electron π-cloud plays an important role in the conformational preferences of the peptide. In addition for conformers containing an aromatic residue it is essential to perform calculations at the MP2 level of theory to get the correct order of stability of the conformers due to this interaction.

Mehdizadeh, Chass and Farkas[23] published a study on the conformational effects in the Ac-Gly-Gly-Gly-NHMe tripeptide motif. It was concluded that the central glycine residue can affect the conformation of the two adjacent terminal residues. Additionally, each of the terminal residues can affect significantly the central residue. However, the effect of one terminal residue on the other terminal residue is minimal. The results of this study are in agreement with the current study, as for the most stable conformers of Gly-Gly-Gly large structural differences are observed when the φ_2 and ψ_2 (which define the internal bonds of the central glycine residue) angles are different.

Most of the low-lying conformers of Gly-Gly-Gly have an extended structure and some of them contain an OHO interaction. From the MP2 single-point calculations for the 20 most stable conformers, it was observed that the stability of the conformers that contain the OHO interaction as well as conformers that have the peptide backbone perpendicular to the N-terminus (conformers 13 and 14) was decreased with MP2. So MP2 prefers conformers that have a free OH hydroxyl group of glycine (3). This finding can be compared with the results obtained for Tyr-Gly-Gly where the MP2 single-point energy calculations showed that MP2 prefers conformers with folded structures due to the effect of dispersion forces between the aromatic ring and the peptide backbone. In addition for Tyr-Gly-Gly a significant variation in the relative energies resulting from the B3LYP and MP2 single-point calculations was observed. The most stable conformers resulting from the MP2 single-point calculations included only folded conformers whereas the most stable conformers resulting from the B3LYP optimizations included some conformers with extended structures. The structures resulting from single-point energy calculations for Gly-Gly-Gly do not show as large differences as those observed for Tyr-Gly-Gly. This is due to the absence of the aromatic ring in Gly-Gly-Gly, which means dispersion is less important in Gly-Gly-Gly.

MP2 single-point energy calculations were performed for the most stable conformers resulting from the B3LYP geometry optimizations. The comparison between the results obtained from the B3LYP and the MP2 single-point calculations show only small differences in the order of stability of the most stable conformers. This confirms that dispersion is less important. For this reason we did not perform geometry optimizations at the MP2 level of theory for the Gly-Gly-Gly tripeptide.

To explore the conformational landscape of Gly-Gly-Gly using the hierarchical selection scheme we performed single-point calculations for all possible conformations except the conformers with no hydrogen bonds. However the complete structural characterization of peptides is still a difficult task. The large flexibility of the peptides and the necessity to use high-level quantum chemical methods makes it impossible to perform calculations for all possible conformers if we consider larger peptides (tetra-peptide, penta-peptide) and for tripeptides

containing larger residues (as we saw in the previous chapters for Tyr-Gly-Gly). Experimental guidance is needed in order to locate the most stable conformers of the peptide.

References

1. van Mourik, T.; Gdanitz, R. J., *J. Chem. Phys.* **2002,** 116, 9620-9623.
2. van Mourik, T., *J. Chem. Phys.* **2004,** 304, 317.
3. Millet, A.; Korona, T.; Moszynski, R.; Kochanski, E., *J. Chem. Phys.* **1999,** 111, 7727.
4. Rappe, A. K.; Bernstein, E. R., *J. Phys. Chem. A,* **2000,** 104, 6117.
5. Salpieto, S.; Viskolcz, J.; Csizmadia, I. G., *J. Mol. Struct. (Theochem)* **2003,** 666-667, 89.
6. Torii, H.; Tasumi, M., *J. Raman Spectrosc.* **1998,** 29, 81.
7. Nguyen, P.; Stock, G., *J. Chem. Phys.* **2003,** 119, 11350.
8. Mu, Y.; Stock, G., *J. Phys. Chem. B.* **2002,** 106, 5294.
9. Mu, Y.; Kosov, D.; Stock, G., *J. Chem. Phys. B.* **2003,** 106, 5294.
10. Wang, Z.; Duan, Y., *J. Comput. Chem.* **2004,** 25, 1699.
11. Gould, I.; Cornell, W.; Hillier, I., *J. Am. Chem. Soc.,* **1994,** 116, 9250.
12. Cammi, R.; Mennucci, B.; Tomasi, J., *J. Phys. Chem. A.* **1999,** 103, 9100.
13. Cammi, R.; Mennucci, B.; Tomasi, J., *J. Phys. Chem. A.* **2000,** 104, 5631.
14. Cossi, M.; Rega, N.; Scalmani, G.; Barone, V., *J. Chem. Phys.* **2001,** 114, 5691.
15. Cossi, M.; Scalmani, G.; Rega, N.; Barone, V., *J. Chem. Phys.* **2002,** 117, 43.
16. Mennucci, B.; Tomasi, J., *J. Chem. Phys.* **1997,** 106, 5151.
17. Moller, C.; Plesset, M. S., *Phys. Rev.* **1934,** 46, 618.
18. Improta, R.; Barone, V., *J. Comp. Chem.* **2004,** 25, 1333.
19. Vargas, R.; Garza, J.; Hay, B.; D., D., *J. Phys. Chem. A.* **2002,** 106, 3213.
20. Percel, A.; Farkas, O.; Jakli, I.; I., T.; Csizmadia, I. G., *J.Comput.Chem.,* **2003,** 24, 1026.
21. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A., *J. Comput. Chem.* **1986,** 7, 230.
22. Brooks, B. R.; R.E., B.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., *J. Comput. Chem.* **1983,** 4, 187.
23. Mehdizadeh, A.; Chass, G. A.; Farkas, O.; Perczel, A.; Torday, L. L.; Varro, A.; Papp, G. J., *J. Mol. Struct.* **2002,** 588, 187.

24. Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J. L.; Marcoccia, J. F.; Csizmadia, I. G., *J. Am. Chem. Soc.*, **1991,** 113, 6256.

25. Perczel, A.; Kajtar, M.; Marcoccia, J. F.; Csizmadia, I. G., *J. Molec. Struct. (Theochem)* **1991,** 78, 291.

26. Dewar M.J.S, Z. E. C., Healy, E.F., Steward J.J.P, *J. Am. Chem. Soc.*, **1985,** 107, 3902.

27. Stewart J.J.P, *J. Comput. Chem.*, **1989,** 10, 209.

28. Stewart J.J.P, *J. Comput. Chem.*, **1989,** 10, 209.

29. Caminati, W.; Di Bernardo, S., *J. Mol. Struct.* **1990,** 240, 253.

30. Bakker J., M., Plützer C., Hünig I., Häber, T., Compagnon, I., Helden G., V., Meijer, G., Kleinermanns, K., *ChemPhysChem* **2005,** 6, 120-128.

31. Řeha, D., Valdés H.,Vondráek, J.,Hobza, P.,Abu-Riziq A., Bridgit Crews, B., De Vries, M., S., *Chem. Eur. J.* **2005,** 11, 1-16.

8

GENERAL CONCLUSIONS

The aim of this thesis was to develop a method that is able to explore the potential energy surface of small peptides using electronic structure methods. The method developed (chapter 2) explores the potential energy surface of peptides and hierarchically increases more accurate electronic structure theory methods including geometry optimization at the HF/3-21G*, B3LYP/6-31+G*, MP2/6-31+G* and single-point energy calculation at the MP2 level of theory.

The hierarchical selection method was employed to locate the most stable conformers of a number of small peptides. Figure 36 shows the most stable structure obtained for the Tyr-Gly dipeptide, Tyr-Gly-Gly tripeptide, Tyr-Gly-Gly-Phe tetrapeptide and Tyr-Gly-Gly-Phe-Leu pentapeptide. It was found that the tyrosine (1) residue plays an important role in the stability of the structures of these peptides[1, 2]. For the dipeptide folded ("book") structures are the most stable ones (structures resulting from a small (<90°) C_4-C_7-C_8-C_9 torsion angle) (chapter 3). The most stable Tyr-Gly-Gly conformers are stabilised by a characteristic "circle" conformation (structures resulting from a small (<90°) $C_8C_7C_2N_1$ torsion angle) (chapter 4). Both peptides contain characteristic OH•••O interactions between the hydroxyl group of the tyrosine (1) and the peptide backbone.

The most stable conformer of the Tyr-Gly dipeptide based on MP2 single-point energy calculations is stabilised by an OH•••O interaction between the hydroxyl hydrogen of glycine (2) and the carboxyl oxygen of tyrosine (1) residue. The most stable conformer of the Tyr-Gly-Gly tripeptide is mainly stabilised by two characteristic OH•••O interactions. The first interaction is between the hydroxyl hydrogen of tyrosine (1) and the carboxyl oxygen of glycine (3). The second interaction is between hydroxyl hydrogen of glycine (3) and the carboxyl oxygen of glycine (2). The most stable conformer of the Tyr-Gly-Gly-Phe tetrapeptide is stabilised mainly by three characteristic interactions. The first interaction is between the hydroxyl hydrogen of tyrosine (1) and the hydroxyl oxygen of phenylalanine (4). The second and third interaction is observed between the amino

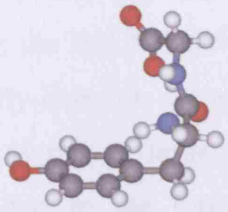
hydrogen of glycine (3) and the carboxyl oxygen of tyrosine (1) and between the amino hydrogen of phenylalanine (4) and the carboxyl oxygen of glycine (2). All of the most stable conformers found for Tyr-Gly, Tyr-Gly-Gly and Tyr-Gly-Gly-Phe have an additional NH•••N hydrogen-bonding interaction between the amide hydrogen of glycine (2) and the nitrogen atom of the N-terminus of the peptide. Finally the most stable conformer of the Tyr-Gly-Gly-Phe-Leu pentapeptide is stabilised by five characteristic interactions. The first interaction is between the hydroxyl hydrogen of tyrosine (1) and the carboxyl oxygen (C=O) of phenylalanine (4). The second and the third interaction are the two NH•••O interactions that occur in the most stable conformer of Tyr-Gly-Gly-Phe: namely those between the amino hydrogen of glycine (3) and the carboxyl oxygen of tyrosine (1) and between the amino hydrogen of phenylalanine (4) and the carboxyl oxygen of glycine (2). Additionally the carboxyl oxygen of glycine (2) is also involved in a hydrogen-bonding interaction with the amide hydrogen of leucine (5) forming a bifurcated bond. Finally another NH•••O interaction is stabilising the most stable Tyr-Gly-Gly-Phe-Leu: this interaction is between the nitrogen amide of glycine (2) and the carboxyl oxygen of leucine (5).

In table 14 we list the main interactions stabilizing the most stable conformer of the different peptides. For example, for the tetrapeptide (chapter 5, table 14) we listed the most stable conformer with five hydrogen-bonding interactions. The fifth interaction is a weak CH•••O and is not listed in the table below as an interaction that stabilizes the conformation of the peptide. As we can see from table 14 there is a correlation between the number of the amino acid residues of the peptide and the number of the hydrogen-bonding interactions in the most stable conformer. Additionally we can observe that from the dipeptide to the tripeptide, an additional OH•••O is stabilizing the global minimum, while from the tripeptide to the tetrapeptide and the pentapeptide additional NH•••O interactions are stabilizing the most stable conformers of the peptides.

PEPTIDE	HBONDS	INTERACTIONS
Tyr-Gly	2	OHO/NHN
Tyr-Gly-Gly	3	R-OHO/OHO/NHN
Tyr-Gly-Gly-Phe	4	RP-OHO/NHO/NHO/NHN
Tyr-Gly-Gly-Phe-Leu	5	RP-OHO/NHO/NHO/NHO/NHO

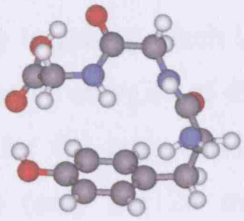
Table 14. Main interactions stabilising the most stable conformers of Tyr-Gly, Tyr-Gly-Gly, Tyr-Gly-Gly-Phe, Tyr-Gly-Gly-Phe-Leu.

It was found that there is a correlation between the number of hydrogen-bonding interactions in the most stable Tyr-Gly dipeptide, Tyr-Gly-Gly tripeptide, Tyr-Gly-Gly-Phe tetrapeptide and Tyr-Gly-Gly-Phe-Leu pentapeptide conformers. For example all of the 20 most stable Tyr-Gly conformers have two hydrogen-bonding interactions, whereas all of the 20 most stable Tyr-Gly-Gly conformers have three or four hydrogen-bonding interactions. Additionally, all of the 20 most stable Tyr-Gly-Gly-Phe conformers have four or five hydrogen-bonding interactions in their structure whereas all of the 20 most stable Tyr-Gly-Gly-Phe-Leu conformers have five or six hydrogen-bonding interactions in their structure. This indicates that dipeptide conformers that contain less than two hydrogen-bonding interactions can be discarded from the scan whereas tripeptide conformers with less than three hydrogen-bonding interactions, tetrapeptide conformers with less than four hydrogen-bonding interactions and pentapeptide conformers with less than five hydrogen-bonding interactions can be discarded.



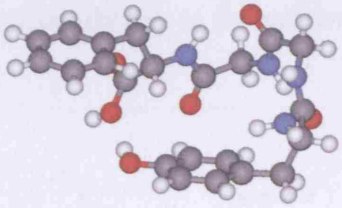
OH•••O, NH•••N

Tyr-Gly



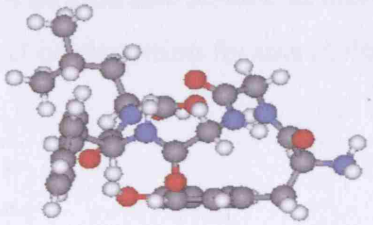
R-OH•••O, OH•••O, NH•••N

Tyr-Gly-Gly



R-OH•••O, 2(NH•••O), NH•••N

Tyr-Gly-Gly-Phe



R-OH•••O, 4(NH•••O)

Tyr-Gly-Gly-Phe-Leu

Figure 36. The global minimum found for Tyr-Gly, Tyr-Gly-Gly, Tyr-Gly-Gly-Phe and Tyr-Gly-Gly-Phe-Leu (from single-point MP2-6-31+G* calculations with inclusion of scaled (0.976) B3LYP/6-31+G* ZPEs for Tyr-Gly, Tyr-Gly-Gly, and Tyr-Gly-Gly-Phe).

Figure 37 shows the relative energies of the twenty most stable Tyr-Gly, Tyr-Gly-Gly, Tyr-Gly-Gly-Phe and Tyr-Gly-Gly-Phe-Leu conformers (based on the DFT geometries). For the most stable conformers found for the Tyr-Gly dipeptide a significant variation was observed between the optimized structures at the B3LYP and the MP2 level of theory. These variations are expected to be caused by dispersion effects from the phenyl group of tyrosine (1). MP2 optimizations give more folded structures. Additionally two conformers optimised to completely different structures (conformers 4 and 6). In conformer 4 a change of the φ angle from 179 to 74° allows a very distant interaction between the carboxylic C=O and the tyrosine OH.

Conformer 6 has become much more folded (the 'book' effectively 'closes'), allowing an interaction between the carboxyl hydroxyl of glycine and the tyrosine OH. For Tyr-Gly-Gly and Tyr-Gly-Gly-Phe the variation between the relative energies at the B3LYP and the MP2 level of theory is getting much larger. For Tyr-Gly-Gly-Phe-Leu the variation between the relative energies at the B3LYP and MP2 is not as large as the variation found for the tetrapeptide. This is probably due to the low number of conformers (only the 20 most stable conformers) found from Hartree-Fock optimizations that were selected and optimised at the B3LYP level of theory. In contrast, only minor differences were observed between MP2 and B3LYP for the most stable conformers found for Gly-Gly-Gly. This is probably due to the absence of an aromatic residue in this peptide and consequently due to the much smaller effect of dispersion for this molecule.

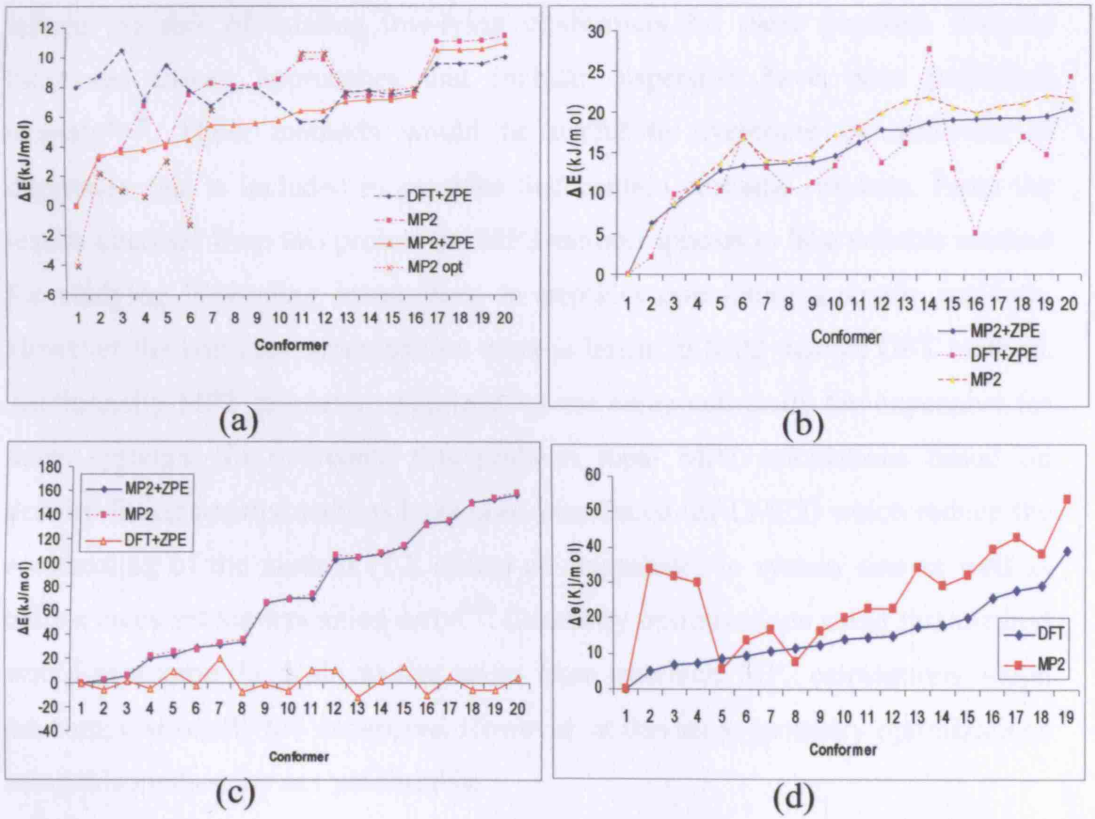


Figure 37. Comparison of the B3LYP, B3LYP+ZPE, MP2+ZPE relative energies of the 20 most stable Tyr-Gly (a), Tyr-Gly-Gly (b), Tyr-Gly-Gly-Phe (c), Tyr-Gly-Gly-Phe-Leu (d) conformers (based on the DFT geometries). The ZPEs were computed with B3LYP/6-31+G* and scaled by 0.976.

The application of faster semi-empirical methods can predict similar structures as Hartree-Fock geometry optimizations. For the dipeptide both PM3 and AM1 gave structures similar to those optimised with the HF/3-21G*. For the tripeptide only AM1 gave structures similar to those obtained at the HF/3-21G* level of theory. So it is recommended to start optimizations using HF/3-21G* from the optimized structures resulting from semi-empirical methods.

It was found that B3LYP is not good enough to describe the conformational preferences of peptides containing aromatic residues. More accurate methods need to be employed for the accurate description of the peptide energetics. However due to computational restraints it is too expensive to use more accurate methods. So, the exploration of the conformational features of these peptides remains a difficult task. A combination of different approaches should be considered to reduce the risk of missing low-lying conformers for these peptides. Density functional theory approaches that include dispersion have been published recently[3, 4]. These methods would be useful to overcome the problem of dispersion that is included in peptides that contain aromatic residues. From the results obtained from this project the MP2 method appears to be a suitable method for studying H-bonding interactions in peptides containing aromatic residues. However the basis set superposition error is larger in MP2 than in DFT method. Additionally MP2 geometry optimizations are computationally too expensive for these systems. To overcome this problem local MP2 calculations based on density-fitting approximations have been introduced (df-LMP2) which reduce the cost scaling of the method (1-2 orders of magnitude) in system size as well as reduce basis set superposition error[5-8]. Geometry optimizations using this method would overcome the problem that arises from canonical MP2 calculations which are computationally too expensive. However, at this stage geometry optimizations using this method are not yet feasible.

References

1. Toroz, D.; Van Mourik, T., *Molec. Phys* **2006,** 104, (4), 559.
2. Toroz, D.; Van Mourik, T., *submitted.*
3. Grimme, S., *J. Comput. Chem.* **2004,** 25, 1463-1473.
4. Grimme, S., *J. Chem. Phys.* **2006,** 124, 034108.
5. Hetzer, G.; Pulay, P.; Werner, H.-J., *Chem. Phys. Lett.* **1998,** 290, 143.
6. Hetzer, G.; Schutz, M.; Stoll, H.; Werner, H.-J., *J. Chem. Phys.* **2000,** 113, 9443.
7. Schutz, M.; Hetzer, G.; Werner, H.-J., *J. Chem. Phys.* **1999,** 111, 5691.
8. Werner, H.-J.; Manby, F. R.; Knowles, P. J., *J. Chem. Phys.* **2003,** 118, 8149.