Analytical functional representation of quantum chemical potential energy curves and surfaces

Ph.D. Dissertation

Anita Rágyanszki

Supervisor:
Dr. Béla Viskolcz

Doctoral School of Chemistry

Faculty of Science and Informatics
University of Szeged

Szeged, Hungary
2015

"The distance between those abstruse quantum mechanical calculations and the patient bed may not be as great as believed"

<div align="right">

Albert Szent Györgyi
Introduction to Submolecular Biology
Academic Press 1960

</div>

Disclaimer

Every science starts as an empirical science which is based on an empirical observation and therefore it has a very limited predictive value. An empirical science become an exact science after its mathematization has been completed.

This process happened to physics with Newton in the 17th century. The mathematization of chemistry has been initiated by Schrödinger in 1926, and it has not been completed yet.

It would be practical to start the mathematization of chemistry by starting with molecular conformation change since such a process does not involve bond breaking and bond making. However, it would be extensively useful because it may lead to a full understanding of the secret of protein folding which could not be uncoded for centuries.

The present dissertation aims to find mathematical functions for the representation of conformation changes of small peptide residues in order to take the first steps of a bottom up solution of protein folding. However, the full mathematization of chemistry is not claimed this dissertation.

Abbreviations

A	Matrix of the critical points
Ala	N- Acetyl- Alanine N- methylamide
BO	Born-Oppenheimer Approximation
B3LYP	Exchange-correlation functional
CCSD	Coupled cluster method
C_3	Propane
ΔE	Relative Energy
DFT	Density Functional Theory
fs	Femtosecond
G	Hessian matrix
Gly	N- Acetyl- Glycine N- methylamide
HF	Hartree-Fock
L	Vector lenghts
MM	Molecular Mechanics
MD	Molecular Dynamics
MP2	Møller–Plesset perturbation theory
PEC	Potential Energy Curve
PES	Potential Energy Surface
PEHS	Potential Energy HyperSurface
R^2	Statistical characterization of accuracy (R-square)
RMSE	Root Mean Squared Error
SA	Simulated annealing
SMD	Truhlar and coworkers' solvation model
TDS	Time-delayed summation
QM	Quantum Mechanics
Val	N- Acetyl- Valine N- methylamide

Table of Contents

1. Introduction

1.1. Historic Predicament

"The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved"[1]

Paul Dirac

Computer technology plays an important role in theoretical and computational chemistry. This specific technology related to processor speed and architecture, computer memory, networking, graphical displays, software development and data storage and data transmission. Almost a century ago Schrödinger's quantum theory[2], in 1926, described the electron distribution in the hydrogen atom. However, for a long time the computer technology was not sufficiently developed to solving complicated cases or complex systems. During the last half century, the computer technology has improved a lot. For instance, in the 1960s, the third generations of computers have appeared[3]. These computers included integrated circuit and semiconductors. In that decade the FORTRAN programming language was relatively new. FORTRAN, the formula translation language, well suited to linear algebraic problems[4] and it was influenced by developments in computational chemistry. Consequently, in those years the use of computers began to enter quantum chemistry. In the 1960's the quantum chemist believed that all of the problems in chemistry (*Figure 1*) could be solved by computation within a few decades.
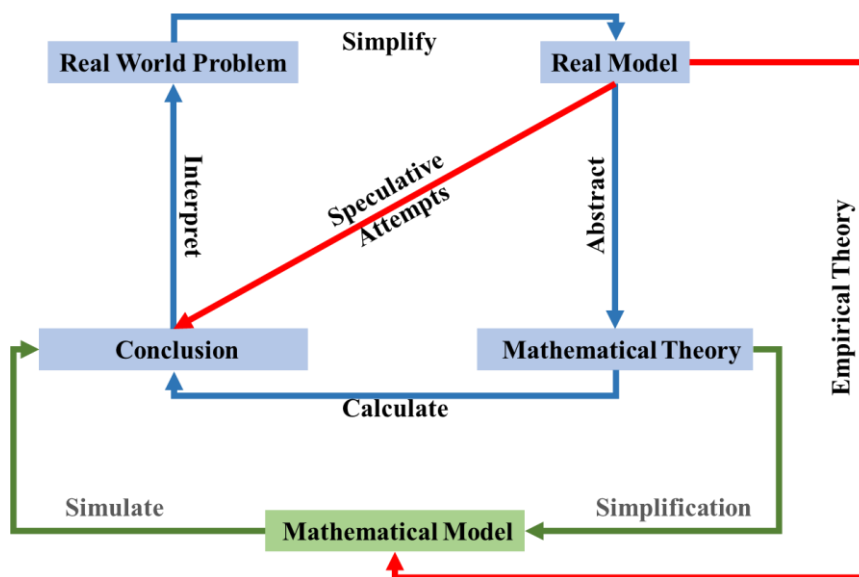

Figure 1 *The cycle of the problems in mathematical modelling*

Theoretical chemistry has become a great field within many new research areas such as in computational chemistry, biochemistry, mathematical chemistry and molecular modelling. Many questions have been answered in the last fifty years. However, one unsolved problem is the biggest conundrums of our century in the field of theoretical chemistry, is the protein folding problem. The upper square connecting the blue boxes (*Figure 1*) represents a work flow of the mathematical model. In contrast to that, the lower rectangle involving the computer model, is used to replace the details of the TDS (Time-delayed summation) computations. For example, Kepler's law[5], prior to Newton's law[6] of mechanics, for the motion of the planets around the Sun covered identical segment of the ellipse of equal time (*Figure 2*) is an example for such a model before rigorous computation was possible.
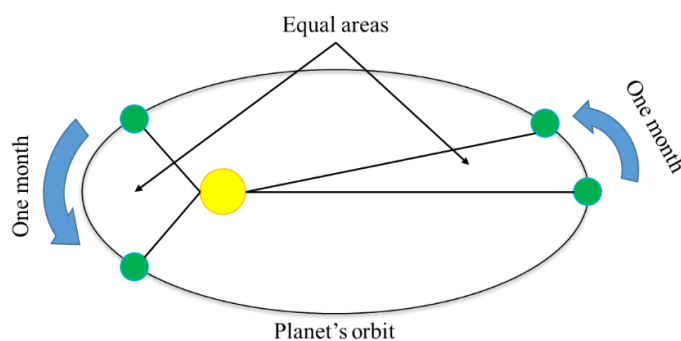

Figure 2 planets move around the Sun in an elliptical orbit

"The most vitally characteristic fact about mathematics is, in my opinion, its quite peculiar relationship to the natural sciences ... In modern empirical sciences it has become more and more a major criterion of success whether they have become accessible to the mathematical method or to the near-mathematical methods of physics. Indeed, throughout the natural sciences an unbroken chain of successive pseudomorphoses, all of them pressing toward mathematics, and almost identified with the idea of scientific progress, has become more and more evident. Biology becomes increasingly pervaded by chemistry and physics, chemistry by experimental and theoretical physics, and physics by very mathematical forms of theoretical physics. There is a quite peculiar duplicity in the nature of mathematics. One has to realize this duplicity, to accept it, and to assimilate it into one's thinking on the subject. This double face is the face of mathematics, and I do not believe that any simplified, unitarian view of the thing is possible without sacrificing the essence."[7]

John von Neumann

One of the reasons that protein folding has not yet been solved is due to the big data associated with it. Reducing this data set is a possible approach towards simplification. In order to make these simplifications, we first have to deeply understand mathematical properties of conformational spaces.

The other reason for this delay is due to the fact that we still need many mathematical experience, for example, in the field of the functional analyses, numerical analyses, mathematical modeling, differential equations, optimization, nonlinear least squares, multidimensional approximation of functions and statistical classification. Consequently, the problem of the protein folding became a diverse mathematical challenge[8,9].

1.1. The protein folding problem

The protein folding problem is one of the 125 largest unsolved problems in science in addition to "what is the Universe made of" or "are we alone in the Universe"[10]. There were several attempts to solve this century old unanswered question[9,11–14]. To solve this problem we need many hundreds of supercomputer years. An average protein could containing hundreds of amino acids each of which are containing several dozens of atoms (*Figure 3*). This is the reason that the protein folding problem is too complex and therefore it remained unresolved.
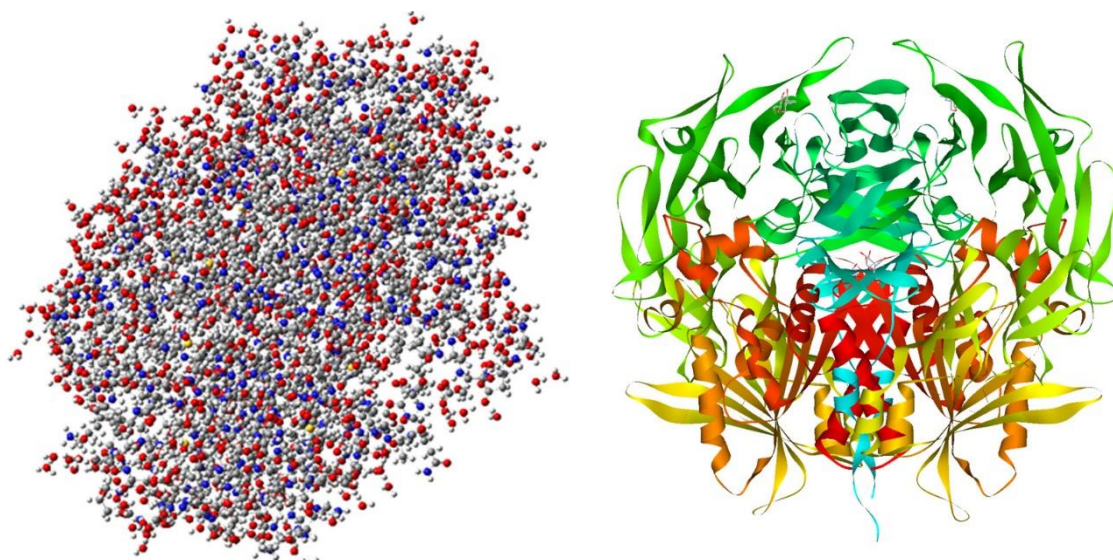

Figure 3 *The molecular structure of Dipeptidyl peptidase and the simplified graphical representation*

The left hand side of *Figure 3* shows an actual molecular structure of Dipeptidyl peptidase protein[15]. Clearly, the mathematical representation of such biological architecture is not trivial and the mechanics whereby the polypeptide chain will reach this folded structure

is not a real simplification. For this reason, graphical models are used to sown in the right hand side in *Figure 3*. The application of *Figure 1* for the current problem is depicted in *Figure 4* The lower part of this figure shows the role of MM which is the classical model of the QM problems.
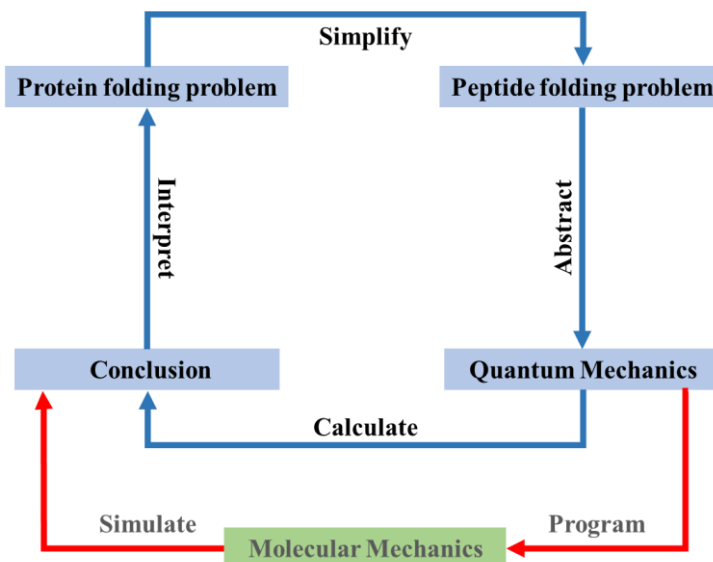
Figure 4 *The cycle of the problem solving in quantum chemistry*

For example the actual anharmonic vibrational potential of a bond stretch is simplify by a parabolic potential (*Figure 5*). This could be one of the models in the lower green box of *Figure 4*.
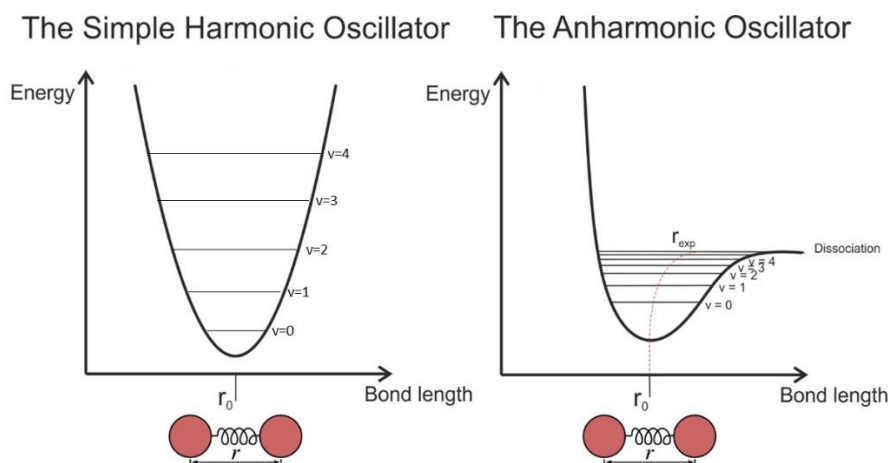
Figure 5 *The simple harmonic potential and the anharmonic potential energy curves*

Proteins are large polymeric molecules which consist of amino acid residues. The protein folding problem is simply a definition of how the different amino acids determine the three dimensional atomic structure of the protein. If we wish to give a solution of it

we need to solve three different sub-problems: the folding code-the thermodynamic question [16], the protein structure prediction-the computational problem[17] and the folding mechanism-the direction and the speed of folding and refolding change[18,19].

The above problem has been cited, poetically, in the following way, in 1904:

"I always feel like running away when any one begins to talk about proteids in my presence. "In my youth I had a desire to attack these dragons, but now I am afraid of them. -They are unsolved problems of chemistry; and let me add, they are likely to remain such for generations to come. -Yet every one who knows anything about chemistry and physiology, knows that these proteids must be understood, before we can hope to have a clear conception of the chemical processes of the human body."[20]

Ira Ramsen

1.2. The peptide folding

"All things are made of atoms, and everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

Richard P. Feynman

Each poly-peptide consist of several of the twenty different natural (DNA coded) amino acids and form a long flexible chain. The structure of the proteins are defined by four ways. The first step, which defines the sequence, is the primary structure (*Figure 6*) of the amino acid specifying a long flexible chain. The amino acids in this chain link to each other by peptide bonds, which are a covalent bonds. Sometimes they are also interconnected by disulphide bridges which are also covalent linkages[21].

Gln——Arg——Phe——Ser——Arg

Figure 6 *The primary structure of the opiorphin pentapeptide*

Further structural features (*Figure 7*) of the proteins are secondary structural elements for example an alpha helices, a beta-pleated sheet, beta-turns or random coils. These structures are stabilized together by hydrogen bonds. In contrast to that, the tertiary structure means the 3-dimensional shape of the whole protein. This tertiary structure (*Figure 8*) is held together by interactions of the side chains by several ways, for example ionic interactions, hydrogen bonds van der Waals dispersion forces and sulphur bridges.
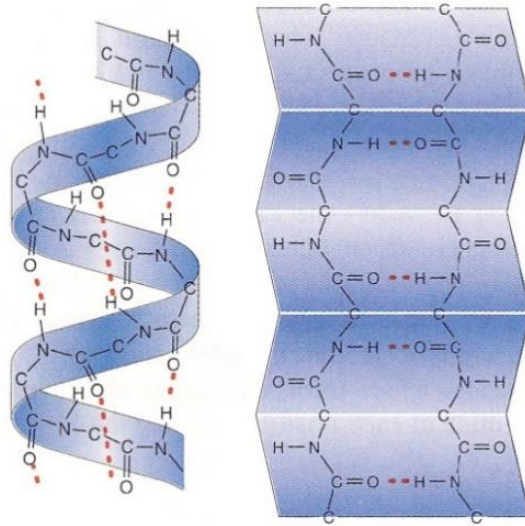
Figure 7 *The secondary structure of the proteins: the alpha helix (in the left hand side) and the beta pleated sheet (right hand side).*
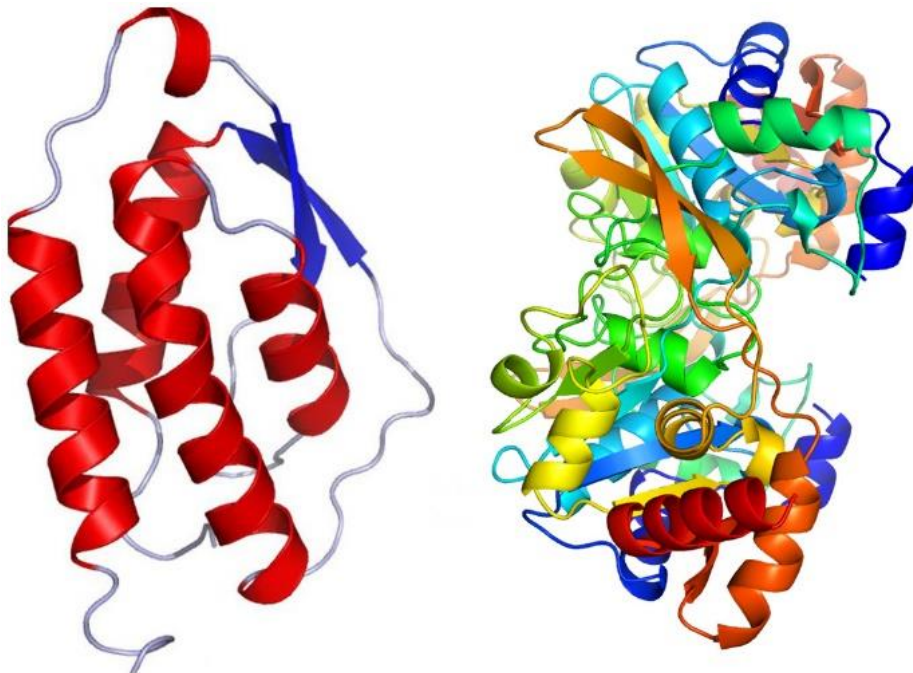
Figure 8 *The tertiary structure (in the left hand side) and the quaternary structure of (right hand side) the proteins*

Very often, the proteins make up protein subunits, which are multiple polypeptide chains. This subunits interact with each other and they form a protein complex. This is what we known as quaternary structure (*Figure 8*) of the protein [22].
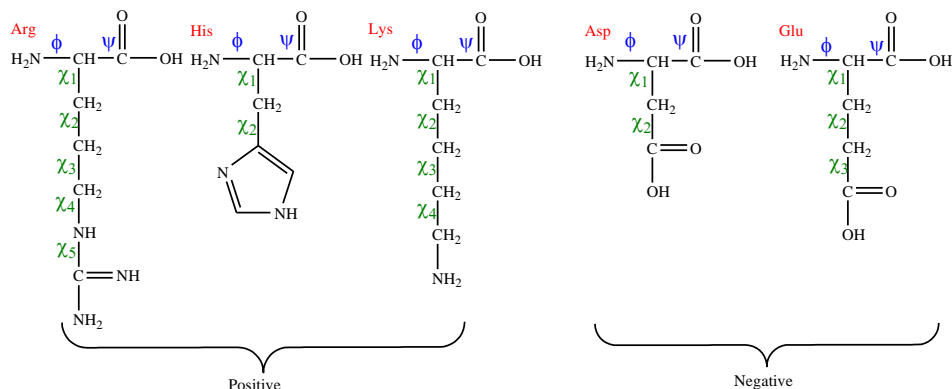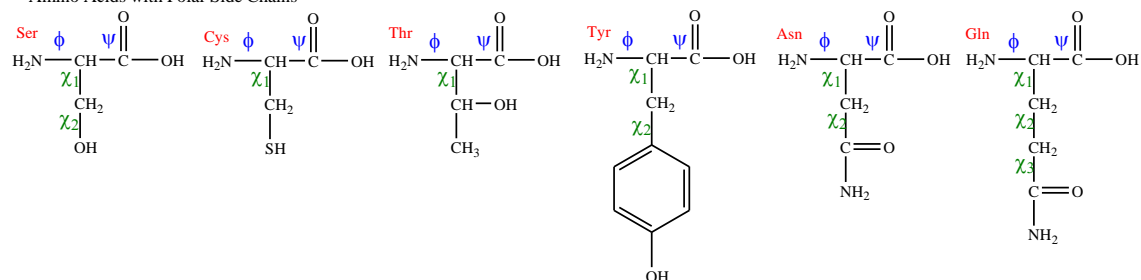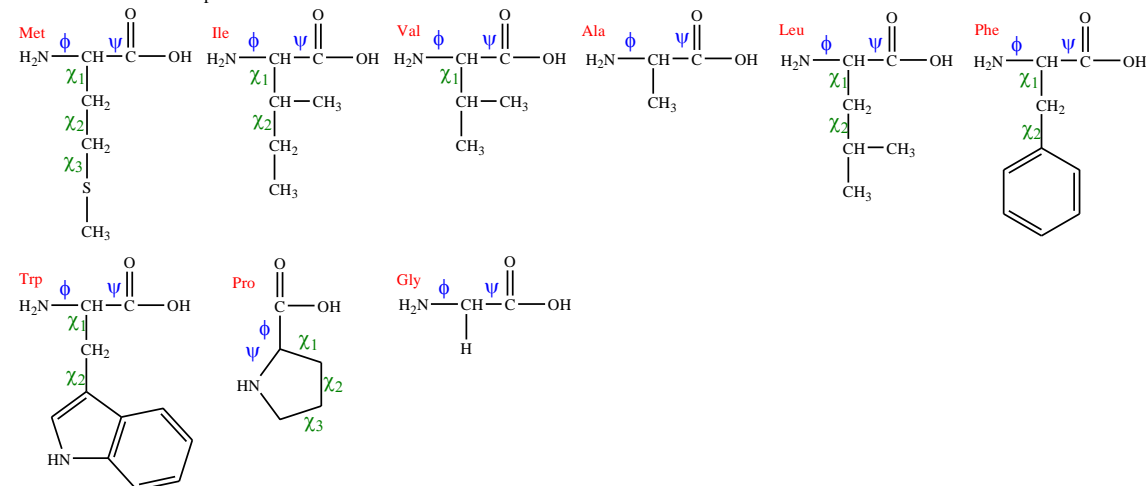
Figure 9 *The twenty different amino acids*

About half the atoms of the amino acid residues are hydrogen atoms and the others are carbon, nitrogen, oxygen and sulfur atoms (*Figure 9*). The side chains (R) of the 20 different amino acids contain 1-18 atoms[23]. The flexibility may also involve several dihedral angles of the side chain variation up five dimensions (5D)[24]. The backbone, in *Figure 10*, has two dihedral angles and each of this is a 3-fold rotational potential, consequently, it is expected to exhibit $3^2=9$ conformers.
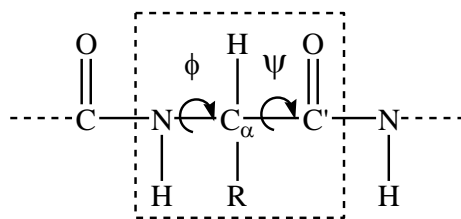
Figure 10 *Amino acid residue with the side chain R*

A peptide unit, which includes three amino acid residues (n=3) with the side chains, has six backbone torsion angles (discounting ω_i angles) (*Figure 11*). This many torsion angles, in ideal case, cause $9^3=729$ stable geometries [1a] and 4374 transition states [1b] and these numbers are increasing rapidly, if a new amino acid is added.

$$N_0 = \left(3^2\right)^n = 9^n \tag{1a}$$

$$N_1 = nN_0 = n \cdot 9^n \tag{1b}$$

In [1] the N_0 is the number of minima, N_1 is the number of transition states and n is the number of amino acid residues.
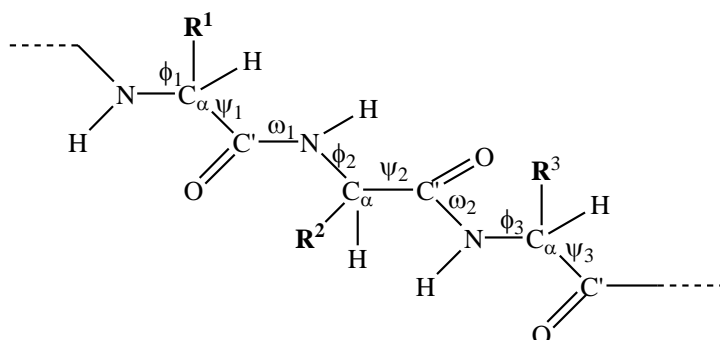
Figure 11 *Three peptide units with the side chains R^i and dihedral angles of ω_i, ϕ_i and ψ_i*

In general, amino acids are of two relative configurations but natural amino acids are of the same L-chirality.

The Ramachandran plots[25] (*Figure 12*) show the torsion angles (φ, ψ) for all possible residues. The torsion angles describe the conformational parameters and they are important for controlling peptide folding. The fundamentals of the plot are the Ramachandran angles. In the protein, the torsion angle between the C_α and the N-terminus is called φ, and the torsion angle between the C_α and the C-terminus is called ψ. The plot describes the rotations of the peptide backbone, where φ and ψ are independent, and specify the potential energy surface.
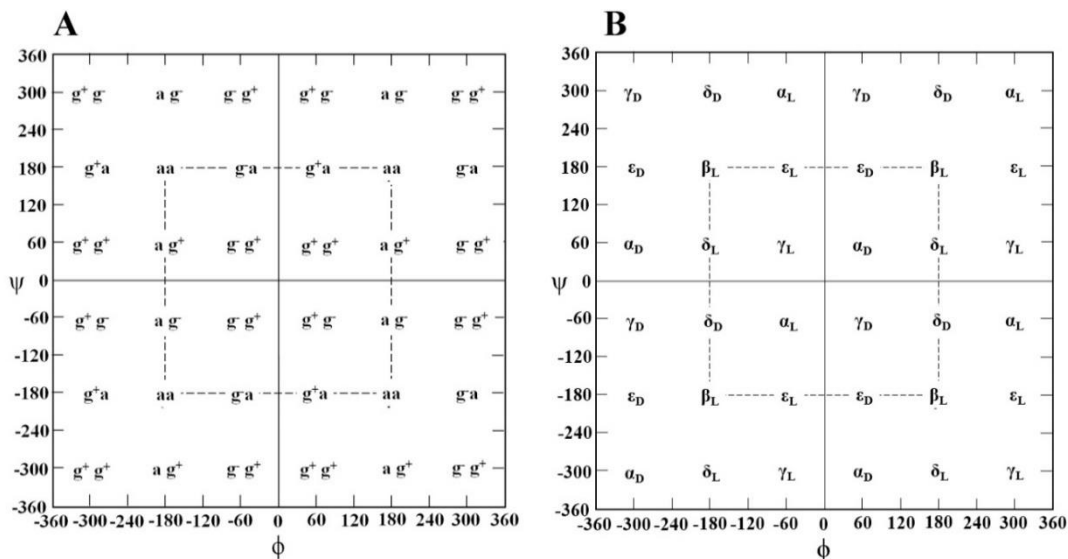
Figure 12 *Schematic topology of conformational PES map showing the -360° to 360° range of dihedral angles of double rotors. **A**: Saturated molecules like propane (CH_3-CH_2-CH_3), which might be characteristic to side-chain orientation **B**: Peptide Backbones (Ramachandran Map).*

1.3. Energy calculations by Quantum Mechanics

Nature consists of molecules which are made of atoms and the atoms are built up from nuclei and electrons. Quantum mechanics is a tool which describes how molecules work at the atomic molecular level. In the first decade of the 20th century, quantum mechanics has led to a revolution to understanding the properties of the atoms, the nuclei and the electrons. The Schrödinger equation[2] [2], the fundamental equation of the quantum mechanics, cannot be solved analytically except for the simplest system, the hydrogen atom.

The time independent Schrödinger equation:

$$\hat{H}\Psi = E\Psi \qquad [2]$$

where $\hat{H}$ is a Hamiltonian, E is energy, and Ψ is an N-body wave function.

Usually the number of the particles (N) are too large therefore we cannot solve this equation analytically.

A very reasonable solution of the time independent Schrödinger equation is the Born-Oppenheimer approximation. In the case of this approximation the nuclei is heavier than the electron. This approximation allows that, to define molecules by a geometry structure determined by the nuclear positions. The Born-Oppenheimer[26] approximation gives a reasonable version of the many-body Hamiltonian [3]:

$$\hat{H} = \sum_i -\frac{1}{2}\nabla_i^2 + \sum_i \sum_a \frac{Z_a}{|r_i - d_a|} + \frac{1}{2}\sum_i \sum_{j \neq i} \frac{1}{|r_i - r_j|} + \frac{1}{2}\sum_a \sum_{b \neq a} \frac{Z_a Z_b}{|d_a - d_b|} \qquad [3]$$

If the nuclei are in a fixed positon, the nuclear repulsion which is the last term becomes a constant and the kinetic term is not shown since it vanishes from the total Hamiltonian. If we assign the electronic Hamiltonian only then we obtain the potential energy by adding nuclear repulsion term to the electronic energy. Thus, the electronic Schrödinger equation must be solved for various nuclear positions and this leads to a potential energy surface.

One of the simplest theories of the solution of the many-body Hamiltonian is the Hartree-Fock theory, which is the basis of the Molecular Orbital (MO) Theory [27–29]. This is based on a single Slater determinant of spin-orbitals, but due to the involvement of double occupancy of orbitals it does not include the correlation of the electrons. The electronic structure described by a single slater determintant [4].

$$\phi_{HF} = \hat{A}\varphi_1(1)\varphi_2(2)\varphi_3(3)\ ...\varphi_n(n) = D_o \qquad [4]$$

where D_o is the Slater determinant, φ_i are the spin-orbitals and $\hat{A}$ is the antisymmetrizer operator.

In contrast to the MO Theory, the Density functional theory (DFT)[30–33], the most successful method to calculate the electronic structure which is derived from density function and can be used for solutions without experimental data. DFT is given by two theorems which were introduced by Hohenberg and Kohn[30].

The first theorem is: non-degenerate ground states, the external potential is determined, to within an additive constant, by the electron density. The ground state energy can be obtained variationally: the density that minimizes the total energy is the exact ground state density[30,34].

To create the molecular orbitals we need basis functions, also called basis set, which are a set of functions. The basis functions are a linear combination of the Gaussian expanded function of the atomic orbitals. One of the most used basis sets are the Pople basis sets. This is includes the set of STO-*NG* (where N primitive Gaussian type orbitals fitted to a single Slater type orbital) minimal basis set and also there are a large number of split valance basis sets. The Slater basis set has one basis function for the core and two or more for the valance. For example Pople's split valance double-zeta basis set is called 6-31G. In this case the core orbital made of 6 Gaussians and the valance (each of 2s, 2p$_x$, 2p$_y$, 2p$_z$) described by 2 orbitals the first is contracted from 3 Gaussians and the second is

represented by 1 Gaussian. To polarization the non-hydrogen atoms we need to add a d polarization function (6-31G* or also called 6-31G(d)), or to polarization the hydrogen atoms we need to add a plus p polarization function (6-31G** or also called 6-31(d,p))[35].

1.4. Conformational Potential energy surfaces as a conceptual tool of chemistry
1.4.1. The landscape of the Conformational Potential Energy Surfaces

The Potential Energy Surfaces (PES) is a function of two independent variables (x, y) and they are playing an important role in chemistry.

$$E = f(x, y) \tag{5}$$

They contain information about the molecular structure, thermodynamic properties and chemical reactivities. By exploring a PES, we can find minima, first-order saddle points, which are the transition states (TS) for reactions, as well as maxima[36].

A typical illustration of a segment of a PES that contains a central minimum. At the four adjacent edges there are saddle points, or TS, and four maxima are located at the four corners as shown in *Figure 13*.
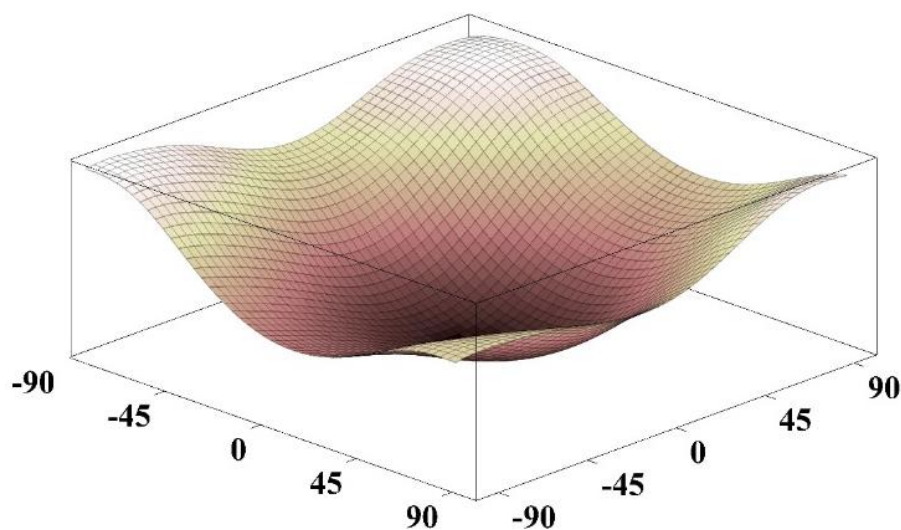
Figure 13 The landscape of a model conformational potential energy surface

These critical points, minima, saddle points and maxima, can be characterized by the first and the second derivatives of the functional form of the PES shown in equation [5] and [6].

$$E = E(\phi) \qquad\qquad \text{a curve} \tag{6a}$$

$$E = E(\phi, \psi) \qquad\qquad \text{a surface} \tag{6b}$$

$E = E(\phi, \psi, \chi)$ a hypersurface   [6c]

The first derivatives *Figure 14* vanish at the location of the critical point. The type of any critical point, let it be minimum, TS or maximum, can be assessed from the second derivative of the corresponding function of [6]. In the case of equation [6a] the second derivative will be positive for a minimum and it will be negative for a maximum.
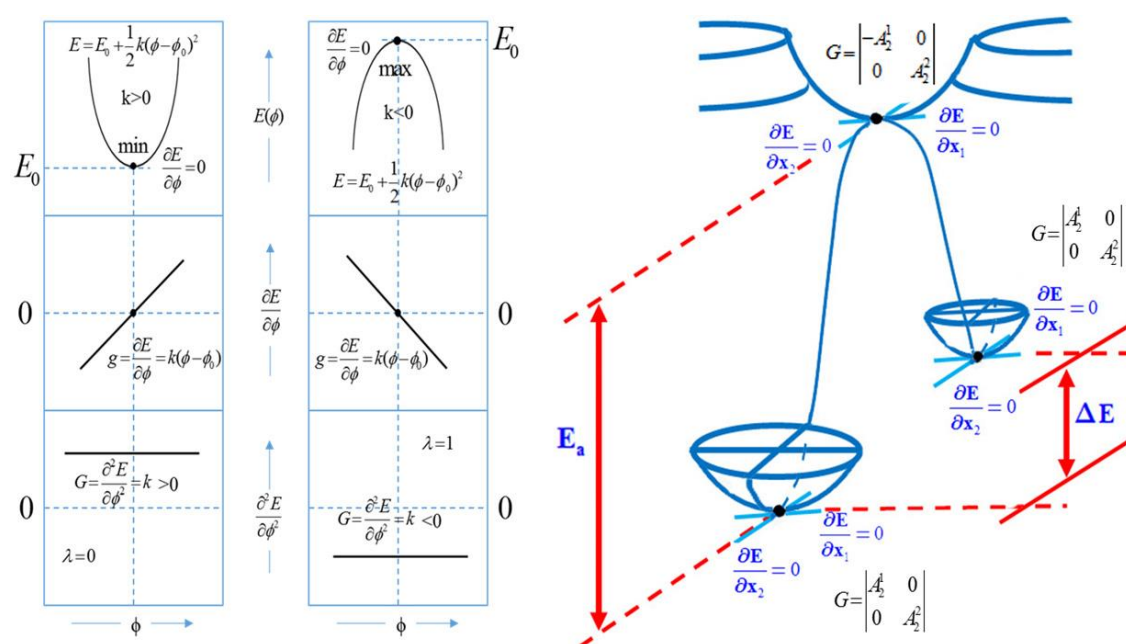
Figure 14 *The characteristic of the critical points, minimum and maximum, of a potential energy curve and surface. The g is the gradient, G is the Hessian, k is the force constant and the λ is the index of the critical points.*

In the case of equation [6b] both partial first derivatives will be zero at the critical point location and the two second partial derivatives will be either positive or negative. If the critical point a minimum, both second derivative are positive and if it is a maxima both second derivatives are negative. In contrast to these for the TS one of the second derivatives is positive and the other one is negative.

The number of negative second derivatives are called the index of the critical point (λ). For minima λ=0, for TS λ=1 and for maxima λ=2. Very often only the index (λ) of the critical point is shown in graphical representations[24].

Although all chemical changes may be described by a PES, nevertheless conformational PES are representing a special case which will be the subject of the present dissertation. They are special in the sense, that these PES are periodic (*Figure 15*).
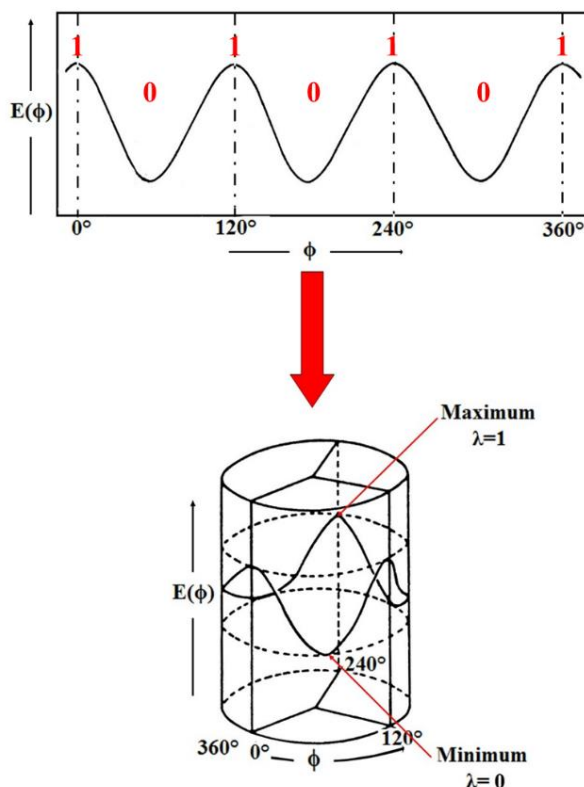
Figure 15 *The periodicity of a Potential Energy Curves*

Such periodicity is due to the fact that a rotation about the dihedral angle after 360° will be identical to the 0°. In the case of ethane the rotational potential about the C-C bond can be plotted either on a sheet or on the envelope as a cylinder. At the bottom on the topological features of this process circle which has a single minimum and a single maximum if we take the height as dependent variable [37].

For these reason the independent variables like x, y or ϕ, ψ maybe represented either by Cartesian or circular coordinates as shown by *Figure 16*.
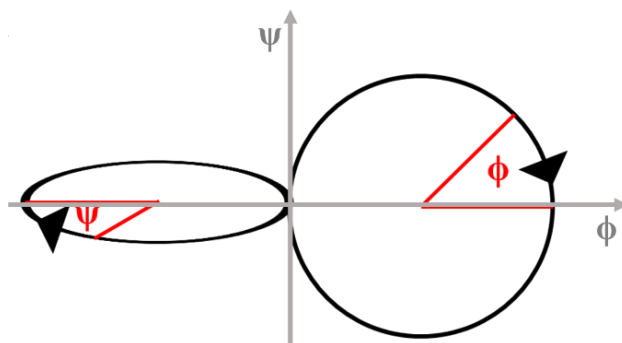
Figure 16 *Periodicity of rotation about a pair of dihedral angles (ϕ, ψ)*

A double rotor represents a more complicated problem since in the case of one dimensional PES one can roll it up only once. However, in the case of PES of two independent variables we may role up the surface along each of the independent variables. In the first step we generate a cylinder and in the second step we roll up the cylinder in a two-dimensional torus or a simply a torus[38,39]. This is illustrated in *Figure 17*. The torus is the topological representation of the two dimensional conformational PES.
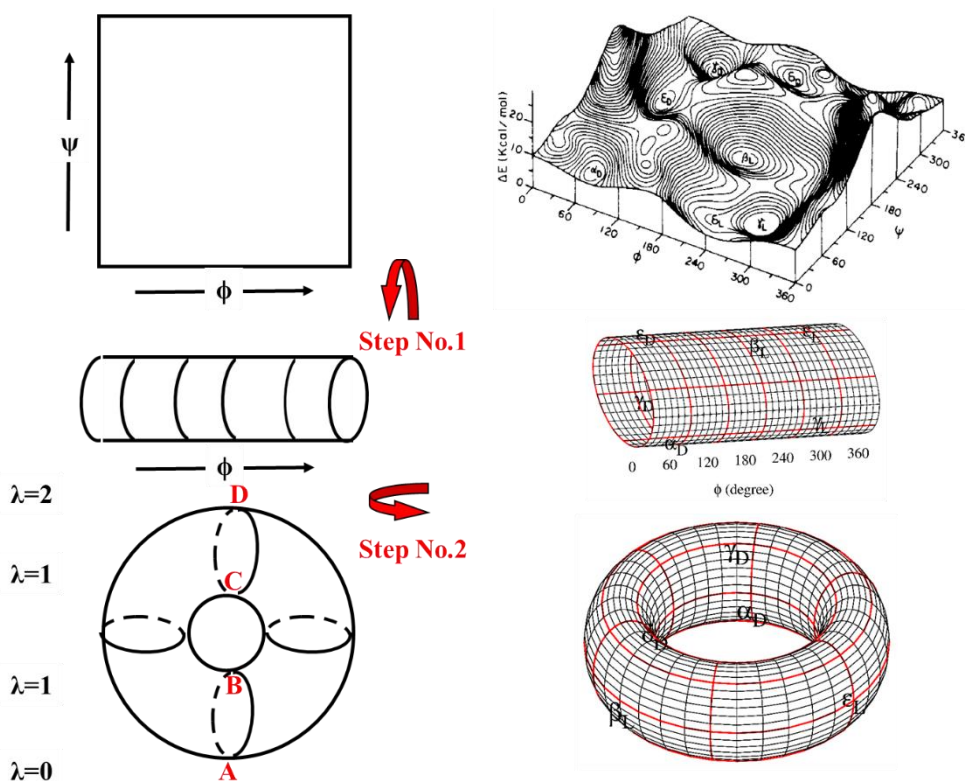
Figure 17 *A torus representation of the two dimensional PES.*

This figure clearly shows that by taking the height on the torus as dependent variable, we can recognize points A, B, C and D which correspond to one minimum ($\lambda=0$) two TS ($\lambda=1$) and one maximum ($\lambda=2$). Special toroidal representation is shown *Figure 18* which the shape of the torus is distorted proportionally to the energy values of the surface[40].
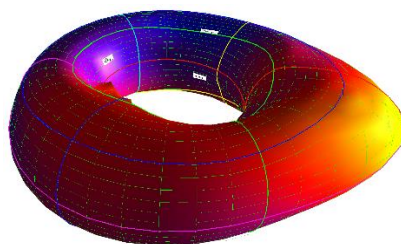
Figure 18 *Special toroidal representation*

16

Due to the periodicity of conformational change, the unit cells of the PES may occur more than the once on a full PES. If the rotation is about a tetrahydrocarbon (sp^3) then the periodicity is expected three-fold. Each unit cell would contain a single minimum which may occur three times along a full rotation.

The following 3 examples represent (*Figure 19, 20*) the one- (A), two- (B) or three (C) dimensional cases involving ethane, propane and iso-butane.
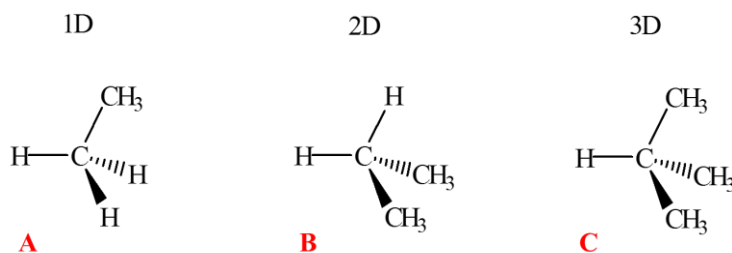
Figure 19 *The three example which represent the one-(A), two-(B) or three (C) dimensional rotational cases, namely ethane (1D), propane (2D) and iso-butane (3D).*
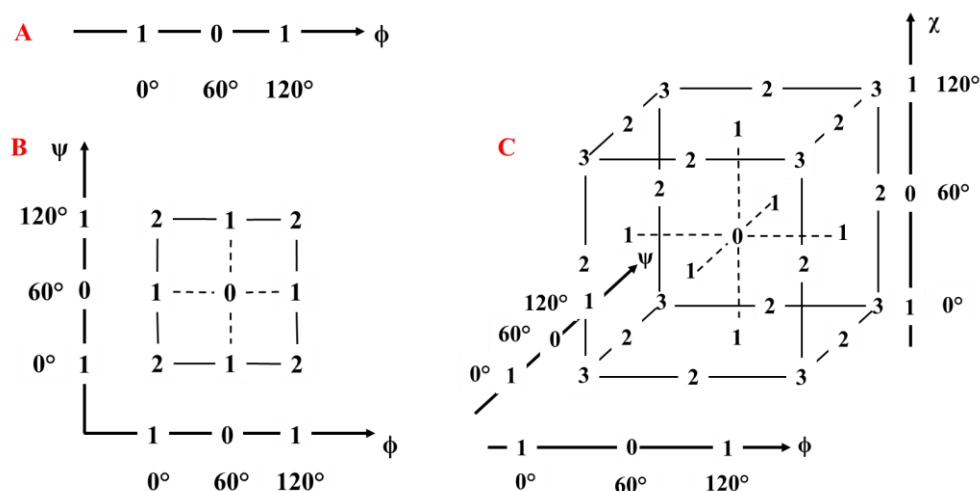
Figure 20 *Topological representation of unit cells of a potential energy curves (A), surface (B) and hypersurface (C) for the rotation of a single (A), two (B) and three (C) CH₃ groups.*

In the full (0°-360°) *PES* the one-dimensional case would have 3 minima, the two-dimensional $3^2=9$ minima and the three-dimensional is expected $3^3=27$ minima.

The number of critical points, of index λ, in a unit cells correspond to the Betti numbers, $β_λ$, which is the topological name of the binomial coefficient shown in equation[41] [7].

$$\beta_\lambda = \binom{n}{\lambda} = \frac{n!}{\lambda!(n-\lambda)!} \qquad\qquad [7]$$

The Betti numbers are shown in a Pascal triangle up to n=4 in *Table 1*. The application of these for curves, surfaces and hypersurfaces are summarized in *Table 2*.

***Table 1** Betti numbers for conformational PESs*

n	β_λ				
1			1	1	
2		1	2	1	
3	1	3	3	1	
4	1	4	6	4	1
	β_0	β_1	β_2	β_3	β_4

***Table 2** The application of potential energy curves, surfaces and hypersurfaces*

Type of potential energy function

Index of critical point	n=1 Curve E=(ϕ)	n=2 Surface E=(ϕ, ψ)	n=3 Hypersurface E=(ϕ, ψ, χ_1)	n=4 Hypersurface E=($\phi, \psi, \chi_1, \chi_2$)
$\lambda=0$	minimum	minimum	minimum	minimum
$\lambda=1$	maximum	saddle	1° saddle	1° saddle
$\lambda=2$	-	maximum	2° saddle	2° saddle
$\lambda=3$	-	-	maximum	3° saddle
$\lambda=4$	-	-	-	maximum

There are two alternative sum rules: one for the Betti numbers[42], β_λ, and one for the actual number of critical points, N_λ, of the full surfaces. These are shown in the equation [9] and [11].

For the Betti numbers:

$$\beta_0 - \beta_1 + \beta_2 - \ldots = \sum_{\lambda=0}^{n} (-1)^{\lambda} \beta_{\lambda} = 0 \qquad [8]$$

where

$$\beta_{\lambda} = \binom{n}{\lambda} \qquad [9]$$

For the number of critical points:

$$N_0 - N_1 + N_2 - \ldots = \sum_{\lambda=0}^{n} (-1)^{\lambda} N_{\lambda} = 0 \qquad [10]$$

where

$$N_{\lambda} \geq \beta_{\lambda} \qquad [11]$$

The relationship between the Betti numbers and the actual number of critical points are shown in equation [12].

$$N_{\lambda} = \kappa . \beta_{\lambda} \qquad [12]$$

The scaling factor, κ, is the arithmetical product of periodicities (m_i) of the interval mode of rotation (*Figure 19*).

For ideal surfaces (Mezey equation)[43–45]:

$$N_{\lambda} = \left[\prod_{i=1}^{n} m_i \right] \beta_{\lambda} \qquad [13]$$

where m_i is the topological periodicity of the component potential curve along the i-th variable.

m=4   m=3   m=2
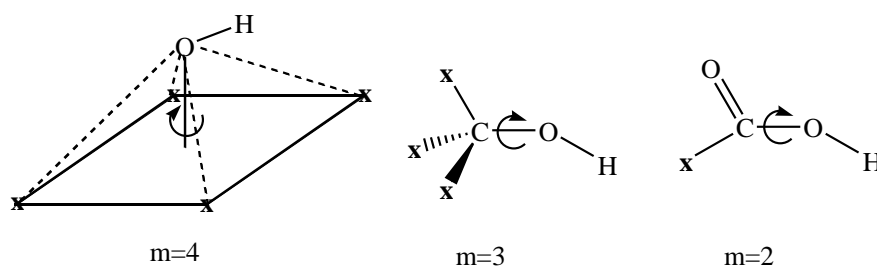
Figure 21 Molecular structure will pre-determine the periodicity (m_i). The variation of m for an –OH rotation.

If $m_1 = m_2 = m_3 = \ldots$ then

$$N_{\lambda} = m^n \beta_{\lambda} = m^n \binom{n}{\lambda} \qquad [14]$$

For non-ideal surfaces[46]:

$$N_\lambda = \left[\prod_{i=1}^{n} m_i \right] \beta_\lambda + \Delta N_\lambda \qquad\qquad [15]$$

if $\Delta N_\lambda > 0$ for critical points created, and $\Delta N_\lambda < 0$ for critical points annihilated.

A single critical point cannot be annihilated alone only by pairs. If one which is annihilated will appear with negative sign, then the other must appear with a positive sign so that the alternative sum rule, will not be violated[47].

Different cases for the creation and the annihilation are illustrated in *Figure 22 and 23*.
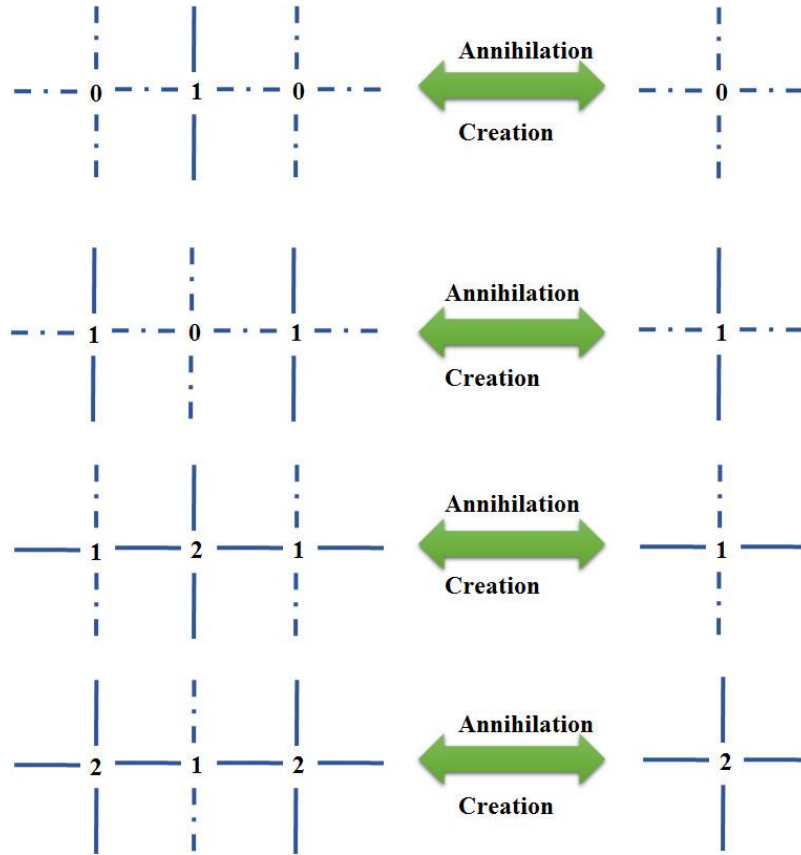
Figure 22 Different cases for the creation and the annihilation of critical points

Non-ideal PES

$N_0-N_1+N_2=0$
$8-17+9=0$

Ideal PES

$N_0-N_1+N_2=0$
$9-18+9=0$

Non-ideal PES

$N_0-N_1+N_2=0$
$10-19+9=0$
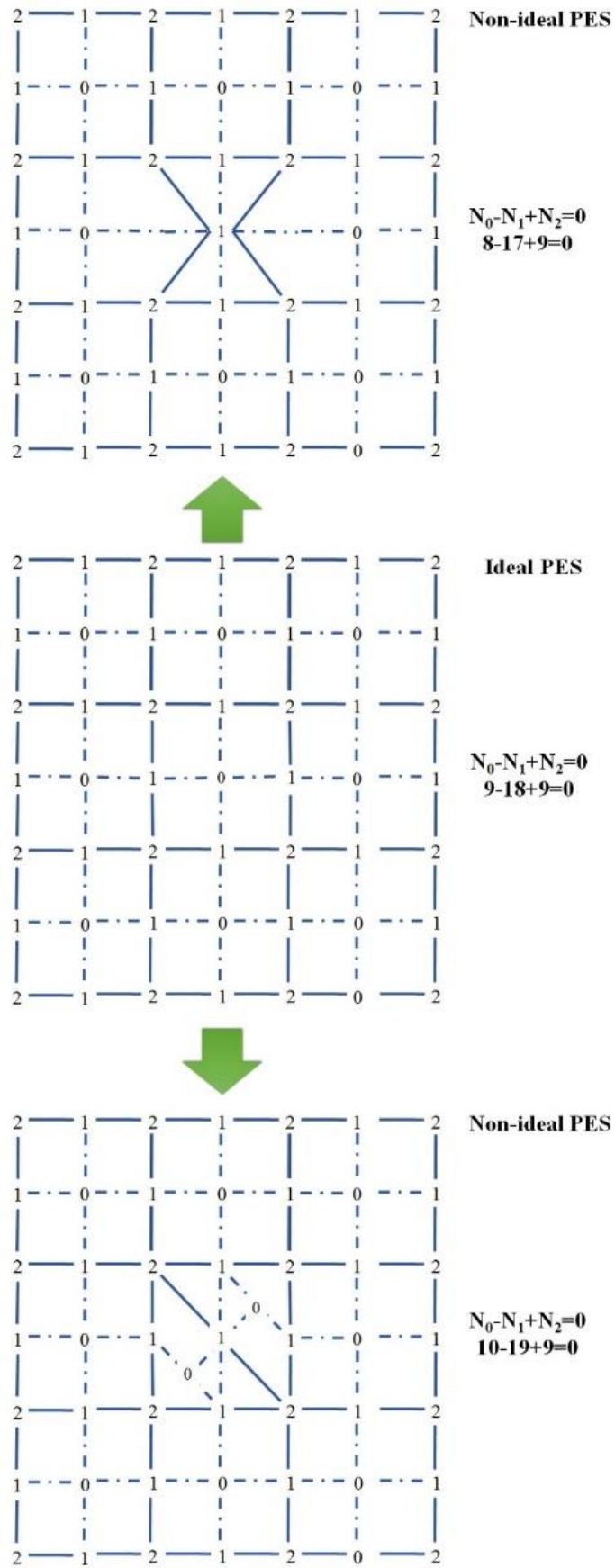
Figure 23 *The difference of Ideal central and two Non-ideal PES (top and bottom)*

1.4.2. Previous attempts of fitting curves, surfaces and hyper-surfaces

Conformational analysis of organic molecules has been initiated before the middle of the 20th century. Simple hydrocarbons like ethane (H_3C-CH_3), propane ($CH_3-CH_2-CH_3$) and n-butane ($CH_3-CH_2-CH_2-CH_3$) were the basic molecules that exhibited such structural characteristics that provided the basis of conformational analysis. Methyl rotations (-C-CH_3) and ethylene rotations (-CH_2-CH_2-) were very similar in a variety of compounds as illustrated in *Figure 24* and *25*[48].

The minima for methyl rotations occurred at -60°, +60° and 180° for ethane, propane and for the anti (a) orientation of n-butane. These are considered ideal values. For the gauche (g+ or g-) orientation in n-butane, the methyl rotation remained in the same vicinity (52.3°) [48] thus it is still regarded as ideal. The two methyl rotations in propane also produced an ideal surface where 3x3=9 minima occurred regularly at -60°, +60° and 180° values of the dihedral angles. In n-butane the dihedral angles associated with the rotation about the central C-C bond were -69.5°, +69.5° and 180°. On the basis of this it would have been easy to assume that this is the general rule for conformational analysis.
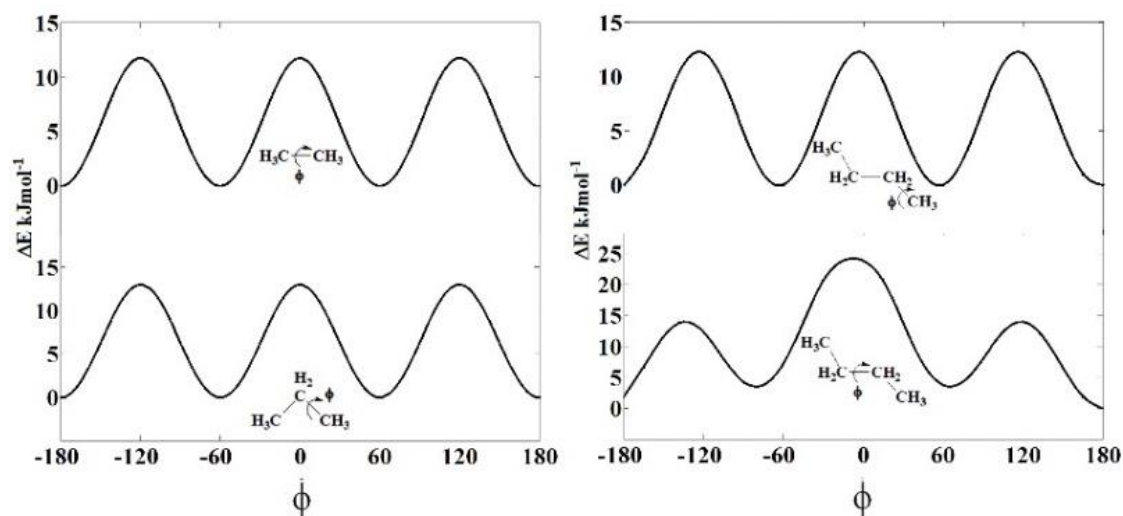
Figure 24 *Modification of conformational potential energy curve by increasing degree of substitution.*

In 1963 Ramachandran[25] attempted to perform conformational analysis on simple peptides and the pattern for such a double rotor –CONH-CH_2-CONH- turned out to be quite different from the propane surface. More recently, it was observed that even as a simple compound as n-pentane behaved in non-ideal fashion[49]. Consequently, it now

appears to be reasonable to assume that surface topology is a function of the rotating moieties.

To establish the topology of the surface, a set of grid points need to be computed. The locations of the minima on the surface represented by the grid points will lead to the topological image of the potential energy surface (PES). Fitting analytic functions to the grid points is a mathematical technique that has already been developed.
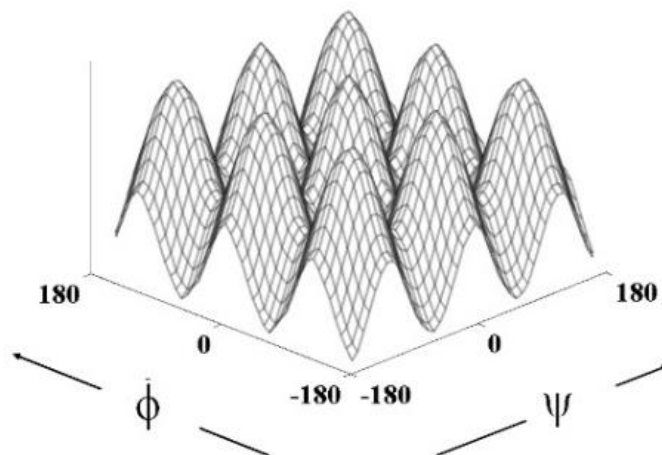
Figure 25 *Conformational potential energy surface (PES) of propane.*

Fitting mathematical functions to potential energy curves was investigated, in the 1970s, by Pople and co-workers[50–53]. Their original driving force appeared to be that such functions might be useful to estimate the rate of conformational changes. Pople and his co-workers attempted to use very simple trigonometric functions to achieve a reasonable accuracy. Later, in 1988, Chung has pointed out that full Fourier expansion is needed[54]. In the meantime Peterson et al.[48] started to use an extended Fourier series in which Gaussian functions were augmenting the trigonometric expansion. Work has also continued using rather large multi-term Fourier expansion[55].

A variety of molecular systems were also fitted by quantitative analytic functions[56–61]. Such surfaces may be used for qualitative visualization. This is usually accomplished by 2D-contour or pseudo 3D plots. These were historically demonstrated in several cases such as $\cdot CH_3$[62], $\cdot CH_2\text{-}S(O)H$[63] and $\cdot CH_2\text{-}S(O_2)H$[64].

2. Purpose and aim

The long term aims for the present dissertation is to find mathematical functions which describe the folding of the peptide residues in order to find eventually the solution to solve the protein folding problem, and to build up the conformational network for the folding.

When proposing such methodologies it is reasonable to start with the description of small compounds, and aim for a bottom up solution.

At first it is necessary to achieve a reasonable accuracy of the fitted functions for the folding of the internal rotation of typical organic functional groups which have only one independent variable.

As the second step it is indispensable to extend the one dimensional mathematical functions to describe the folding of the molecules which have two independent torsional angles and they include peptide bonds.

And last, but not least, the fitted functions, which describe the conformational spaces, can be analyzed to obtain the critical points, minima, transition states and maxima.

Not all possible cases are studied. The present dissertation only outlines a "method" which hopefully shows the direction for further researches.

3. Methodology

"We must emphasize a statement which I am sure you have heard before, but which must be repeated again and again. It is that the sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes the observed phenomena. Furthermore, it must satisfy certain esthetic criteria, that is, in relation to how much it describes, it must be rather simple. Since one cannot tell exactly how 'simple' simple is ... Simplicity is largely a matter of historical background, of previous conditioning, of antecedents, of customary procedures, and it is very much a function of what is explained by it."[65]

John von Neumann

3.1 Chemical background

3.1.1 Investigation of typical organic functional groups with a single rotor

To considering the dihedral torsion or internal rotation of typical organic functional groups, like a methyl group ($H_3C–$). The simplistic view-point is that the rotation is always the same no matter wherever it is located in any molecule.

The first question we wish to investigate is how large such a nearest neighbour interaction is in compounds H_3C-X, where X may be chosen horizontally or vertically along the periodic table:

$H_3C\!-\!CH_2$ $H_3C\!-\!NH_2$ $H_3C\!-\!OH$ $H_3C\!-\!SH$

0 lone pair   1 lone pair   2 lone pairs   2 lone pairs
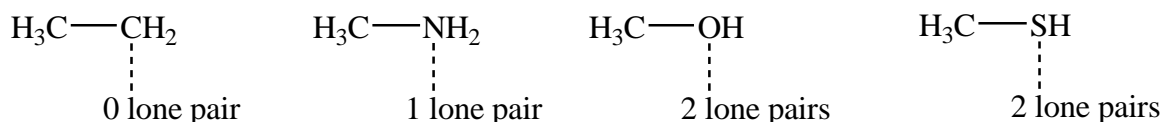
It will explore is if it is possible to quantify the changes in energy for the rotation about homonuclear diatomic units such as (X-X) as the number of lone pairs of X increases, as in the series:

$H_3C\!-\!CH_3$ $H_2N\!-\!NH_2$ $HO\!-\!OH$ $HS\!-\!SH$

0-0 lone pair   1-1 lone pairs   2-2 lone pairs   2-2 lone pairs
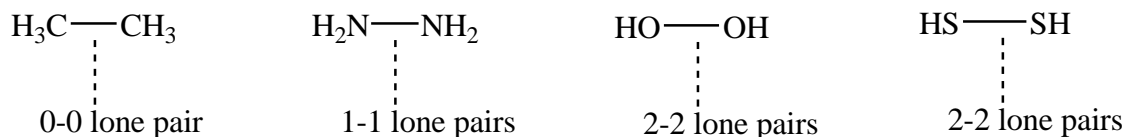
Around a single bond (C-O or O-O) in compounds containing more than one σ-bonds results in variable chemical environments.

$H_2N\!-\!OH$ $H_2N\!-\!SH$ $HO\!-\!SH$

1-2 lone pairs   1-2 lone pairs   2-2 lone pairs
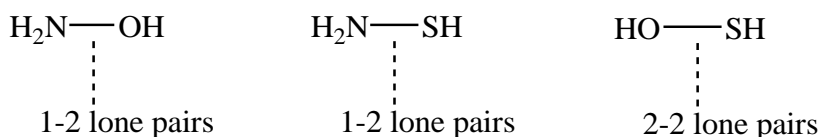
Heteronuclear (X-Y) diatomic units were also studied.

3.1.2 Organic molecules with double rotors

The molecules that were studied are shown in *Figure 26 and 27*. Of the five compounds the first two (I and II) has saturated rotating moieties attached to the central CH_2 skeleton. The second set (III, IV and V) contains amide bonds (-CONH-). Compounds, III, IV and V have special structures: which is actually glycine diamide - CONH-CH_2-CONH-, alanine diamide-CONH-CH-Me-CONH- and valine diamide - CONH-CH-CH-Me_2-CONH- .
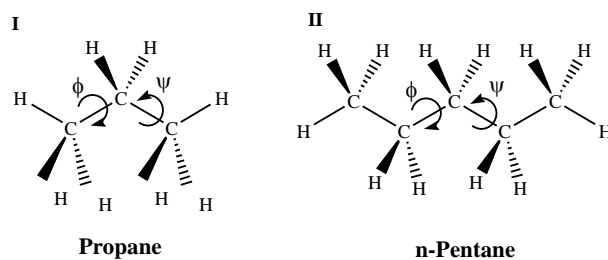
Figure 26 *Molecules with rotating sp³ ligand were studied as double rotors to generate the Potential Energy Surface (PES).*
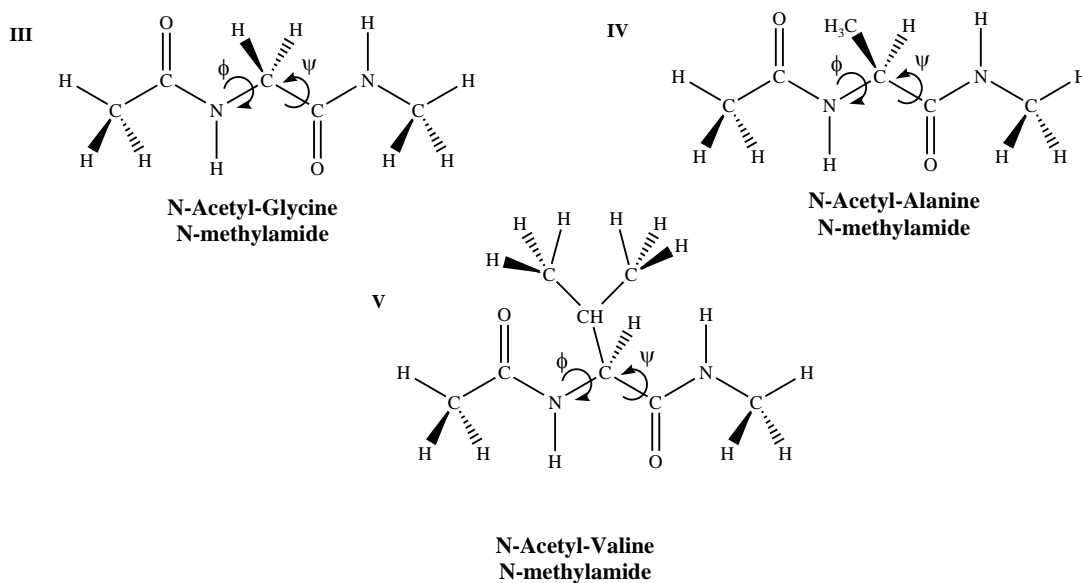
Figure 27 *Molecules with rotating symmetric sp² ligand studied as double rotors to generate the Potential Energy Surface (PES).*

Clearly in the first two compounds (I and II) the rotating moieties connected to the central methylene carbon are of tetrahedral moieties. In contrast to that the last three compounds (III-V) have flat, trigonal planar, rotating moieties with heteroatoms.

These set of five compounds (I-V) represent increasing complexity of the potential energy surface. Of course the increasing complexity in appearance may lead to increasing complexity of the mathematical function to be fitted.

3.2 Mathematical background

3.2.1 Calculation of the grid points

The relative energies, ΔE, associated with internal rotation were calculated by quantum mechanics using the B3LYP/6-31G(d) implementation of the density functional theory in gas phase in the Gaussian09[66,67] software package. Since the Ramachandran type peptide conformational potential energy surface is not only the most important chemically and biochemically, but also a most complicated topologically, an extensive level of theory study was carried out using the alanine-diamide (*Figure 28*). In this study the B3LYP/6-31G(d) level of theory proved to be the best compromise.

1	3-21G	4	6-31G	7	6-31++	10	6-311G(d,p)
2	4-21G	5	6-31G(d)	8	6-31++G(d,p)	11	6-311++G
3	4-31G	6	6-31G(d,p)	9	6-311G	12	6-311++G(d,p)
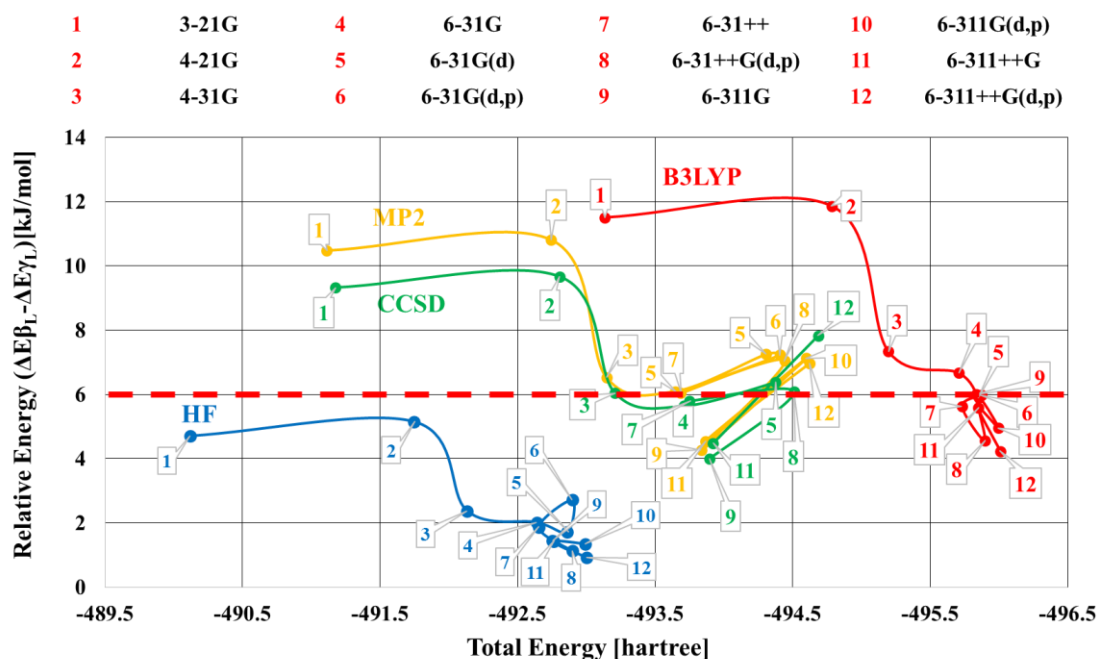
Figure 28 A case study of the basis set size effects using the alanine diamide. The horizontal broken line corresponds to 5.97 kJ/mol, the value obtained at B3LYP/(6-31G(d)) level of theory.

The calculations were carried out for a range of dihedral angles for ϕ and ψ of the molecules in the interval $[-\pi, \pi]$ with grid points at 15° increments in order to generate the curve, surface and hypersurface.

Number of grid points $\left(\dfrac{360}{15}+1\right)^2 = (24+1)^2 = 25^2 = 625$ was calculated for each surface.

3.2.2 Functional analysis

The functional representation of the internal rotations are based on the finite even Fourier series [16] which are periodic functions in $[-\pi, \pi]$ interval[68], are able to represent

the periodicity of the PEC or PES. Fourier-series with the general form is shown in equation [16].

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos nx \right) \qquad [16]$$

where f(x) is a periodic function in [2π] interval and a_0 is a constant term of the series .

The functional analyses were carried out using the Matlab[69] software. A Levenberg-Marquardt algorithm[70–72], a nonlinear least square method with a local minimizer, was performed to fit a Fourier expansion[68]. A nonlinear least square method is defined as an equation that is nonlinear, or a combination of linear and nonlinear in the coefficients[73]. The model is particularly sensitive to the starting points.

The model is given by an equation:

$$y = f(X, \beta) + \varepsilon \qquad [17]$$

where y is a vector of responses, f is a function of β and X, β is a vector of coefficients, X is matrix for the model and ε is a vector of errors [69,71].

Each fitting was made an iterative approach which required the following steps:

1. Estimating the coefficients initially.
2. Creating the fitted curve or surface for the estimated set of coefficients.
3. Adjusting the coefficients and determining for the better fit.
4. Iterate the process (2) until the fit reaches the specified convergence criteria.[69]

3.2.2.1 Background of the Potential Energy Curve fitting

Let ΔE be the energy change associated with rotation of a functional group relative to the lowest energy. ΔE can be expanded in a Fourier-series [18]:

$$\Delta E(\varphi) = a_0 + \sum_{n=1}^{m} \left(a_n \cos n\omega\varphi + b_n \sin n\omega\varphi \right) \qquad [18]$$

where a_0 is a constant, m is the number of terms in the expansion and ω is the fundamental frequency of the signal.

3.2.2.2 Background of the Potential Energy Surface fitting

Of the surfaces used, each surface had a minimal set of functions [19], which includes a summation up to 6, yielding 6x4=24 terms of simplified Fourier expansions[68] with two independent variables.

$$E_a(\varphi,\psi) = a_0 + \sum_{m=1}^{6} (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi) \qquad [19]$$

where a_0 is the constant term in the series, ω is the fundamental frequency of the signal and m is the sequence number of the terms.

For the sake of achieving higher accuracy, a set of Gaussian functions were fitted [20] to the recognisable critical points.

$$E_b(\varphi,\psi) = \sum_{m=1}^{9} A_m e^{\left(-\left(\frac{c_\varphi(\varphi-\varphi_{0m})^2}{2\sigma_{\varphi m}{}^2} + \frac{c_\psi(\psi-\psi_{0m})^2}{2\sigma_{\psi m}{}^2}\right)\right)} \qquad [20]$$

where m is the sequence number of the terms, A is the amplitude, φ_0 and ψ_0 are define the center and the σ_φ and σ_ψ are the φ and ψ extension of the ellipsoid.

For the sake of the transformation of coordinate system an extended two dimensional Fourier-series [21] is needed.

$$E_c(\varphi,\psi) = \sum_{m=1}^{6} f_1 \cos(m\omega\varphi + m\omega\psi) f_2 \cos(m\omega\varphi - m\omega\psi) + f_3 \cos(m\omega\varphi + m\omega\psi f_4 \sin(m\omega\varphi - m\omega\psi) \quad [21]$$
$$+ f_5 \sin(m\omega\varphi + m\omega\psi) f_6 \cos(m\omega\varphi - m\omega\psi) + f_7 \sin(m\omega\varphi + m\omega\psi) f_8 \sin(m\omega\varphi - m\omega\psi)$$

In the summation m is the sequence number of terms in the expansion and ω is the number of the terms.

3.2.3 Determination of critical points

3.2.3.1 First derivative for the determination of the critical points

After fitting the functions to the grids, the next step of the analysis was to generate the partial derivatives. In the case of propane the following functions were obtained:

$$\frac{\partial E_{C3}}{\partial \phi} = \frac{c_1 \cos\left(\frac{p_1 \phi}{q_1}\right)}{d_1} - \frac{c_2 \cos\left(\frac{p_2 \phi}{q_2}\right)}{d_2} - \frac{c_3 \sin\left(\frac{p_2 \phi}{q_2}\right)}{d_3}$$
$$- \frac{c_4 \sin\left(\frac{p_2 \phi}{q_1}\right)}{d_4} + \frac{c_5 \sin\left(\frac{p_1 \phi}{q_1}\right)}{d_5} \qquad [22]$$

$$\frac{\partial E_{C3}}{\partial \psi} = \frac{r_1 \cos\left(\frac{t_1 \psi}{z_1}\right)}{s_1} - \frac{r_2 \cos\left(\frac{t_2 \psi}{z_2}\right)}{s_2} - \frac{r_3 \cos\left(\frac{t_2 \psi}{z_2}\right)}{s_3}$$
$$- \frac{r_4 \sin\left(\frac{t_2 \psi}{z_2}\right)}{s_4} - \frac{r_5 \sin\left(\frac{t_2 \psi}{z_1}\right)}{s_5} + \frac{r_6 \sin\left(\frac{t_1 \psi}{z_1}\right)}{s_6} \qquad [23]$$

From the gradient vectors of propane [24]:

$$\langle g_{C3}| = \left(\frac{\partial E_{C3}}{\partial \phi}, \frac{\partial E_{C3}}{\partial \psi} \right) \qquad [24]$$

the length (L_{C3}) of the gradient vector was calculated according to [25].

$$L_{C3} = \sqrt{\left(\frac{\partial E_{C3}}{\partial \phi} \right)^2 + \left(\frac{\partial E_{C3}}{\partial \psi} \right)^2} \qquad [25]$$

This vector length was taken as dependent variables over ϕ and ψ, and this L_{C3} was plotted as a function of ϕ and ψ as shown in *Figure 29*. At the critical points, the gradient vector must be zero

$$\langle g| = (0,0)$$

Consequently, at these points, the length of the gradient vectors (L) must also vanish.

$$L = 0$$

It is clearly seen from *Figure 29* that this $L_{C3}(\phi, \psi)$ surface does indeed touch zero at the locations of the critical points.
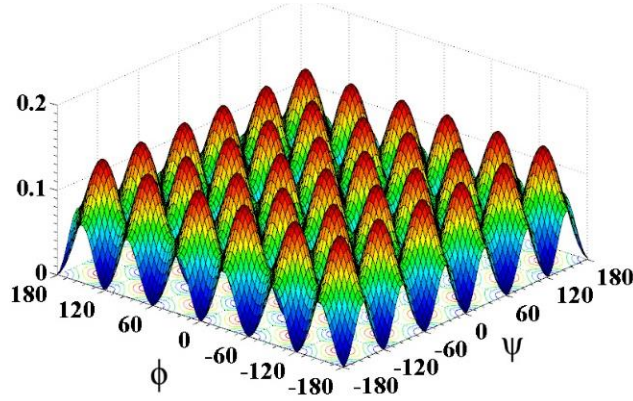
Figure 29 *Length of gradient vectors as functions of dihedral angles L=f (ϕ, ψ) for propane.*

3.2.3.2 Second derivative for the classification of critical points

In the next phase of the procedure it is necessary to calculate the elements of the 2x2 Hessian matrix (**G**) for each of the critical points. This is shown in [26].

$$\mathbf{G} = \begin{vmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{vmatrix} = \begin{vmatrix} \dfrac{\partial^2 E}{\partial^2 \phi^2} & \dfrac{\partial^2 E}{\partial \phi \partial \psi} \\ \dfrac{\partial^2 E}{\partial \psi \partial \phi} & \dfrac{\partial^2 E}{\partial \psi^2} \end{vmatrix} \qquad [26]$$

Since the fitted energy function $E_{C3}(\phi,\psi)$, given in equation [19] respectively, are continuous, and the gradient vector components given in equation [27] and [28] are also continuous functions, the off-diagonal elements of the Hessian **G** are expected to be identical. The following set of equations summarizes the analytic functions of the second derivative for propane (C_3).

$$\frac{\partial^2 E_{C3}}{\partial \phi^2} = \frac{c_1 \cos\left(\frac{p_1\phi}{q_1}\right)}{d_1} - \frac{c_2 \cos\left(\frac{p_2\phi}{q_1}\right)}{d_2} - \frac{c_3 \sin\left(\frac{p_2\phi}{q_2}\right)}{d_3}$$
$$+ \frac{c_4 \sin\left(\frac{p_2\phi}{q_2}\right)}{d_4} - \frac{c_5 \sin\left(\frac{p_1\phi}{q_1}\right)}{d_5}$$

[27]
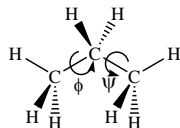
$$\frac{\partial^2 E_{C3}}{\partial \psi^2} = \frac{r_1 \cos\left(\frac{t_1\psi}{z_1}\right)}{s_1} - \frac{r_2 \cos\left(\frac{t_2\psi}{z_1}\right)}{s_2} - \frac{r_3 \cos\left(\frac{t_2\psi}{z_2}\right)}{s_3}$$
$$+ \frac{r_4 \sin\left(\frac{t_2\psi}{z_2}\right)}{s_4} + \frac{r_5 \sin\left(\frac{t_2\psi}{z_1}\right)}{s_5} - \frac{r_6 \sin\left(\frac{t_1\psi}{z_1}\right)}{s_6}$$

[28]

$$\frac{\partial^2 E_{C3}}{\partial \phi \partial \psi} = 0$$

[29]

$$\frac{\partial^2 E_{C3}}{\partial \psi \partial \phi} = 0$$

[30]

Note that the mixed second derivatives [29], [30] vanish for the symmetric trigonometric functions used.

Based on these equations, the elements of the Hessian matrix may be calculated by substituting the actual ϕ, ψ value pairs to the locations of the critical points of the functions. Let us take for example the *anti*, *anti* (a, a) conformation of propane.

$$\mathbf{G}_{C3}(a,a) = \begin{vmatrix} 0.019 & 0 \\ 0 & 0.019 \end{vmatrix}$$

[31]

For a minimum there is no negative diagonal elements, therefore the index (λ) of the critical point is zero ($\lambda=0$). For a saddle point there is one negative diagonal element therefore $\lambda=1$ and for a maximum there are two negative diagonal elements ($\lambda=2$)

For $\lambda=0$
$$\mathbf{G}_{C3} = \begin{vmatrix} A_1^2 & 0 \\ 0 & A_2^2 \end{vmatrix}$$

[32]

For λ=1
$$G_{C3} = \begin{vmatrix} -A_1^2 & 0 \\ 0 & A_2^2 \end{vmatrix}$$
[33]

For λ=2
$$G_{C3} = \begin{vmatrix} -A_1^2 & 0 \\ 0 & -A_2^2 \end{vmatrix}$$
[34]

3.2.4 Statistical assessment of function fitting

3.2.4.1 R-square

The R^2 is the square of the correlation between the response values and the predicted response values [35].

$$R^2 = \frac{SSR}{SST}$$
[35]

where the SSR is the sum of squared regression and the SST is the total sum of squares. *The R^2 can take on any value equal or less than 1. If a value of the R^2 is the closest to 1 then it indicaties a better fit*

3.2.4.2 Root Mean Squared Error

It is an estimate of the standard deviation of the random component in the data.

$$RMSE = \sqrt{MSE}$$
[36]

where MSE is the mean square error.

$$MSE = \frac{SSE}{v}$$
[37]

where SSE is the sum of squares due to error.

If the RMSE values closest to 0 indicating a better fit.

3.3 An overview of the limitation of current computational capabilities

Treating oligopeptides is a problem which is beyond the capability of the present day computer technology to handle. Therefore, at the best what we can do is an estimation how many of the minima of an oligopeptide may be of low enough energy to have any biological significance. The extension of the present method to larger system is not a trivial project. The possibilities of the current methods are limited to determine the global minimum energy structure. For example, if we were to consider a pentapeptide, like the opiorphin (Gln[1]-Arg[2]-Phe[3]-Ser[4]-Arg[5])[74–77], which is an endogenous human pentapeptide, first we need to estimate the size of the problem. We attempted to answer this question using this following procedure.

In this pentapeptide there are 5 pairs of rotors (ϕ and ψ) corresponding to 10 independent variables to the backbone conformation change. As far as the sidechain conformation is concerned there are 3+5+2+2+5=17 dihedral variable, so the PES maybe represented by the following function:

$$E = f \overbrace{(x_1, x_1, ...x_{10},}^{Backbone} \overbrace{x_{11}, x_{12}, ...x_{27})}^{Sidechain} \tag{38}$$

In general we may estimate the minima of these rotational mode, therefore we might expect the following number (N) of the minima for the PEHS.

$$N = \overbrace{3^{10}}^{N_{Backbone}} \times \overbrace{3^{17}}^{N_{Sidechain}} = \overbrace{3^{27}}^{N_{Total}} \tag{39}$$

Characterizing the structure of opiorphin was attempted using a molecular modelling tools. The extended conformer for the wild type was constructed by the *tleap* module of the AmberTools 12 program package[78].

In order to find the global minimum, the sampling of the conformation space was carried out with a simulated annealing (SA)[79] protocol, using the AmberSB99 force field[80,81] and the improved generalized Born implicit solvent model. After the initial geometry optimization the temperature was raised from 300 K to 1000 K for 1000 fs. Then, the system was equilibrated at 1000 K within 4000 fs. Finally, in three steps, the temperature was cooled down first from 1000 K to 500 K within 1000 fs, then from 500 K to 200 K within 2000 fs, and finally from 200 K to 50 K within 7000 fs[79]. A final minimization was also performed in order to find the closest local minimum on the potential energy surface. This protocol was repeated 3000 times. Each conformer was checked after SA procedure and conformers with improper chirality or cis peptide bond were omitted and new calculation was performed in order to obtain the total of 3000 structures which were regarded to have low enough energy for biological significance.

The global minimum and the extended conformers were selected for quantum-mechanical calculations. All geometry optimizations were carried out using the B3LYP functional[82,83] as implemented in the Gaussian package[66], using 6-31G(d) basis set [82] applying a tight convergence criterion. The integral equation formalism of the polarizable continuum model with the radii and non-electrostatic terms of Truhlar and co-workers (SMD)[84] was used to mimic the surrounding aqueous environment. The normal mode analysis was carried out on the structures optimized at the same level of theory, and the calculated harmonic frequencies were scaled by a factor of 0.9608[85]. In order to check the

sensitivity of the geometry to the changing the level of theory, the global minimum structure found by AmberSB99 and the corresponding B3LYP/6-31G(d) optimized geometry were compared.

Beyond the structural characterization of intramolecular hydrogen bonds, the B3LYP/6-31G(d) electron density of the B3LYP optimized structure was also analysed by Bader's Quantum Theory of Atoms in Molecules (AIM) [86] in order to identify the nature of all intramolecular interactions including hydrogen bond and van der Waals interactions.

4. Results and discussion

4.1 Results of curve fitting

A Fourier series [18] was fitted to the ΔE data for the two families of compounds (H_3C-X and X-Y where X, Y = O, N, S). The results are shown in *Table 3* and the plots obtained for the first family are given in *Figure 30.* It appears that the fit is very simple in the case of the H_3C-X family as a single term (n = 1) will present the data very well. We note that the b_1 coefficient is much smaller than the a_1 coefficient, suggesting the PEC is close to being an even function. It is also worth noting that a_1 is very close to ΔE_0. The quality of the fit is estimated by calculating the deviation between the fitted function and the raw data as a function of the dihedral angle. We find the maximum deviation to be smaller than 0.1 kJ/mol.

For the second family (X-Y) two or three terms are required. Again we note from *Table 3* that the "b" coefficients are much smaller than the "a" coefficients. The $\Delta E(\varphi)$ plots for the X-Y family are shown in *Figure 31*. This figure shows, that H_3C-O rotations on *Figure 30* can be represented fairly accurately by single term (n = 1) while the rotation about a O-O bond requires a two term (n = 2) expansion in the Fourier-series.

Table 3 *Fits of the Fourier series [18] for relaxed torsional potential energy curves (PEC), ΔE (kJ/mol). The data are arranged according to the number of terms, n, in the Fourier series.*

	a_0	a_1	b_1	a_2	b_2	a_3	b_3	ω	R^2
C-X family (n=1)									
C-C	5.86	5.99	$-1.07*10^{-1}$	-	-	-	-	0.052	0.9936
C-N	4.97	4.83	$6.42*10^{-1}$	-	-	-	-	0.052	0.9963
C-O	2.97	2.99	$2.50*10^{-1}$	-	-	-	-	0.052	0.9996
C-S	2.72	2.75	$-4.30*10^{-5}$	-	-	-	-	0.052	0.9968
X-Y family (n=2)									
O-O	12.3	17.1	-1.62	8.12	-1.55	-	-	0.0175	0.9994
S-S	15.2	5.65	$-5.80*10^{-1}$	14.2	-2.95	-	-	0.0175	0.9972
S-O	13.5	7.02	$-6.789*10^{-1}$	12.7	-2.48	-	-	0.0175	0.9987
X-Y family (n=3)									
N-O	28.7	3.33	-1.37	-11.5	2.06	-	-	0.0175	0.9994
N-S	19.6	4.31	$-5.44*10^{-1}$	-17.0	$-8.16*10^{-1}$	-1.97	$-3.47*10^{-1}$	0.0175	0.9994
N-N	12.1	12.2	$-2.52*10^{-1}$	11.7	$-4.85*10^{-1}$	4.42	$-2.74*10^{-1}$	0.0175	0.9904

Figure 30 *Fits of [18] with n = 1 to the torsional potential energy curves (PEC), ΔE (kJ/mol), for fully hydrogenated C-X families, (X = C, N, O, S). Data obtained at the B3LYP/6-31G(d) level of theory.*

It appears from *Figure 32* that hydrazine has a very shallow minimum at the anti-conformation[87], which increases the demand for more terms in the Fourier expansion. The observation that the "b" coefficients in [18] are much smaller than the "a" coefficients in most cases along with the closeness to ΔE_0 and to a_1 for $n = 1$ inspired us to adopt a simple even fitting function:

for $\Delta E(\varphi)$ as

$$\Delta E(\varphi) = \frac{\Delta E_{max}}{n+1} + \sum_{m=1}^{n}\left(\frac{\Delta E_{max}}{2^m}\cos\omega\varphi\right) \tag{40}$$

where ΔE_{max} is the highest energy barrier, n is the number of the terms in the series and ω is the fundamental frequency which is depend on the periodicity. For example, the H_3C-X family can be represented by an expansion with one term as:

$$\Delta E = \frac{\Delta E_{max}}{2} + \frac{\Delta E_{max}}{2}\cos(0.052\varphi) \tag{41}$$

For the X-Y family -excluding the N-N and N-S members- the function is:

$$\Delta E = \frac{\Delta E_{max}}{3} + \frac{\Delta E_{max}}{2}\cos(0.0175\varphi) + \frac{\Delta E_{max}}{2^2}\cos(0.017\cdot 2\varphi) \tag{42}$$

In the case of N-N and N-S we obtain the function:

$$\Delta E = \frac{\Delta E_{max}}{4} + \frac{\Delta E_{max}}{2}\cos(0.0175\varphi) + \frac{\Delta E_{max}}{2^2}\cos(0.017\cdot 2\varphi) + \frac{\Delta E_{max}}{2^3}\cos(0.017\cdot 3\varphi) \tag{43}$$

It appears from *Table 3* that not all σ-bonded functional groups behave in an identical way during internal rotation (torsion). Some of them like H_3C-X exhibit rather high symmetry with three identical transition states[54]. These can be represented with a single ($n = 1$) cosine function. Rotation about other σ-bonds, where the three TS are not identical or where only a pair of non-identical TS is in the potential, require a longer Fourier expansion ($n = 2$ or 3). All of the potentials however were fitted very accurately with R^2 values ranging from 0.9904 to 1 (*Table 3*).
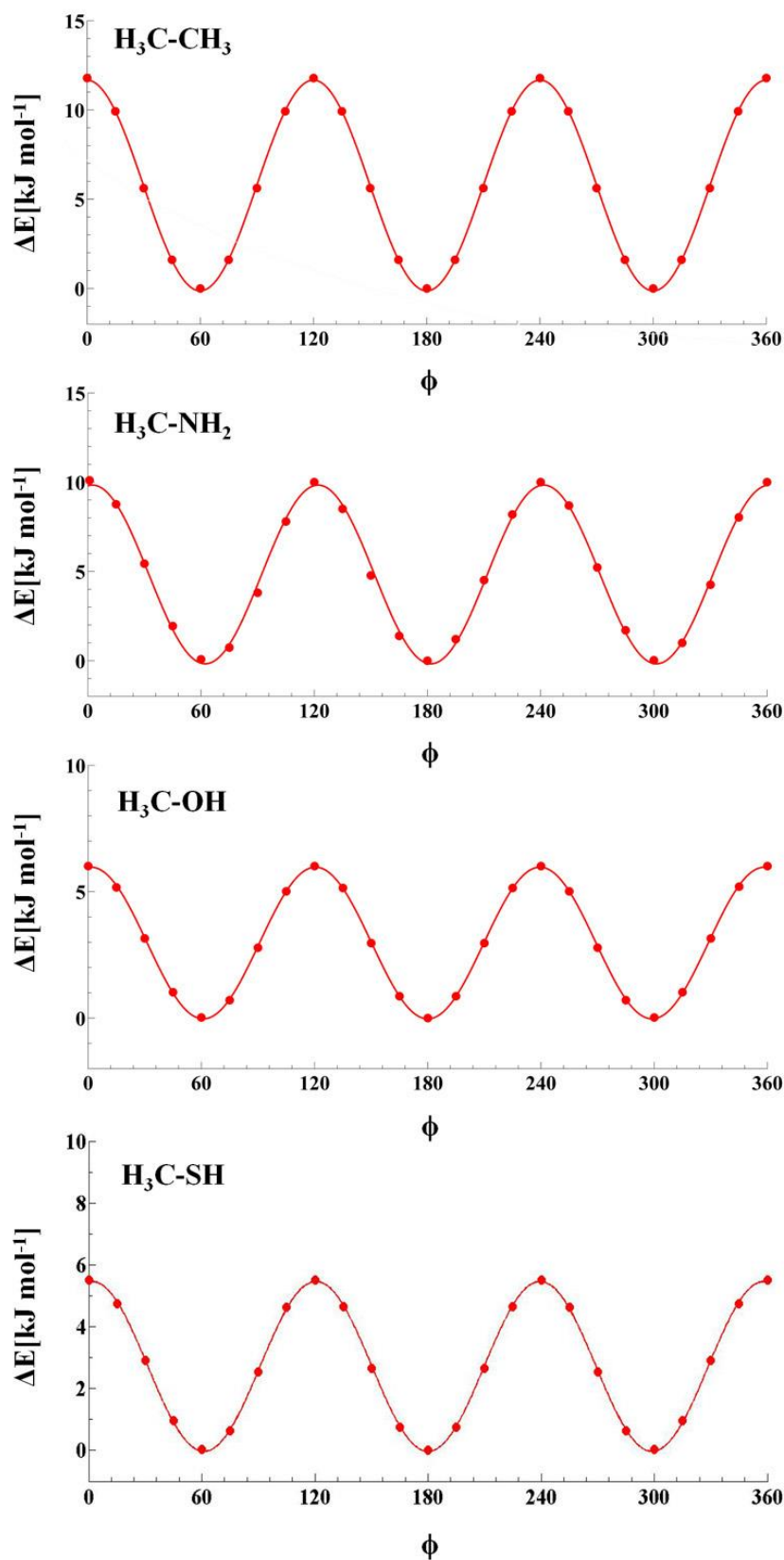
Figure 31 *Fits of [18] with n = 2 to the torsional potential energy curves (PEC), ΔE (kJ/mol), for fully hydrogenated X-Y families, (X, Y = N, O, S). Data obtained at the B3LYP/6-31G(d) level of theory.*
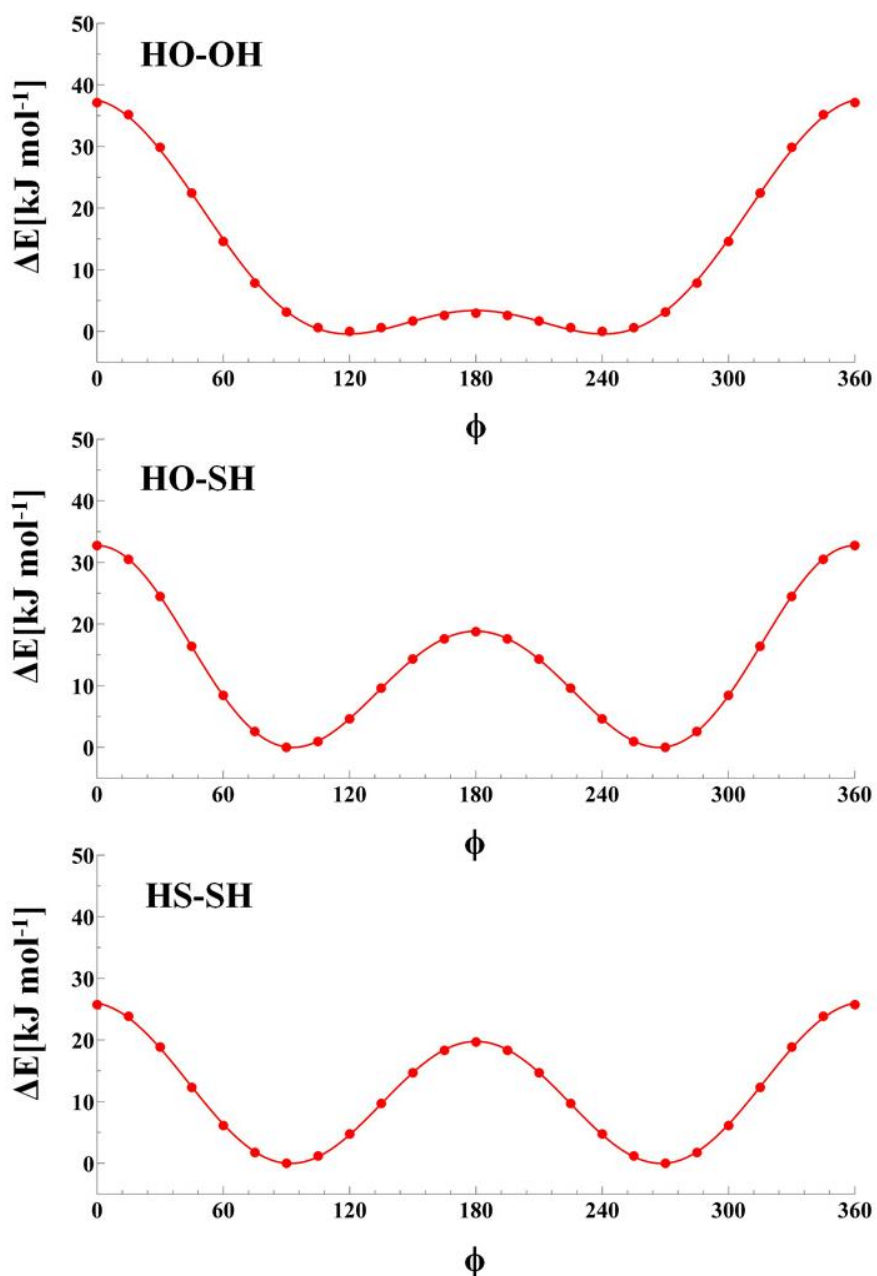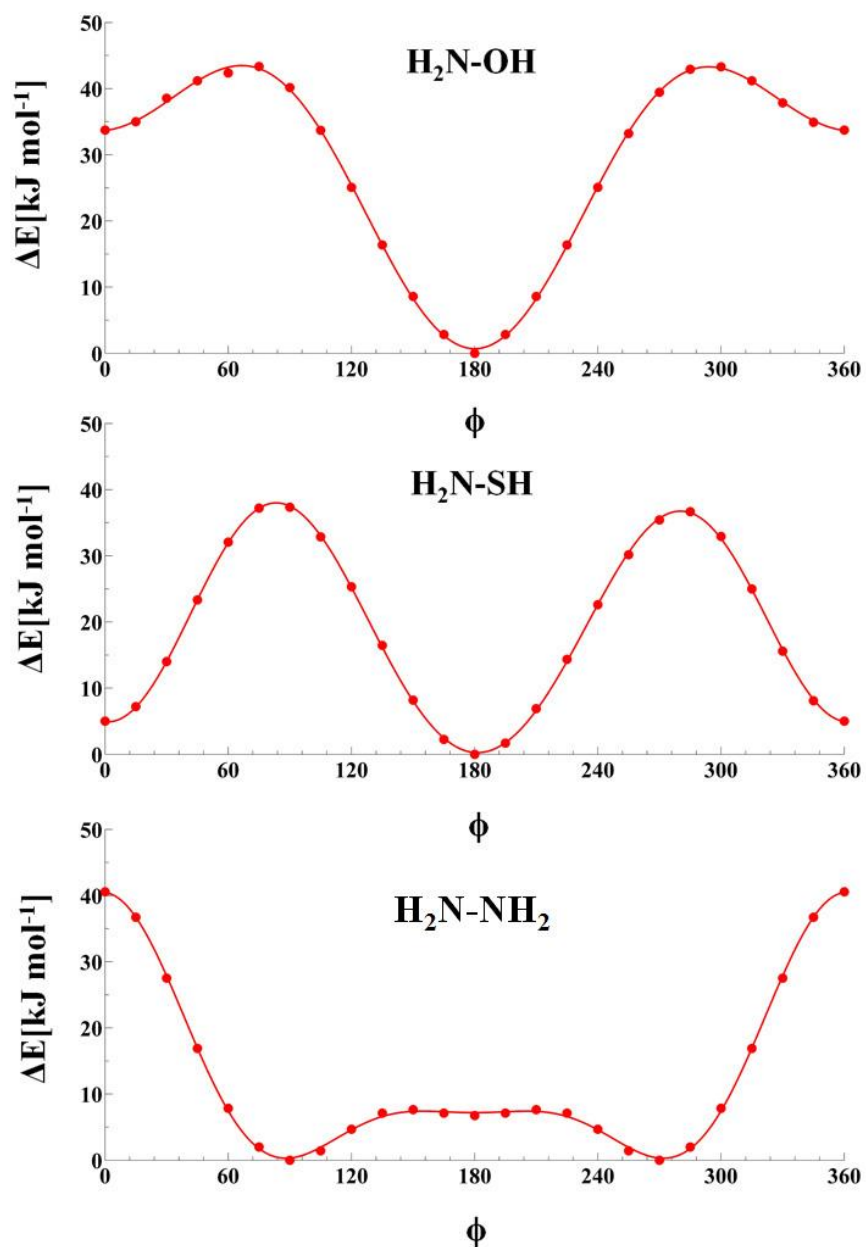
Figure 32 *Fits of [18] with n = 3 to the torsional potential energy curves (PEC), ΔE (kJ/mol), for fully hydrogenated X-Y families, (X, Y = N, O, S). Data obtained at the B3LYP/6-31G(d) level of theory.*

Table 4 *Comparison of the maxima (or transition states) obtained by fitting [41], [42] and [43] to the ΔE data (kJ/mol) from DFT [B3LYP/6-31G(d)] calculations to the force field data for fully hydrogenated C-X and X-Y families, (X,Y = O, N, S). Data for N-acetyl Alanine methylamide are also presented.*

Method	A: Higher TS		B: Lower TS	
	DFT	Force field	DFT	Force field
C-X family (n=1)				
C-C	11.78	12.05	-	-
C-N	10.11	9.35	-	-
C-O	6.01	4.82	-	-
C-S	5.51	5.17	-	-
H₃C-Ala	15.62	14.91	-	-
X-Y family (n=2)				
S-O	47.86	13.96	18.77	11.42
N-O	43.29	41.86	42.93	37.67
S-S	34.42	53.02	23.08	47.86
O-O	37.12	35.7	2.92	4.41
X-Y family (n=3)				
N-S	36.66	22.74	36.66	21.81
N-N	40.56	14.28	7.64	5.30

The TS obtained by molecular mechanics (MM) simulations agreed reasonably well (*Table 4*) for the C-X family. In the case of the X-Y family, noticeable deviations were observed in the cases of the S-O, S-S, N-O and N-N linkages. These differences amounted to 33.9 kJ/mol, -18.6 kJ/mol, 13.9 kJ/mol and 26.3 kJ/mol, respectively. This is understandable because traditional force field software were written to simulate protein structure and protein folding. Thus discrepancies may be expected in cases of bonds that are not common in proteins. Since MM simulations, in general, aim to reproduce geometries and relative stabilities of minima, the achieved TS energy approximation are surprisingly good.

4.2 Results of surface fitting

Figures 33 and 34 show the conformational PES for compounds I-V. The optimized parameters of the five molecules and the accuracy of the fitted surface are listed in *Table 5* and *6*. The parameters of the fitted functions are listed in in *Table S1* and *Table S2* (Supplementary Materials).
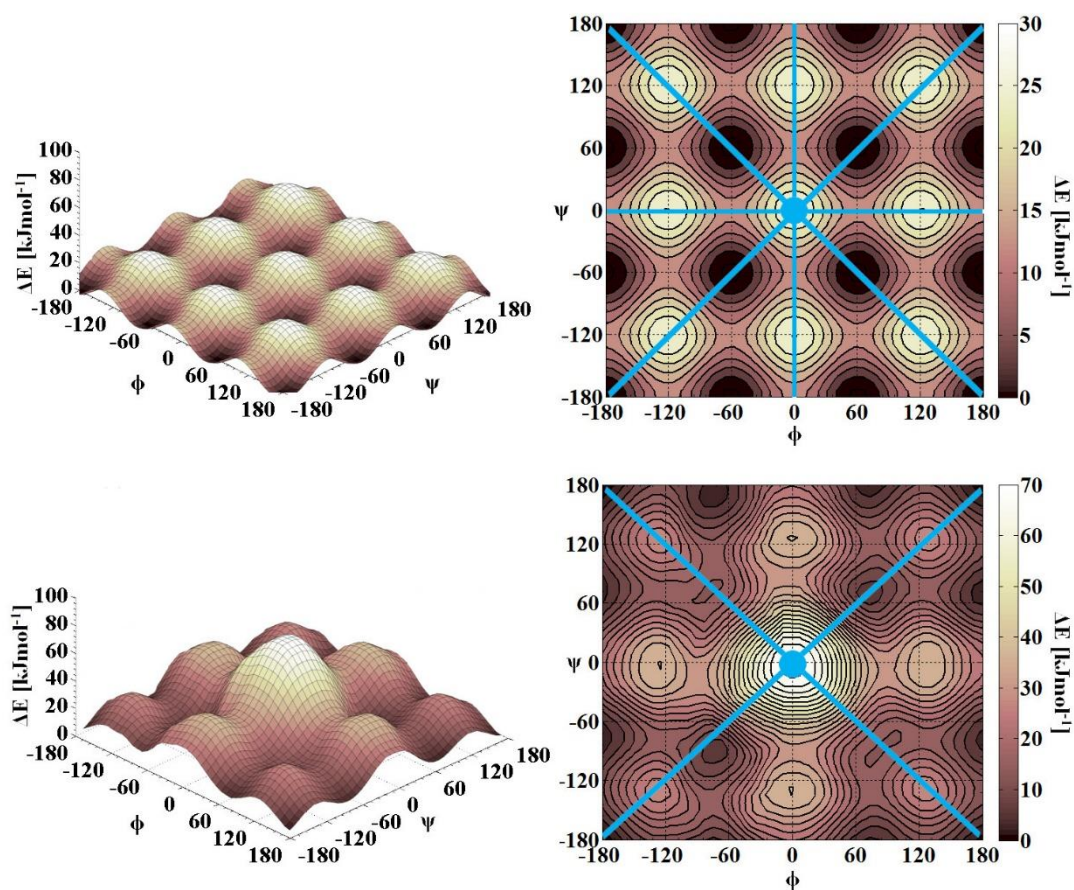
Figure 33 *Potential energy surfaces for the compounds I and II in Figure 22 calculated for the gas phase at the level B3LYP/6-31G(d)and the symmetry axis.*
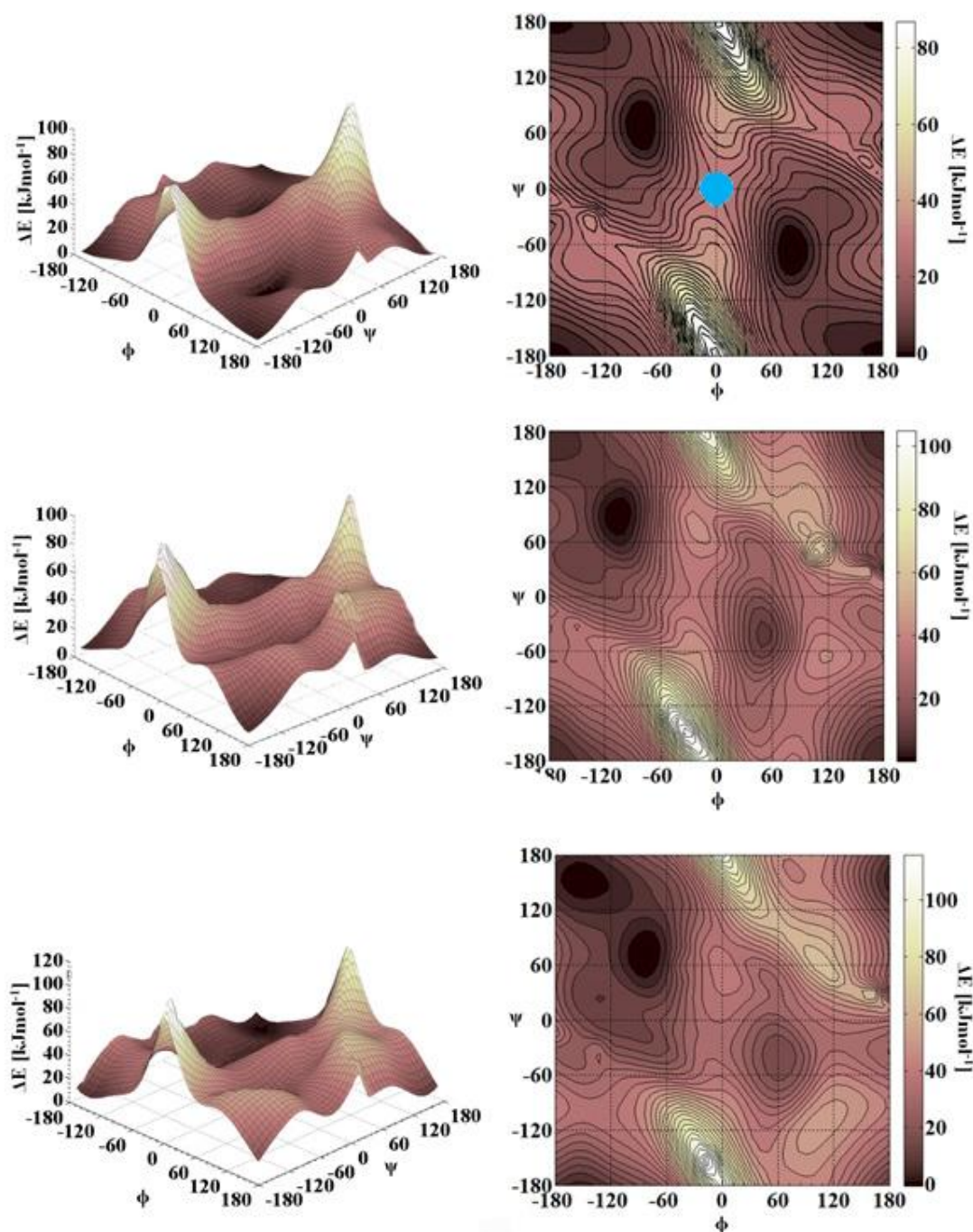
Figure 34 *Fully relaxed potential energy surfaces for the compounds III, IV and V in Figure 23 calculated for the gas phase at the level B3LYP/6-31G(d).*

The inhomogeneity of the molecules main chain causes a reduction in the number of the symmetry axes and the surface becomes less-symmetric, consequently, the function which describes that surface is increasingly more complicated. Consequently, the R^2 values are gradually reduced from unity, the differences between calculated and RMSE of the fitted functions are also reduced in a gradual fashion.

Table 5 indicates that the PES which are highly symmetrical in the case of propane (I) and n-pentane (II) can be fitted by a simple two dimensional Fourier expansion [19]. In contrast the PESs of the molecules that have strong intramolecular interactions can only be described with functions that are more complicated. The number of the Fourier expansions depends on the number of the higher maxima on the surface.

This simple two dimensional Fourier expansion usable for propane [44] and n-pentane [45]:

$$E(\varphi,\psi) = a_0 + (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi) \qquad [44]$$

$$E(\varphi,\psi) = a_0 + \sum_{m=1}^{4} (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi) \qquad [45]$$

The last three molecules studied (III-V), contains a peptide bond and has a dissymmetrical structure, which is described by a surface even though there is a symmetry centre only for the glycine diamde. In contrast to that, the chiral alanine diamide (IV) and the valine diamide (V) have no symmetry what-so-ever. This increasing complexity in the structure leads to increasing complexity of the fitted mathematical functions. The surface of glycine diamide, the alanine diamide and the valine diamide needs according to equation [46] 9 terms of Gaussian functions [20], 24 terms of 2 dimensional Fourier expansions [19] and also 24 terms of extended Fourier functions [21]:

$$E(\varphi,\psi) = a_0 + \sum_{m=1}^{9} A_m e^{\left(-\left(\frac{c_\varphi(\varphi-\varphi_{0m})^2}{2\sigma_{\varphi m}^2} + \frac{c_\psi(\psi-\psi_{0m})^2}{2\sigma_{\psi m}^2}\right)\right)}$$

$$+ \sum_{m=1}^{6} (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi)$$

$$+ \sum_{m=1}^{6} f_1 \cos(m\omega\varphi + m\omega\psi) f_2 \cos(m\omega\varphi - m\omega\psi) + f_3 \cos(m\omega\varphi + m\omega\psi \, f_4 \sin(m\omega\varphi - m\omega\psi)$$

$$+ f_5 \sin(m\omega\varphi + m\omega\psi) f_6 \cos(m\omega\varphi - m\omega\psi) + f_7 \sin(m\omega\varphi + m\omega\psi) f_8 \sin(m\omega\varphi - m\omega\psi)$$

$$[46]$$

The molecules were originally described by energy values defined by the two rotated dihedral angles.

Creating a function describing the conformational space of the species drastically reduced the amount of data, and more importantly, made it possible to perform a more involved mathematical analysis on the functions than it is ever doable on a matrix of data grid points.

While the Fourier expansion defines mostly the periodicity of the surface, the Gaussian function defines the minima and the maxima and the extended Fourier series

defines the surface dissymmetry. It seems the complexity of the molecule and the number of the symmetry axes (*Figure 33* and *34*) specifies the complexity and the accuracy of the fitted function. *Table S3* illustrates this point in the cases of propane and n-pentane. The number of functions needed is mostly determined by the complexity of the molecule. Although the fitting becomes more appropriate by applying more functions, there is a limit; fitted surfaces for simple molecules' are only worsened if an excessive number of functions are used.

At this level of theory, peptide minima appear on the surface. In the case of glycine diamide (*Figure 35*) five conformers (β_L, γ_L, γ_D, δ_L δ_D), and in the case of alanine diamide (*Figure 36*) six conformers (α_D, β_L, γ_L, γ_D, δ_L δ_D) instead of the ideal nine conformers shown in *Figure 12B*.

In the case of the valine diamide twenty three conformers (*Table 6*), instead of the ideal twenty seven, were located. Four of the minima, α_L (g-), α_L (g+), δ_L (a) and the ε_L (a), couldn't be located.

Table 5 . *Summary of the minimum energy critical points of the I-IV compounds and accuracy of surface fit.*

	Optimized results				Fitted results							$\Delta E_{opt}-\Delta E_{fit}$
Compound	Unique minima	ϕ	ψ	ΔE (kJmol⁻¹)	ϕ	ψ	ΔE (kJmol⁻¹)	R^2	RMSE	Fitted function	Number of fitted parameters	
I	9	60.36	60.72	0.00	60.36	60.72	0.00	0.998	0.344	(1)	25	0
II	4	64.69	89.66	14.15	67.28	89.90	9.90	0.997	0.992	(1)+(2)	70	4.25
	2	-68.28	-67.90	7.07	68.28	67.90	5.79					1.28
	4	-65.40	180.00	3.62	-65.40	180.00	2.51					1.11
	1	179.23	179.23	0.00	179.23	179.23	-0.63					0.63
III	δ_L, δ_D	-122.78	21.93	10.24	-120.20	19.39	12.65	0.929	4.631	(1)+(2)+(3)	118	-2.41
	β_L	179.99	180.00	2.58	178.8	178.8	2.66					-0.08
	γ_L, γ_D	82.15	-68.54	0.00	81.43	-50.41	2.28					-2.28
IV	δ_D	-165.06	-30.94	29.28	-159.3	-28.79	24.50	0.911	5.725	(1)+(2)+(3)	118	4.78
	α_D	67.92	26.68	24.21	67.17	21.11	28.07					-3.86
	δ_L	-126.55	21.42	13.04	-117.10	19.10	14.71					-1.67
	γ_D	73.52	-57.39	10.88	67.33	-67.17	10.05					0.83
	β_L	-158.33	163.41	5.97	-155.50	167.00	6.06					-0.09
	γ_L	-82.87	72.60	0.00	-86.36	78.89	2.23					-2.23

Table 6 *Summary of the minimum energy critical points of the valine diamide (V) and accuracy of surface fit.*

|		Optimized results				Fitted results							ΔE_{opt}- ΔE_{fit}
Compound	Unique minima	ϕ	ψ	χ	ΔE (kJmol^{-1})	ϕ	ψ	ΔE (kJmol^{-1})	R^2	RMSE	Fitted function	Number of fitted parameters	
V	α_D (a)	62.93	39.08	-164.98	29.08	60.00	38.18	31.19	0.913	6.364	(1)+(2)+(3)	118	-2.11
	α_D (g-)	50.87	39.25	-68.52	29.88	52.87	38.18	31.19					-1.31
	α_D (g+)	50.91	42.45	69.38	34.46	52.73	41.82	35.01					-0.55
	α_L (a)	-86.81	-22.80	173.73	22.46	-85.45	-23.64	16.98					5.48
	β_L (a)	-126.58	132.80	-175.43	7.31	-125.50	136.40	3.95					3.36
	β_L (g-)	-132.08	159.92	-60.83	5.11	-132.70	154.50	0.24					4.87
	β_L (g+)	-152.50	153.74	71.60	5.68	-150.90	150.90	1.45					4.23
	δ_D (a)	-128.55	-65.43	178.76	31.50	-128.50	-63.64	28.42					3.08
	δ_D (g-)	-131.95	-40.53	-67.14	37.58	-132.70	-41.82	32.61					4.97
	δ_D (g+)	-160.47	-36.22	64.15	29.76	-160-50	-34.55	23.91					5.85
	δ_L (g-)	-122.94	18.74	-66.09	10.99	-121.80	20.00	14.92					-3.93
	δ_L (g+)	-130.10	21.79	71.94	16.24	-129.10	20.00	16.91					-0.67
	ϵ_D (a)	77.03	142.69	-177.34	40.48	78.18	143.60	43.02					-2.54
	ϵ_D (g-)	68.84	174.90	-41.88	44.92	67.27	172.70	47.90					-2.98
	ϵ_D (g+)	74.35	164.18	102.17	47.55	74.55	165.50	49.77					-2.22
	ϵ_L (g-)	-122.84	18.70	-66.10	10.99	121.80	20.00	13.14					-2.15
	ϵ_L (g+)	-130.11	21.79	71.94	16.24	-129.10	20.00	16.91					-0.67
	γ_D (a)	73.51	-60.81	173.19	9.10	70.91	-60.64	14.18					-5.08
	γ_D (g-)	59.54	-33.73	-65.04	19.24	60.00	-34.55	17.28					1.96
	γ_D (g+)	62.65	-38.66	72.09	20.76	63.64	-35.44	19.81					0.95
	γ_L (a)	-83.49	78.96	-175.19	0.00	-78.18	81.82	3.28					-3.28
	γ_L (g-)	-83.88	65.26	-71.82	1.85	-85.45	67.27	5.81					-3.96
	γ_L (g+)	-84.28	72.33	60.61	3.74	-85.45	74.55	8.15					-4.41
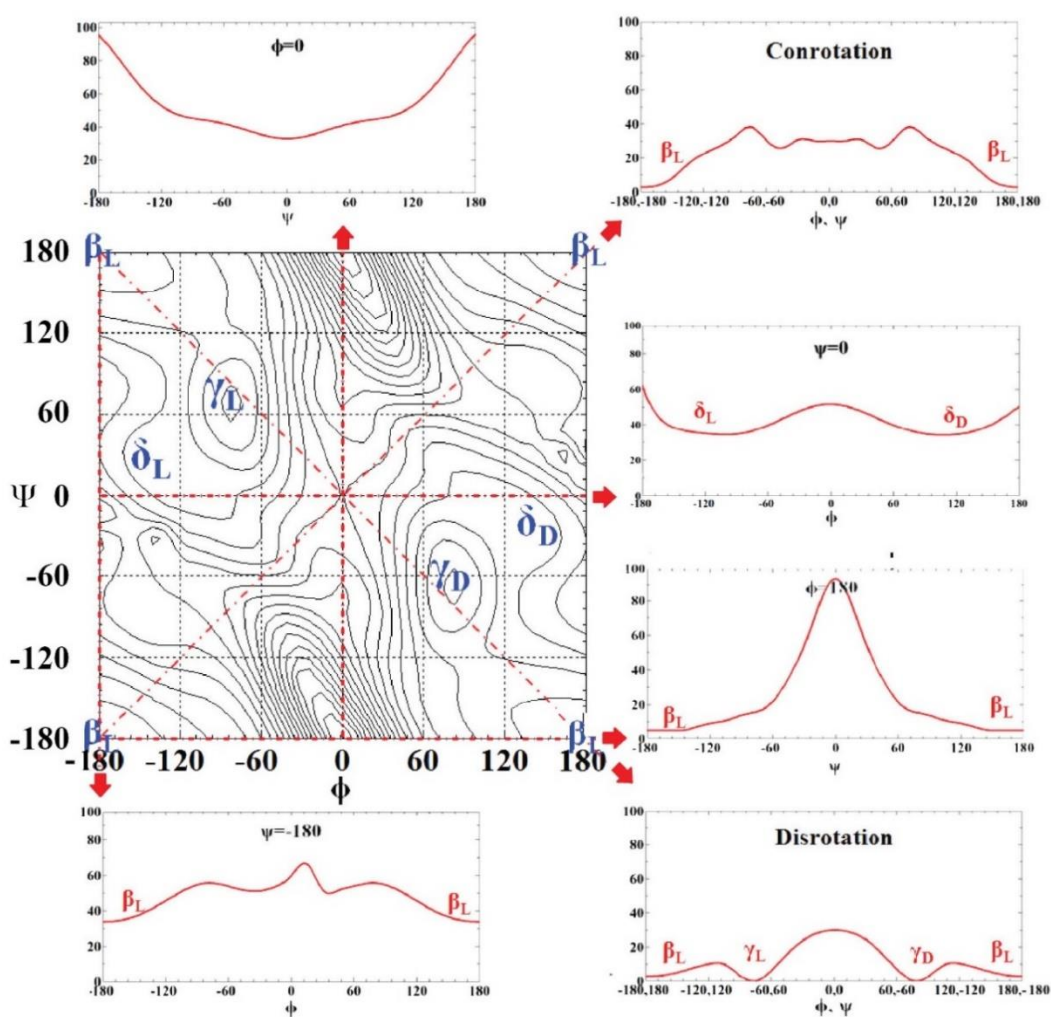
Figure 35 *Topology of Ramachandran map and crossections N-Acetyl-Glycine N-methylamide. Data are calculated for the gas phase at the level B3LYP/6-31G(d).*

The comparison of the fitted surfaces suggested that for a given number of grid points, such as $25^2=625$ simple hydrocarbons can be fitted more accurately than peptides. However, the Ramachandran surfaces of glycine, alanine and valine diamide can be fitted with acceptable R^2 values ($0.91 \leq R^2 \geq 0.93$). These 625 grid points can be represented by mathematical functions containing less than 120 parameters. It is hoped that this may lead to successful big data reduction needed to understand protein folding.
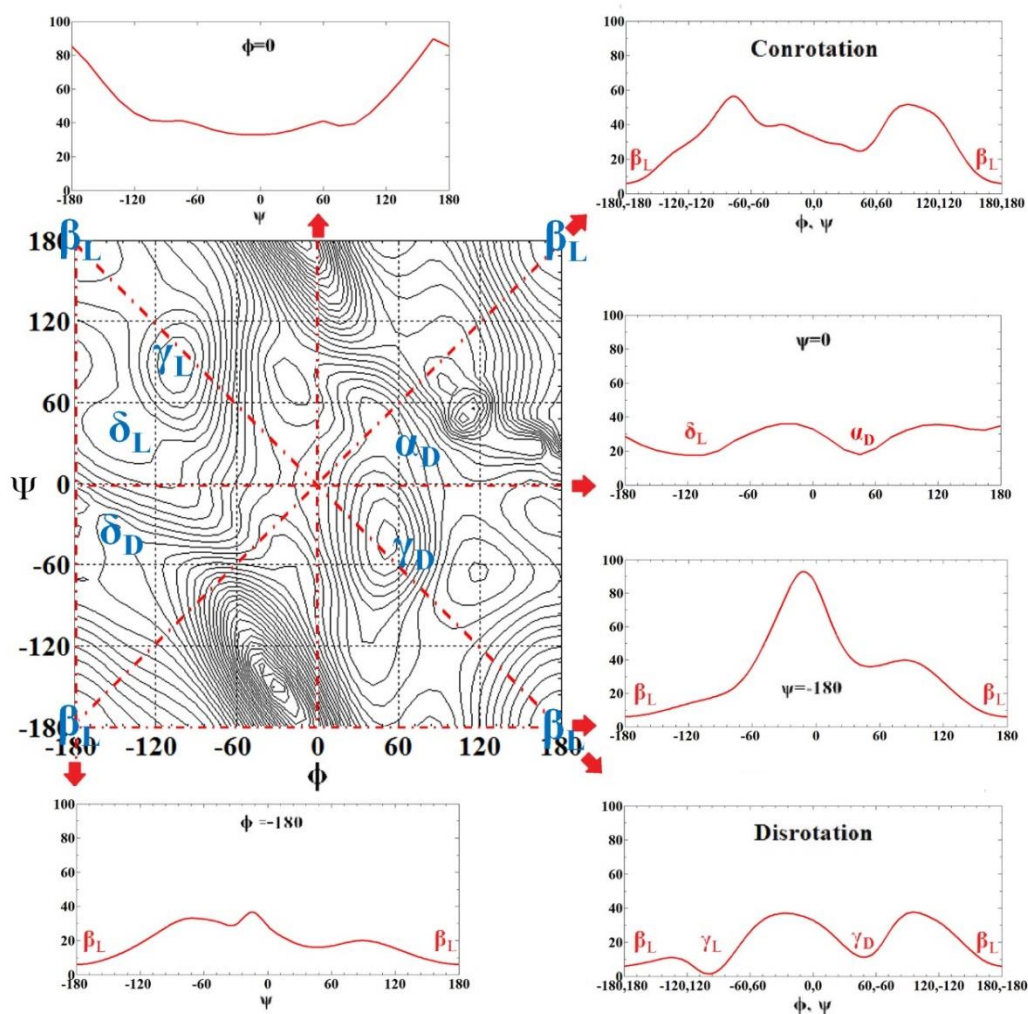
Figure 36 *Topology of Ramachandran map and crossections of N-Acetyl-Alanine-N-methylamide. Data are calculated for the gas phase at the level B3LYP/6-31G(d).*

4.3 Results of critical points determination

The elements for the **G** matrix are summarized in *Table 7* for propane. These results represent the 625 grid point containing PES. The elements of the **G** matrix are summarized in *Table 8* for n-pentane, *Table 9* for glycine diamide and *Table 10* for alanine diamide. In the case of propane, due to the high symmetry of the conformational PES, all critical points are repeated. The possible ϕ and ψ values for minima, TS and maxima are given in [47] and in *Figure 37*.

$$
\begin{array}{lllllllllllllll}
 & \phi & & & & & & & \psi \\
\lambda = 0 & -180° & & -60° & 60° & & 180° & -180° & & -60° & 60° & & 180° \\
\lambda = 1 & -180° & -120° & -60° & 0° & 60° & 120° & 180° & -180° & -120° & -60° & 0° & 60° & 120° & 180° \\
\lambda = 2 & & -120° & & 0° & & 120° & & & -120° & & 0° & & 120° \\
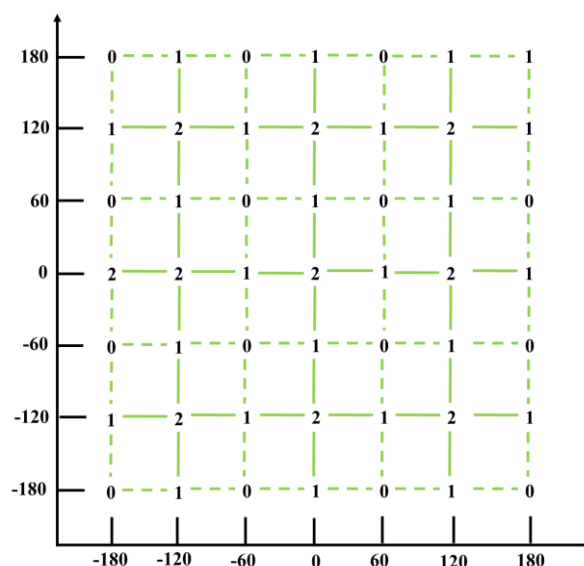\end{array}
$$

[47]

Figure 37 *All critical points in high symmetry conformational PES of propane*

Table 7 *Unique critical points on the conformational PES of propane obtained by fitted analytic function and from first principle direct geometry optimization. Repetitions are are occurring at every 120°.*

Optimized Results			Fitted Results					Index
Conformation		Energy	Conformation		Energy	G		
ϕ	ψ	ΔE [kJ/ mol]	ϕ	ψ	ΔE [kJ/ mol]	G_{11}	G_{22}	λ
60.00	60.00	0.00	60.36	60.72	0.46	0.019	0.019	0
0.00	60.00	12.98	0.18	60.36	13.31	-0.019	0.019	1
0.00	0.00	27.91	0.18	0.18	27.39	-0.019	-0.019	2

Table 8 *Critical points of the conformational PES of n-pentane obtained by fitted analytic function and from first principle direct geometry optimization.*

Optimized Results			Fitted Results					
Conformation		Energy	Conformation		Energy	G		Index
ϕ	ψ	ΔE [kJ/ mol]	ϕ	ψ	ΔE [kJ/ mol]	G_{11}	G_{22}	λ
Minima								
-64.69	89.66	14.15	-67.92	67.92	11.67	0.015	0.017	0
-89.66	64.69	14.15						0
89.66	-64.69	14.15	67.92	-67.92	11.67	0.017	0.015	0
64.69	-89.66	14.15						0
63.69	63.68	7.07	68.28	67.90	7.35	0.015	0.017	0
-63.69	-63.68	7.07	-67.90	-68.28	7.35	0.017	0.015	0
-65.51	177.16	3.62	-65.40	180.00	2.61	0.016	0.013	0
177.16	-65.51	3.62	180.00	-65.40	2.61	0.013	0.016	0
65.51	177.17	3.62	65.40	180.00	2.61	0.016	0.013	0
177.16	65.51	3.62	180.00	65.40	2.61	0.013	0.016	0
179.99	179.99	0.00	179.23	179.23	-0.61	0.018	0.013	0
Transition States								
-118.80	179.17	13.36	-119.81	179.63	14.09	-0.013	0.018	1
118.80	-179.17	13.36	119.81	-179.63	14.09	-0.013	0.018	1
76.47	-76.57	14.59	-	-	-	-	-	1
-76.57	76.47	14.59	-	-	-	-	-	1
119.75	66.89	16.31	118.01	63.06	17.45	-0.013	0.018	1
66.89	119.75	16.31	63.06	118.01	17.45	0.018	-0.013	1
-66.89	-119.75	16.31	-63.06	-118.01	17.45	0.015	-0.013	1
-119.75	-66.89	16.31	-118.01	-63.06	17.45	-0.013	0.015	1
-117.49	69.13	16.27	-119.45	69.00	18.77	-0.013	0.014	1
117.49	-69.13	16.27	119.45	-69.00	18.77	-0.017	0.014	1
-69.13	117.49	16.27	-69.00	119.45	18.77	0.016	-0.017	1
69.13	-117.49	16.27	69.00	-119.45	18.77	0.014	-0.017	1
0.03	179.99	23.81	0.18	179.63	23.93	-0.022	0.013	1
0.03	-179.99	23.81	0.18	-179.63	23.93	-0.022	0.013	1
179.99	0.03	23.81	179.63	0.18	23.93	0.018	-0.013	1
-179.99	0.03	23.81	-179.63	0.18	23.93	0.019	-0.013	1
-4.87	81.10	31.73	-4.50	81.98	33.49	-0.023	0.008	1
4.87	-81.10	31.73	4.50	-81.98	33.49	-0.023	0.008	1
81.10	-4.87	31.73	81.98	-4.50	33.49	0.004	-0.026	1
-81.10	4.87	31.73	-81.98	4.50	33.49	0.013	-0.024	1

Maxima								
120.14	120.14	29.91	120.18	119.81	28.84	-0.017	-0.012	2
-120.14	-120.14	29.91	-119.81	-120.18	28.84	-0.012	-0.017	2
118.58	-118.58	29.94	120.81	-119.81	28.84	-0.017	-0.017	2
-118.58	118.58	29.94	-119.81	120.81	28.84	-0.013	-0.012	2
-120.37	-0.91	40.09	-120.18	0.18	39.42	-0.013	-0.026	2
120.37	-0.91	40.09	120.18	0.18	39.42	-0.017	-0.026	2
0.91	-120.37	40.09	0.18	-120.18	39.42	-0.022	-0.017	2
-0.91	120.37	40.09	0.18	120.18	39.42	-0.022	-0.012	2
0.00	0.00	70.65	0.18	0.18	70.84	-0.022	-0.026	2

It is clear from *Table 8* that the generated PES does show the split up of two minima into two pairs of minima and TS in between, as shown in details in [48a] and [48b] and *Figure 39*.

$$
\begin{array}{cccc}
& \phi,\psi & \phi & \psi \\
& & -90.06 & +64.21 \ (\lambda=0) \\
Expected\ (\lambda=0)\,-60,+60 & & -76.46 & +76.46 \ (\lambda=1) \ observed \\
& & -64.21 & +90.06 \ (\lambda=0)
\end{array}
\qquad [48a]
$$

$$
\begin{array}{cccc}
& \phi,\psi & \phi & \psi \\
& & +90.06 & -64.21 \ (\lambda=0) \\
Expected\ (\lambda=0)\,-60,+60 & & +76.46 & -76.46 \ (\lambda=1) \ observed \\
& & +64.21 & -90.06 \ (\lambda=0)
\end{array}
\qquad [48b]
$$

However, the fitted function shows only the expected two minima not the observed two pairs of minima. These critical points have been optimized, thus they indeed are located on the conformational PES. In the case of n-pentane since the observed TS (λ=1) between the two degenerate minima is less than 0.4 kJmol^{-1} (i.e. 14.58-14.20=0.38 kJmol^{-1}) therefore the fitted function with its R^2=0.9970 and RMSE=0.992 kJmol^{-1} is less accurate than the Gaussian computation.

In order to make an attempt to improve the accuracy of the fitted function, the following two steps have been taken. First of all the PES was recomputed with grid points of 5° increment. This lead to a grand total of $[(360/5)+1]^2=[72+1]^2=5329$ grid points. This more accurate PES shown in *Figure 38A,* and the duplicated minima range, given in equation [48b] is illustrated in *Figure 38B*.
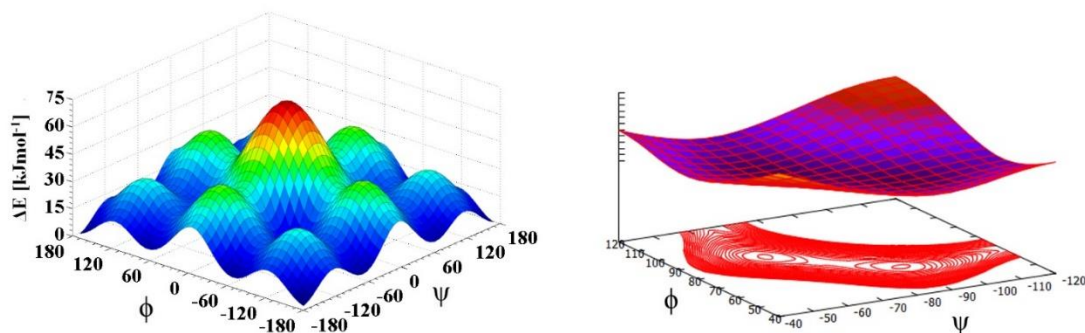
Figure 38 A *PES of n-pentane generated at 5°increments.* ***B*** *Enlarged area of the duplicated minima and TS between then.*

In the second step a more accurate function was fitted. In *Figure 39* the PES of n-pentane, generated from there *ab initio* grid points are shown. The two diagonal type crossections indicate that the TS generated by the steric hindrance split up is very small. For the two 1D crossection, of **a** and **b** (*Figure 39*) the following function was shown.

$$\Delta E_b(\phi) = A_0 + \sum_{k=1}^{n} (A_k \cos k\phi + B_k \sin k\phi) \qquad [49]$$

In the case of **a** five term (n=5) and in the case of **b** 3 term (n=3) expansions were fitted successfully. For the **a** crossection the high energy minimum pair had an energy value of 14.62 kJmol[-1] and the small transition state had 14.10 kJmol[-1]. Thus, the small barrier turned out to be 14.62-14.10= 0.52 kJmol[-1]. This is reasonably close to the *ab initio* value of 0.44 kJmol[-1].
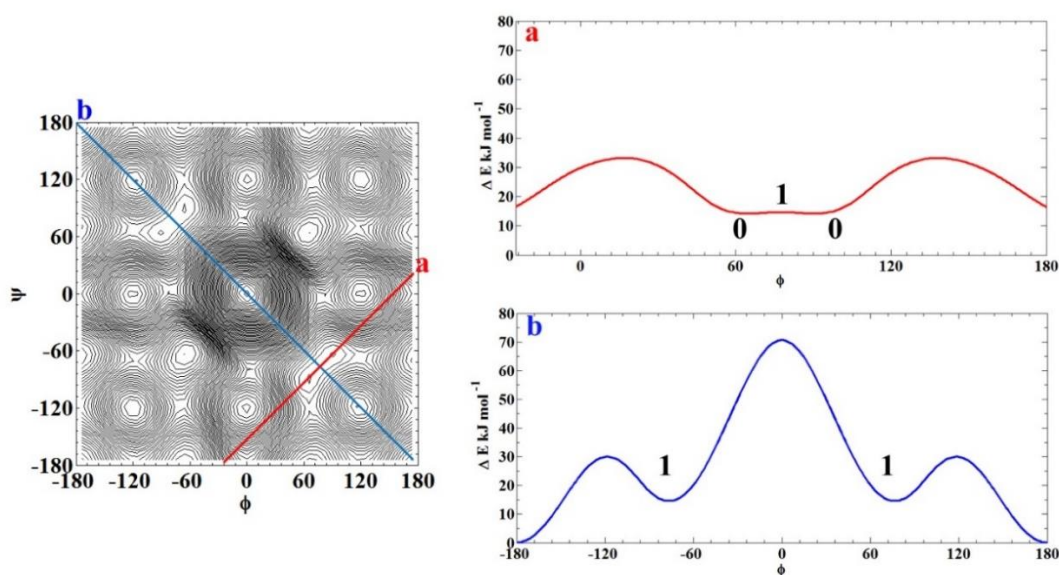

Figure 39 *Enlarged area of the duplicated minima and TS between then.*

Neither the previously fitted surface [45] with R^2= 0.950 nor the more complicated function [50] with R^2=0.9937 were able to reproduce the double minima even though statistically speaking the latter on was more accurate.

$$E(\varphi,\psi) = a_0 + \sum_{m=1}^{9} A_m e^{\left(-\left(\frac{c_\varphi(\varphi-\varphi_{0m})^2}{2\sigma_{\varphi m}^2} + \frac{c_\psi(\psi-\psi_{0m})^2}{2\sigma_{\psi m}^2}\right)\right)}$$ [50]
$$+ \sum_{m=1}^{6} (a_1\cos m\omega\varphi + b_1\cos m\omega\psi + a_2\sin m\omega\varphi + b_2\sin m\omega\psi)$$

Thus, with the higher R^2 value the overall fit was better, but it was not good enough locally to reproduce the double minima.

The elements for G matrix for glycine diamide, and alanine diamide are summarized in *Table 9* and *10*.

Table 9 *Critical points of the conformational PES of glycine diamide obtained by fitted analytic function and from first principle direct geometry optimization.*

Fitted Results				Optimized Results			
Conformation		Energy	Index	Conformation		Energy	
ϕ	ψ	ΔE [kJ/ mol]	λ	ϕ	ψ	ΔE [kJ/ mol]	
Minima							
-120.20	19.39	12.65	0	-122.78	21.93	10.24	δ_L
120.20	-19.39	12.65	0	122.78	-21.93	10.24	δ_D
178.80	178.80	2.66	0	179.99	180.00	2.58	β_L
-81.43	50.41	2.28	0	-82.15	68.54	0.00	γ_L
81.43	-50.41	2.28	0	82.15	-68.54	0.00	γ_D
Transition States							
118.20	-63.64	30.99	1	111.21	56.46	33.64	TS_1
118.20	-63.64	30.99	1	-111.21	-56.46	33.64	TS_2
114.50	-27.27	9.47	1	109.38	-31.26	10.57	TS_3
-114.50	27.27	9.47	1	-109.38	31.26	10.51	TS_4
100.00	-118.20	7.31	1	96.17	-122.41	9.49	TS_5
-100.00	118.20	7.31	1	-96.17	122.41	9.49	TS_6
-1.81	-1.81	29.19	1	0.00	0.00	30.05	TS_7
Maxima							
1.81	180.00	82.11	2	0.00	180.00	92.264	Max_1

Table 10 *Critical points of the conformational PES of alanine diamide obtained by fitted analytic function and from first principle direct geometry optimization.*

Fitted Results				Optimized Results			
Conformation		Energy	Index	Conformation		Energy	
ϕ	ψ	ΔE [kJ/ mol]	λ	ϕ	ψ	ΔE [kJ/ mol]	
Minima							
-159.30	-28.79	24.50	0	-165.06	-30.94	29.28	δ_D
67.17	21.11	28.07	0	67.92	26.68	24.21	α_D
-117.10	19.10	14.71	0	-126.55	21.42	13.04	δ_L
67.33	67.17	10.05	0	73.52	-57.39	10.88	γ_D
-155.50	167.00	6.06	0	-158.33	163.41	5.97	β_L
-86.36	78.89	2.23	0	-82.87	72.60	0.00	γ_L
Transition States							
-114.50	-60.00	35.02	1	-117.286	-56.22	37.20	TS_1
-5.45	12.73	26.24	1	-1.646	-12.175	36.19	TS_2
121.80	-146.60	36.88	1	113.03	-146.263	34.37	TS_3
-103.63	121.52	9.81	1	-105.221	120.044	10.82	TS_4
78.85	85.45	43.42	1	78.219	85.047	48.22	TS_5
129.10	-27.27	39.81	1	135.626	-26.381	35.01	TS_6
-121.80	26.34	19.76	1	-112.938	33.86	13.83	TS_7
Maxima							
-9.78	-172.70	92.66	2	-0.454	-166.606	106.99	Max_1

Although the location of the critical points were determined by a single "unit cell" (ranging from -180° to +180° along both ϕ and ψ), the motions are periodic. *Figures 40* and *41* show a larger range (from -360° to +360° along both ϕ and ψ). These figures illustrate more extensively the periodicity at the conformational subspace. The green lines that interconnect minima are regarded as the network of the conformational subspace. Peptide folding occur along such reaction coordinates. The number of pathways increase dramatically with the increasing molecular size. Protein folding is therefore a consequence of the choice of the lowest energy path on an innumerable network of lines.
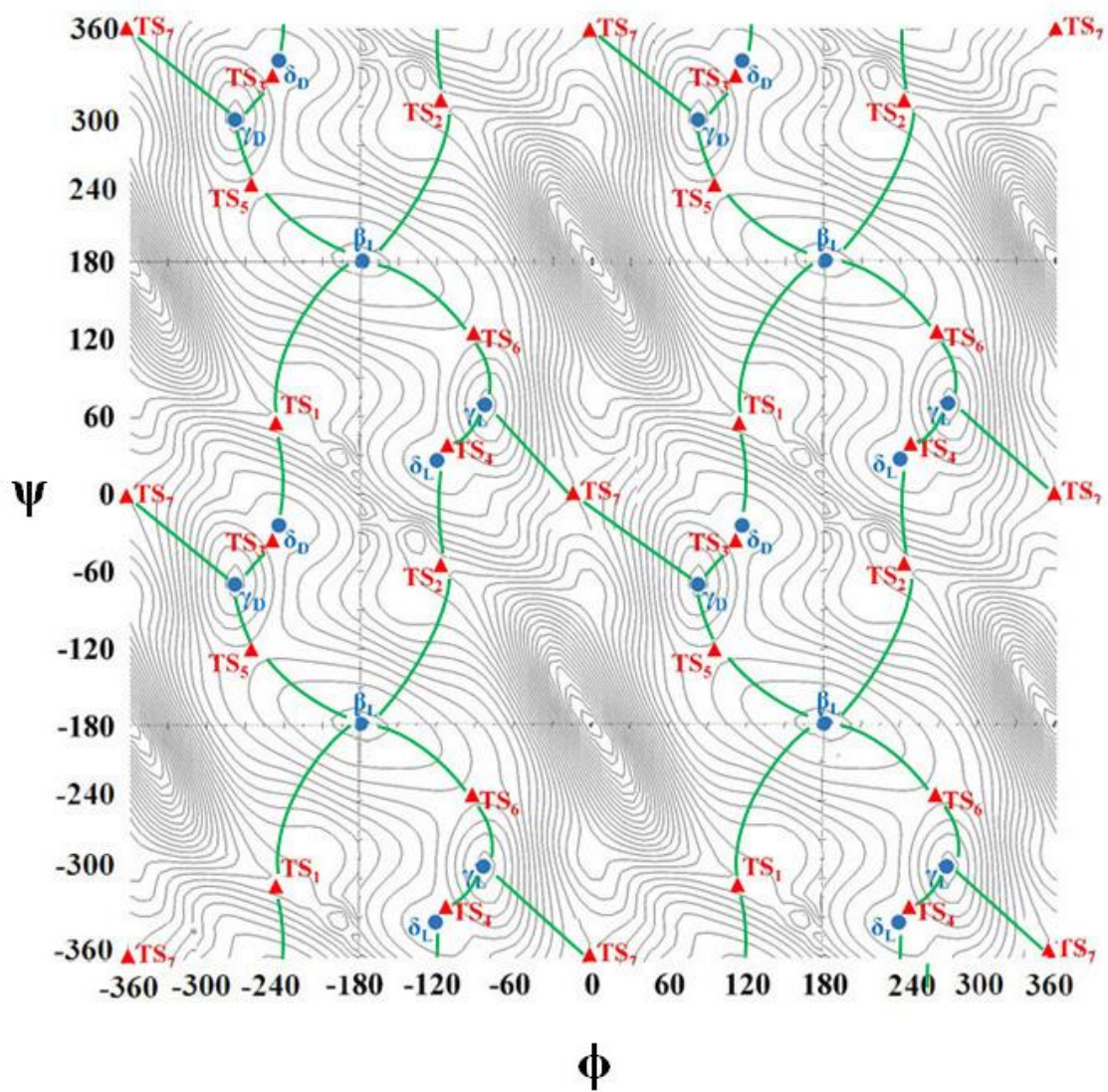
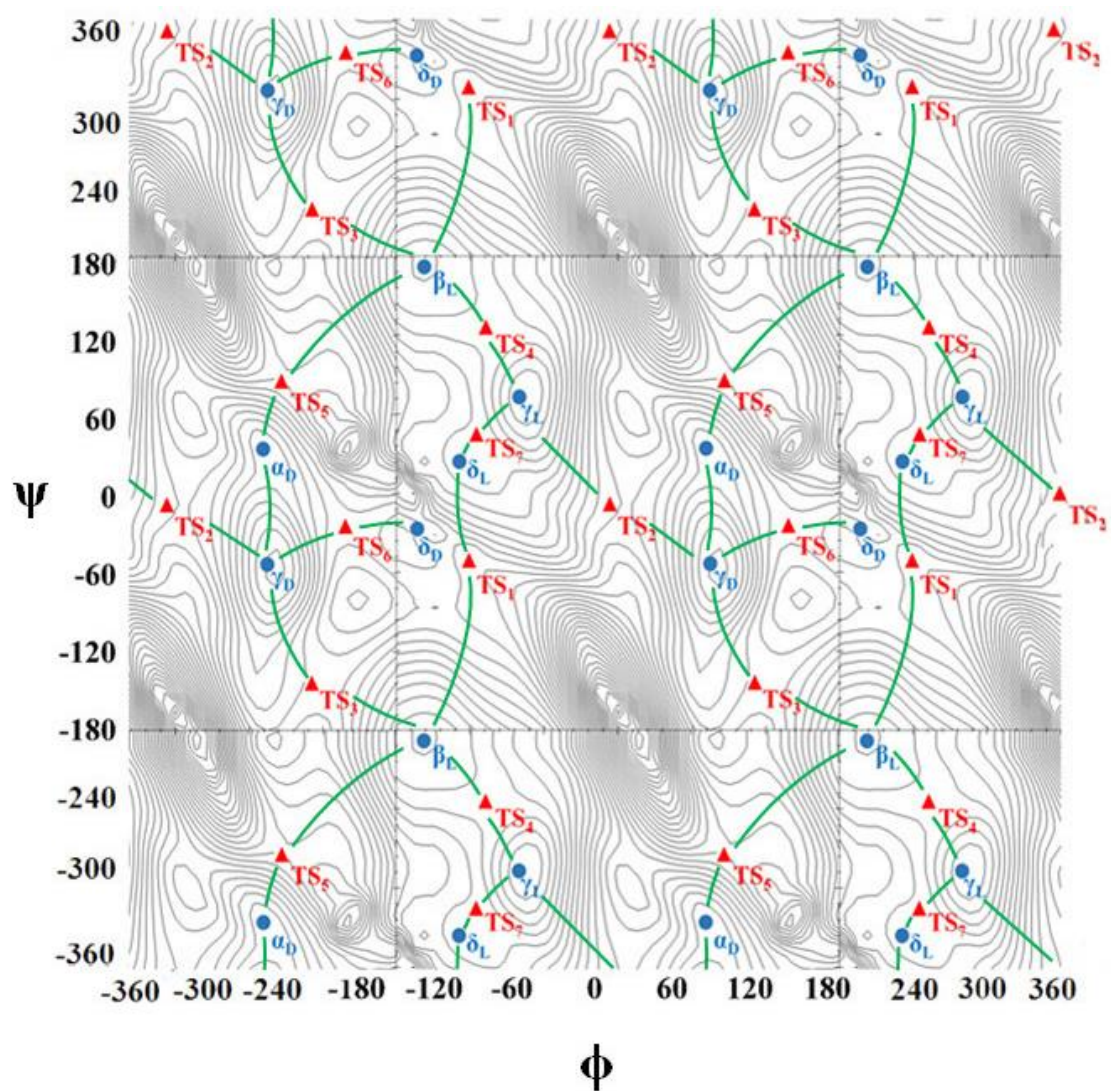**Figure 40** *Zeroth and the First-Order Critical points of the glycine diamide*

Figure 41 *Zeroth and the First-Order Critical points of alanine diamide*

4.4 The results of the minima structures predicting

From the following structural features of opiorphin (*Figure 42*) it is clear that there are $n_{BB}=5*2=10$ backbone dihedral angles [$(\phi i, \psi i)$ $1 \leq n \leq 5$] as independent variables for the backbone.
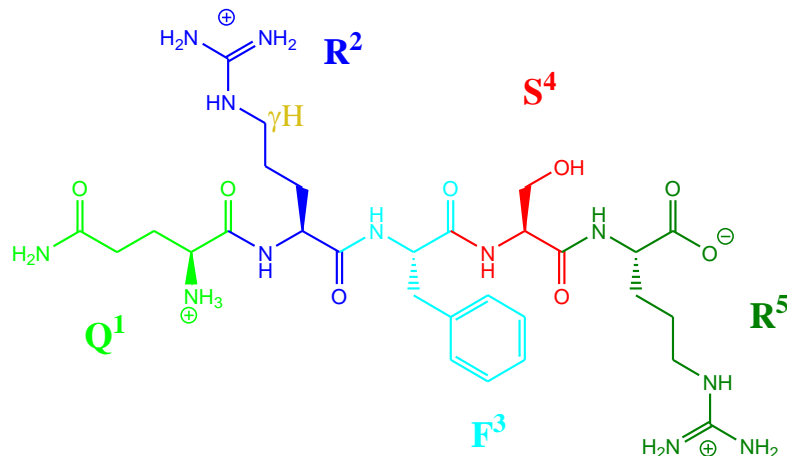

Figure 42 *A schematic structure of the pentapeptide opiorphin (QRFSR)*

Furthermore there are $n_{SC}=3+5+2+2+5=17$ dihedral angles as independent variables for the sidechain. Thus the PES has a grand total of 27 variables:

$$n_{total} = n_{BB} + n_{SC} = 27 \tag{51}$$

$$E = f \overbrace{(x_1, x_1, ...x_{10},}^{Backbone} \overbrace{x_{11}, x_{12}, ...x_{27})}^{Sidechain} \tag{52}$$

The number of minima (N) on the PES maybe be estimated by the following equation.

$$N = \overbrace{3^{10}}^{N_{Backbone}} \times \overbrace{3^{17}}^{N_{Sidechain}} = \overbrace{3^{27}}^{N_{Total}} \tag{53}$$

If one were to attempt to generate a grid for the backbone portion of the PES for

$$\left(\frac{360}{15}+1\right) = 24+1 = 25 \text{ grid points along each variable then}$$

$$25^{10} = 9.54 \times 10^{13} \tag{54}$$

grid point would be needed. In contrast to that, the full PES would require the following number of grid points:

$$25^{27} = 5.55 \times 10^{37} \tag{55}$$

This is beyond what is nowadays normally referee to as "big data".

Nevertheless, with the aid of the preliminary study it was possible to determine the global minimum of the pentapeptide, which is shown in *Figure 43*.
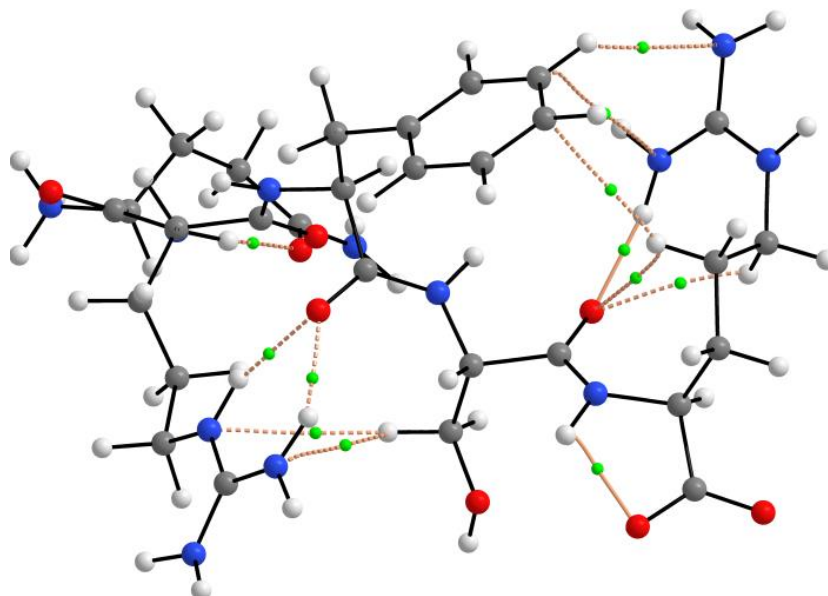

Figure 43 *The global minima of opiorphin. Intramolecular hydrogen bonds are shown by yellow lines and the bond critical points of the hydrogen bonds are denoted by green dots. The bond critical points were obtained by Bader's analysis. Weak interactions are denoted by broken yellow line and strong interaction are denoted by solid yellow line.*

5. Conclusions

The final and far reaching conclusion is that the minima ($\lambda=0$) and the transition states ($\lambda=$) located on a Ramachandran type Potential Energy Surface of a peptide does form a conformational network. Such a network of 0th and the 1st order critical points are associated with the conformational subspace of the peptide or protein structure. Consequently a given network represents all interconversional pathways associated with peptide or protein folding. Clearly, the dimensionality of protein folding is growing very dramatically with the number of amino acids that make up the primary structure of the protein.

Fourier series, comprised by sin and cos terms are capable to fit any periodic function. However, the single variable potential energy functions considered here are close to being even functions which makes a pure cosine expansion satisfactory. Depending on the structural complexity of a given molecule, a 1-term or 2-term or 3-term expansion turned out to be adequate, using only cos expansion. For C-C, C-N, C-O and C-S bonds the 1-term expansion was accurate enough. For O-O, S-S and S-O bonds the 2-term expansion was satisfactory. In contrast, 3-term expansion was necessary to fit the potential energy function associated with the internal rotation about the N-O, N-S and N-N bonds. The C-C rotation potential, fitted by Molecular Mechanics (MM), agreed surprisingly well with the present Fourier series fit for the methyl rotation in ethane as well as in alanine-diamide.

Any conformational PES can be described with an accurate multivariable mathematical function. Depending on the structural complexity of the given molecule, the surface can be characterized by a combination of Fourier series and Gaussian functions.

Molecules, which have two symmetric rotational groups (like CH_3-) and their surfaces have symmetry axes, can be described with a linear combination of the single Fourier series. In the case, when the molecule has peptide bonds, the surface becomes less-symmetric and the fitted functions are increasingly more complicated, thus, Gaussian functions are also needed to fit the surface. If the surface has only central symmetry or has no symmetry at, all the fitting needs a dissymmetrical correction function, in the form of an extended Fourier series.

The present dissertation suggests that the potential energy surfaces and hypersurfaces of flexible molecules, such as peptides, containing several internal bond-rotations, may be reasonably well represented by these types of fitting method. The ultimate purpose is to build a very accurate multi variable mathematical function for the Potential Energy Hypersurface of flexible molecules, such as peptides, of several internal bond-rotations. For such macro molecules the grid points would be in the domain of big data. For example for an oligiopeptide containing 10 amino acids using 15° increments that leads to 25 (360/15+1=25) points would need something like the order of 10^{28} grid points due to the following relationship[24] [56]:

$$N = 25^{2n} = 625^n \approx 10^{28}$$ [56]

A fitted analytic function would reduce this big data into a more manageable set.

As a more concrete example the pentapeptide opiorphin were studied and in this case not the surface fitting but the determination of the critical points required computational resources nearing the limits of computational capabilities.

The accuracy of fitted functions showed about the same accuracy, in terms of R^2 and RMS, for 5° and 15° grids. However, the accuracy of the fit increased with increasing complexity of the fitted increasingly more extensive function. In all cases, fitted analytic functions, using Fourier and Gaussian components could reproduce well defined minima.
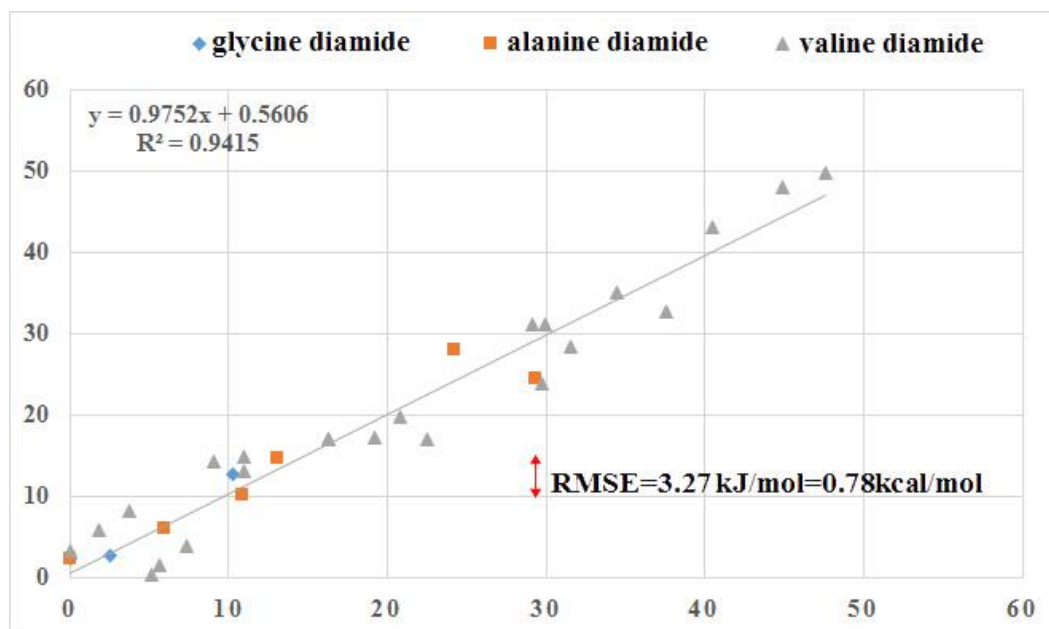
Figure 44 *Comparison of fitted and optimized energy minima glycine, alanine- and valine diamide*

The combined accuracy (RMSE) obtained for glycine, alanine and valine diamides is about 3.27 kJ/mol (or 0.78 kcal/mol) which is somewhat better than the previously accepted chemical accuracy of 1 kcal/mol (*Figure 44*). Consequently, very shallow minima, which are separated by transition states lower than 0.52 kJ/mol (or 0.12 kcal/mol) are not reproducible at the current level of the approach. This was observed in the case of the highest energy minimum of n-pentane. The PES was originally described by 625 energy values defined by the two rotated dihedral angles. Fitting a function describing the conformational space of the species drastically reduced the amount of data (*Table 11*) . In the case of single amino acid diamides the 18.8% reduced data means that 81.2% of data points were eliminated by the fitted parameters.

Table 11 Obtained data reduction by fitting mathematical functions

Compound	Number of grid points	Number of fitted parameters	Reduced amount of data [%]
propane	625	25	4
n-pentane	625	70	11.2
Glycine diamide	625	118	18.8
Alanine diamide	625	118	18.8

The above statistical characterization may be frequently used for analyzing the accuracy of fitting but it may not be all that relevant in determining the pathways of the process of protein folding. Also the $R^2=0.9415$ represents in some way the overall accuracy of the fit. However, the accuracy of the fit or the error of the fit is not homogenous. In certain regions such as near to the maxima of the high mountain ridges the error of the fit may be larger than shallower regions.

This can be clearly seen on *Figure 44*. The two highest energy points of the alanine, corresponding to the α_D and δ_D conformations, are located near the high energy mountain ridge and therefore they show the largest fitting error. Consequently, in any future study special attention should be given to these critical points.

Table 12 shown that the improving Fourier expansion the accuracy is also improving.

Table 12 *As a case study of propane and n-pentane showing the accuracy of the fit is increasing with the complexity of the Fourier expansions.*

	Function	Term	R^2	RMSE [kJ/mol]
propane I	[19]	4	0.996	0.444
	[19]	24	0.998	0.344
n-pentane II	[19]	12	0.950	4.100
	[19]+[20]	33 (24+9)	0.997	0.992

6. Future perspectives

To predicting the total Potential Energy Hypersurface, we must describe two subspaces, the reaction subspace and the conformational subspace (*Figure 45*).

The function of the total PEHS is:

$$E = f(x_1, x_2, x_3, \ldots x_D) \qquad [57]$$

If the number of atoms have the following form:

$$C_k H_l N_m O_n \qquad [58]$$

The degrees of freedom (D) of the system given of N number of atoms:

$$D = 3N - 6 \qquad [59]$$

where the N is:

$$N = k + l + m + n \qquad [60]$$

The number of stretches, bends and the torsion add up to 3N-6.

$$
\begin{array}{ll}
N-1 & strech \\
N-2 & bend \\
N-3 & torsion \\
\hline
3N\text{-}6 & \textbf{Degrees of freedom}
\end{array}
$$

Reaction subspace

Conformation subspace

[61]

Therefore the function of the full potential energy surface has the following form:

$$E = f \overbrace{(x_1, x_2, x_3, \ldots x_{2N-3}}^{\substack{reaction \\ subspace}}, \overbrace{x_{2N-2} \ldots x_{3N-6}}^{\substack{conformational \\ subspace}}) \qquad [62]$$
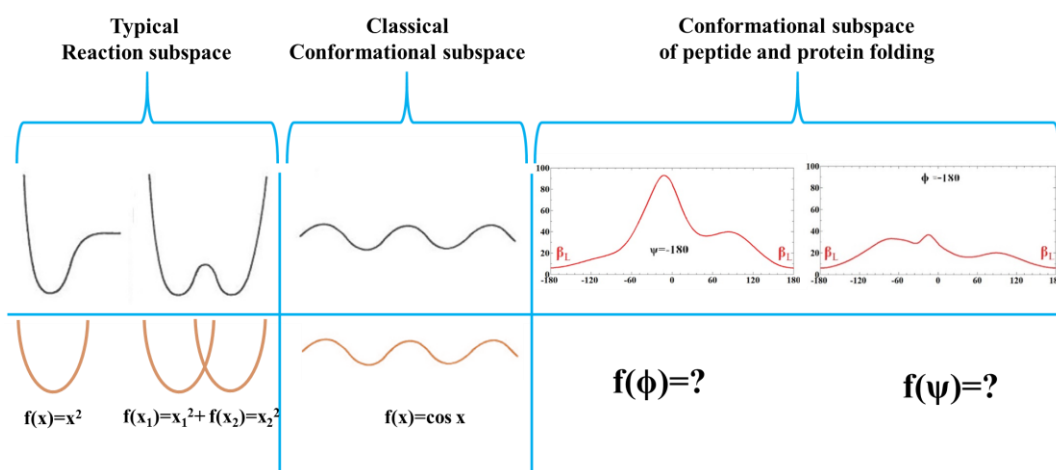
Figure 45 A schematic summary of the type of functions to be used by reaction subspace and conformation subspace

The reaction subspace can be represented by a power series and the conformational subspace can be represented by a trigonometric series. The bottom up approach for protein folding maybe started the simple amino acid diamides but it should be continued studying small oligopeptides such as di-amino acid and tri-amino acid diamides. In a sense these represent the first couple of steps of protein folding biosynthesis as illustrated in *Figure 46 A*. The corresponding energy changes given in *Figure 46 B* where the separation of the reaction subspace and the conformational subspace are clearly indicated.
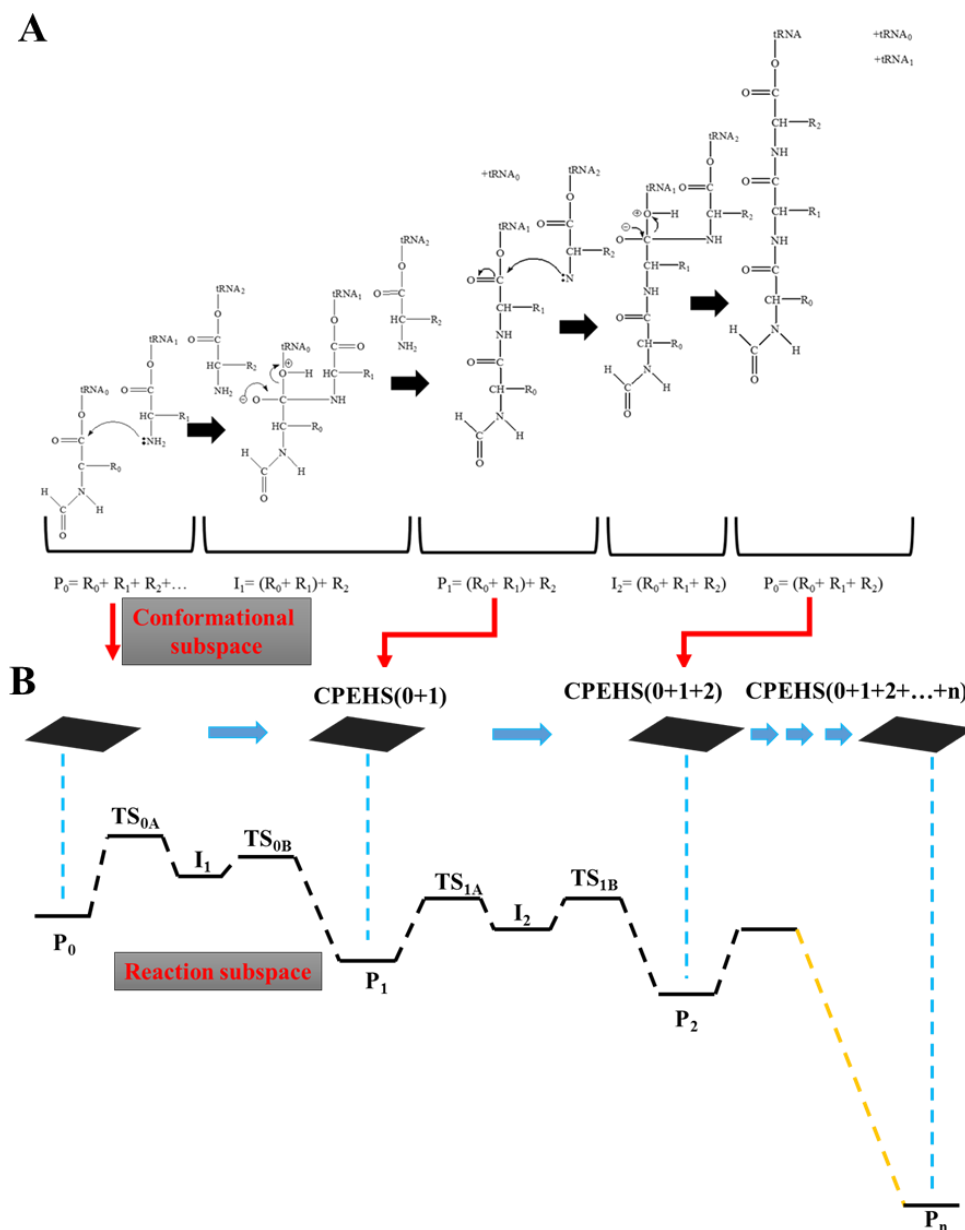
Figure 46 A Reaction of tripeptide formation. B: Schematic free energy levels of protein synthesis and folding.

Since in the peptide the primary sequence of amino acids is quite significant therefore the various dipeptides can be represented by a 3x3 matrix leading to 9 different dipeptides as shown below [56].

$$
\begin{array}{cccc}
 & 1 & 2 & 3 \\
1 & 1,1 & 1,2 & 1,3 \\
2 & 2,1 & 2,2 & 2,3 \\
3 & 3,1 & 3,2 & 3,3
\end{array}
\Rightarrow
\begin{array}{ccc}
gly-gly & gly-ala & gly-val \\
ala-gly & ala-ala & ala-val \\
val-gly & val-ala & val-val
\end{array}
\qquad [63]
$$

Similarly in the tripeptide cases total 3x3x3=27 unique primary structures can be generated which is associated a 3D matrix.
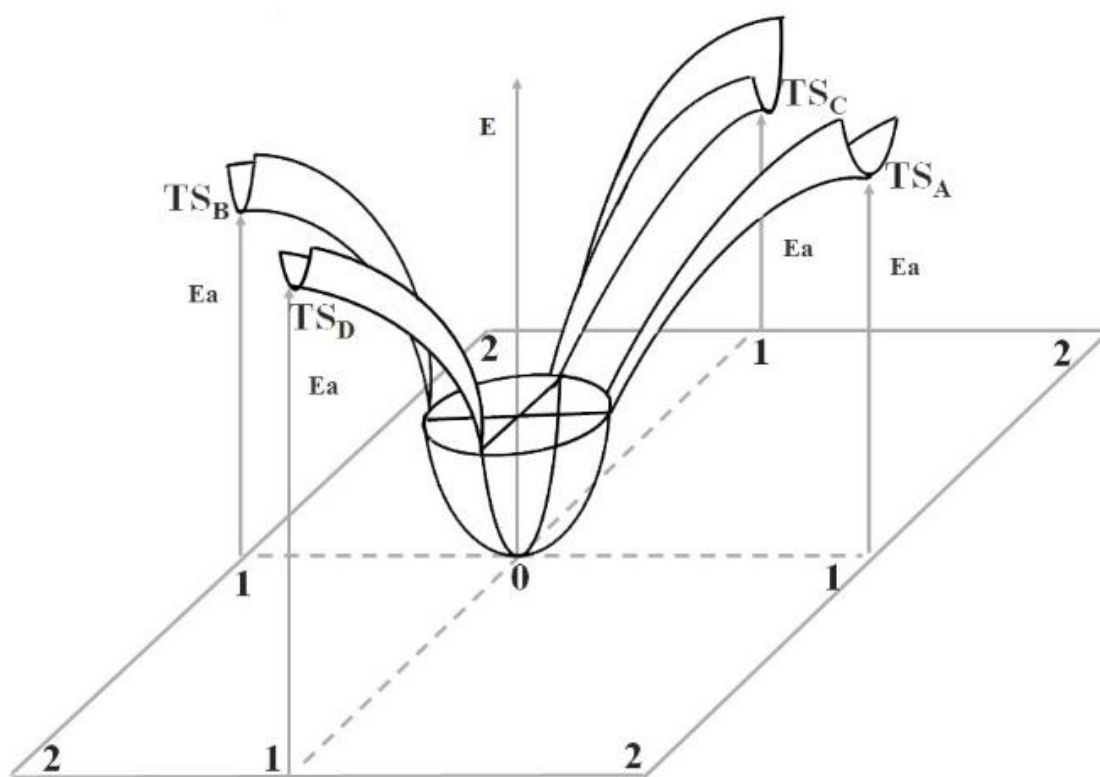
Figure 47 A schematic illustration of the 4 TS surrounding a minimum in a 2D-conformational subspace.

In essence the folding problem is related to the passage of one minimum energy (one foldamer structure) to another minimum energy (another foldamer structure) *via* a transition state (TS). The backbone of a single amino acid diamide is represented by a 2D Conformational Potential Energy Surface and each minimum may be at the cross-road of

2 pathways and therefore each minimum may be surrounded by 2 TS in each of the 2 pathways leading to 2x2=4 TS *(Figure 47)*. If there is a side chain rotation, like in valine then a minimum would be surrounded by 3x2=6 TS.

The ultimate problem is to find the valley down from the various TS to the nearby minima. It is hoped that the extension of the present surface fitting project may be directed to such a direction.

Finding the right methodologies for data set reduction is an arduous work. When proposing such methodologies it is reasonable to start with the description of small compounds, and aim for a bottom up solution. The long term goal would require extensive research as illustrated by *Figure 48.* Final conclusions concerning the mathematical representation of higher peptides can only be made when the study of di- and tripeptides completed.
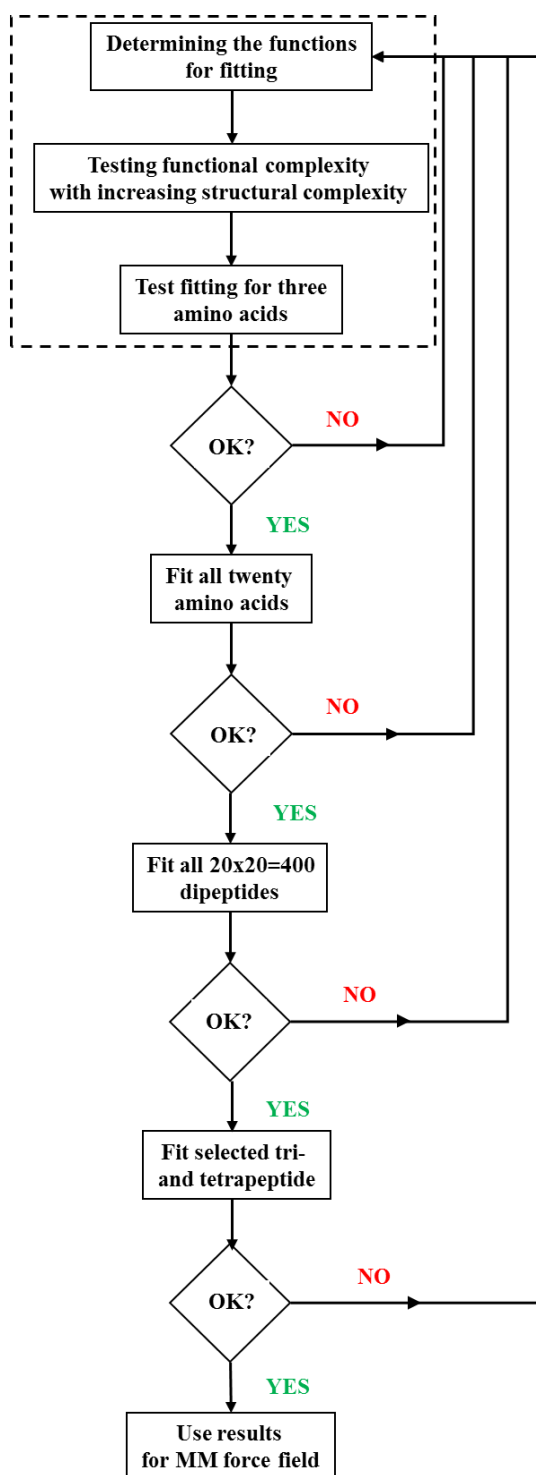
Figure 48 *A schematic flow-chart showing the phases an extensive research from the optimization of fitted function to the use the results in Molecular Mechanics force field*

Acknowledgements

References

[1] P. Dirac, Proc. R. Soc. Lond. **A123**, 714 (1929).

[2] E. Schrödinger, Ann. Phys. **79**, 361 (1926).

[3] F.E. Allen, IBM J. Res. Dev. **25**, 535 (1981).

[4] A.D. McLean and M. Yoshimine, IBM J. Res. Dev. **12**, 206 (1968).

[5] J. Kepler, *Harmonices Mundi* (Linz, 1619).

[6] I. Newton, *Philosophiæ Naturalis Principia Mathematica* (1687).

[7] J. v. Neumann, in *"The Work. Mind"* (University of Chicago Press, 1947), pp. 180–196.

[8] National Research Council, *Mathematical Challenges From Theoretical/Computational Chemistry* (National Academy Press, Washington D. C., 1995).

[9] A. Neumaier, SIAM Rev. **39**, 407 (1997).

[10] Science **309**, 78 (2005).

[11] S. Istrail and F. Lam, Commun. Inf. Syst. **9**, 303 (2009).

[12] F. Torrens and G. Castellano, J. Argent. Chem. Soc. **94**, 27 (2006).

[13] A. FRAENKEL, Bull. Math. Biol. **55**, 1199 (1993).

[14] J. Alonso, G. Chass, I. Csizmadia, P. Echenique, and A. Tarancon, arXiv (2004).

[15] H.B. Rasmussen, S. Branner, F.C. Wiberg, and N. Wagtmann, Nat. Struct. Biol. **10**, 19 (2003).

[16] C. Anfinsen and H. Scheraga, Adv Protein Chem **29**, 205 (1975).

[17] J. Moult, Philos Trans R Soc L. B Biol Sci **361**, 453 (2006).

[18] K.A. Dill and H.S. Chan, Nat Struct Biol **4**, 10 (1997).

[19] C. Levinthal, in *Mossbauer Spectrosc. Biol. Syst. Proc. a Meet. Held Allert. House*, edited by E.M. P. Debrunner, J.C.M. Tsibris (University of Illinois Press, Urbana, 1962), p. 22.

[20] I. Ramsen, New York J. A. Hill Co. (1904).

[21] Berg JM, J. Tymoczko, and L. Stryer, in *Biochemistry*, edited by W.H. Freeman, 5th ed. (New York, 2002).

[22] A. M. Lesk, *Introduction to Protein Architecture: The Structural Biology of Proteins* (Oxford University Press, 2000).

[23] T.S. Young and P.G. Schultz, J. Biol. Chem. **285**, 11039 (2010).

[24] I. Jákli, A. Perczel, B. Viskolcz, and I.G. Csizmadia, in *Protein Model.* (Springer International Publishing, 2014).

[25] G.N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, J Mol Biol **7**, 95 (1963).

[26] M. Born and R. Oppenheimer, Ann. Phys. **20**, 457 (1927).

[27] C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).

[28] J.A. Pople and R. K. Nesbet, J. Chem. Phys. **22**, 571 (1954).

[29] R. McWeeny and G. Dierksen, J. Chem. Phys. **49**, 4852 (1968).

[30] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).

[31] R.G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford Univ. Press, Oxford, 1989).

[32] W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965).

[33] E.D.R. Salahub and M. C. Zerner, *The Challenge of D and F Electrons* (ACS, Washington, D. C., 1989).

[34] Dreizler R. M. and Gross E. K. U., in *Density Funct. Theory* (Springer-Verlag, 1990).

[35] V. Rassolov, J.A. Pople, M. Ratner, P.C. Redfern, and l. A. Curtiss, J Comput Chem. **22**, 976 (2001).

[36] H.P. Hratchian and H.B. Schlegel, in *Theory Appl. Comput. Chem. First Forty Years*, edited by E. Clifford, F. Dykstra, K. Gernot, S. Kwang, and E. Scuseria Gustavo (2005), pp. 195–249.

[37] I.G. Csizmadia, *Basic Principles for Introductory Organic Chemistry* (Quirk Press, Toronto, 1997), pp. 135–141.

[38] A. Perczel, M. Kajtar, J.F. Marcoccia, and I.G. Csizmadia, J. Mol. Struct. THEOCHEM **232**, 291 (1991).

[39] M.R. Peterson, I.G. Csizmadia, and R.W. Sharpe, J. Mol. Struct. **94 (THEOCH**, 127 (1983).

[40] I. Jákli, S.J. Knak Jensen, I.G. Csizmadia, and A. Perczel, Chem. Phys. Lett. **547**, 82 (2012).

[41] J. Milnor, *Singular Points of Complex Hypersurfaces* (Princeton University Press, Princeton, N.J., 1968).

[42] A. Perczel, W. Viviani, and I. Csizmadia, in *Mol. Asp. Biotechnol. Comput. Model. Theor. SE - 3*, edited by J. Bertrán (Springer Netherlands, 1992), pp. 39–82.

[43] P.G. Mezey, Elsevier Sci. Publ. 227 (1987).

[44] P.G. Mezey, Chem. Phys. Lett. **86**, 562 (1982).

[45] P.G. Mezey, Chem. Phys. Lett. **82**, 100 (1981).

[46] J.G. Ángyán, R. Daudel, Á. Kucsman, and I.G. Csizmadia, Chem. Phys. Lett. **136**, 1 (1987).

[47] I.G. Csizmadia, in *New Theor. Concepts Underst. Org. React. SE - 1*, edited by J. Beltrán and I.G. Csizmadia (Springer Netherlands, 1989), pp. 1–31.

[48] M.R. Peterson and I.G. Csizmadia, J. Am. Chem. Soc. **100**, 6911 (1978).

[49] G. Tasi, B. Nagy, G. Matisz, and T.S. Tasi, Comput. Theor. Chem. **963**, 378 (2011).

[50] L. Radom, J.W. Hehre, and J.A. Pople, J. Am. Chem. Soc. **94**, 2371 (1972).

[51] L. Radom, W.A. Lathan, W.J. Hehre, and J.A. Pople, J. Am. Chem. Soc. **95**, 693 (1973).

[52] L. Radom and J.A. Pople, J.. Am. Chem. Soc. **94**, 4786 (1970).

[53] M. Headgordon and J.A. Pople, J Phys Chem-Us **97**, 1147 (1973).

[54] A. Chung-Philips, J. Chem. Phys **88**, 1764 (1988).

[55] T.A.K. Kehoe, M.R. Peterson, G.A. Chass, B. Viskolcz, L. Stacho, and I.G. Csizmadia, J. Mol. Struct. THEOCHEM **666–667**, 79 (2003).

[56] M.R. Peterson and I.G. Csizmadia, J. Amer. Chem. Soc. **101**, 1076 (1979).

[57] T.A. Modro, W.G. Liauw, M.R. Peterson, and I.G. Csizmadia, J. Chem. Soc. Perkin **2**, 1432 (1979).

[58] G.R. DeMare, O.P. Strausz, M.R. Peterson, and I.G. Csizmadia, J. Comp. Chem. **1**, 141 (1980).

[59] M.R. Peterson, G.R. DeMare, I.G. Csizmadia, and O.P. Strausz, J. Mol. Struct. **86**, 131 (1981).

[60] M.R. Peterson and I.G. Csizmadia, Theor. Org. Chem. **3**, 190 (1982).

[61] M.R. Peterson, G.R. DeMare, I.G. Csizmadia, and O.P. Strausz, J. Mol. Struct. **92**, 239 (1983).

[62] R.E. Kari and I.G. Csizmadia, J. Chem. Phys. 1443 (1969).

[63] S. Wolfe, A. Rauk, and I.G. Csizmadia, Can. J. Chem. 113 (1969).

[64] S. Wolfe, A. Rauk, and I.G. Csizmadia, J. Am. Chem. Soc. 1567. (1969).

[65] J. v. Neumann, in *Unity Knowl.*, edited by L. Leary (1955), pp. 157–164.

[66] K.H. Frisch et al., (2009).

[67] A. D. Becke, J. Chem. Phys **98**, 5648 (1993).

[68] G.P. Tolstov, *Fourier Series* (Courier-Dover, 1976).

[69] MATLAB 2013b, The MathWorks Natick (2013).

[70] K. Levenberg, Quart. Appl. Math. **2**, 164 (1944).

[71] J.J. More, Lect. Notes Math. **630**, 105 (1978).

[72] D. Marquardt, SIAM J. Appl. Math. **11**, 431 (1963).

[73] H.L. Harter, in *Encycl. Stat. Sci.* (John Wiley & Sons, Inc., 2004).

[74] A. Wisner, E. Dufour, M. Messaoudi, A. Nejdi, A. Marcel, M.N. Ungeheuer, and C. Rougeot, Proc Natl Acad Sci U S A **103**, 17979 (2006).

[75] M. Pinto, C. Rougeot, L. Gracia, M. Rosa, A. García, G. Arsequell, G. Valencia, and N.B. Centeno, ACS Med. Chem. Lett. **3**, 20 (2012).

[76] A. Kotynia, E. Kamasz, H. Czapor, and J. Brasuń, Tetrahedron Lett. **51**, 2486 (2010).

[77] Y. Tong, M. Tar, V. Monrose, M. DiSanto, A. Melman, and K.P. Davies, J Urol **178**, 338 (2007).

[78] D.A. Case, T.A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, M.-J. Hsieh, G. Cui, D.R.

Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, and A. Kovalenko, (2012).

[79] B. Leitgeb, Chem Biol Drug Des **79**, 313 (2012).

[80] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins **65**, 712 (2006).

[81] H. Nguyen, D.R. Roe, and C. Simmerling, J. Chem. Theory Comput. **9**, 2020 (2013).

[82] D.B. Axel, J Chem Phys **98**, 1372 (1993).

[83] P.J. Stephens, F.J. Devlin, C.F. Chabalowski, and M.J. Frisch, J. Phys. Chem. **98**, 11623 (1994).

[84] A. V Marenich, C.J. Cramer, and D.G. Truhlar, J. Phys. Chem. B **113**, 6378 (2009).

[85] R.D. Johnson III, NIST Stand. Ref. Database Number 101, Release 16a http://cccbdb.nist.gov/ (2013).

[86] R.F.W. Bader, Chem. Rev. **91**, 893 (1991).

[87] L. Song, M. Liu, W. Wu, Q. Zhang, Y. Mo, J. Chem. Theory Comput. **1**, 394 (2005).

Summary

The Potential Energy Surfaces (PES), as a function of two independent variables, playing an important role in chemistry. They are containing information about the molecular structure, thermodynamic properties and chemical reactivities. To predicting the total Potential Energy Hypersurface, we must to describe two subspaces, the reaction subspace and the conformational subspace. Although all chemical changes maybe described by a PES, nevertheless conformational PES are representing a special case which will be the subject of the present dissertation.

The long term aims for the present dissertation is to find mathematical functions which describe the folding of the peptide residues. In order to find eventually the solution to solve the protein folding problem and to build up the conformational network for the folding but it is reasonable to start with the description of small compounds, and aim for a bottom up solution.

The aims for the present dissertation:

➢ to achieve a reasonable accuracy of the fitted functions for the folding of the internal rotation of typical organic functional groups which have only one independent variable,

➢ to extend the one dimensional mathematical functions to describe the folding of the molecules which have two independent torsional angles and they are include peptide bonds,

➢ the fitted functions, which are describe the conformational spaces, can be analyzed to obtain the critical points, minima, transition states and maxima.

The relative energies, ΔE, associated with internal rotation were calculated by quantum mechanics using the B3LYP/6-31G(d) implementation of the density functional theory in gas phase in the Gaussian09 software package. The calculations were carried out for a range of dihedral angles ϕ and ψ of the molecules in the interval $[-\pi, \pi]$ with grid points at $15°$ increments in order to generate the curve, surface and hypersurface.

The energy changes associated with internal rotation of a functional group in a molecule depend on the topology of the chemical environment. Energies obtained from electronic structure and force-field calculations have been analyzed by Fourier expansions [1]. The findings show that rotation around bonds, connecting atoms without

lone pairs, can be described with a one term trigonometric function. In contrast to that, two or three terms are needed if the connected atoms have lone pairs.

$$\Delta E(\varphi) = a_0 + \sum_{n=1}^{m}\left(a_n \cos n\omega\varphi + b_n \sin n\omega\varphi\right) \qquad [1]$$

where a_0 is a constant, m is the number of terms in the expansion and ω is the fundamental frequency of the signal.

The analysis inspires adoption of a simplified Fourier expansion that reproduces the data well, suggesting that Fourier-type-series with few terms are useful in describing any internal rotation analytically.

The potential energy surface associated with internal rotation of a pair of geminally located functional groups was studied using electronic structure calculations. The functional groups were attached to a methylene carbon and were chosen as saturated hydrocarbons, and heteroatom containing moieties like amide bond. For the majority of the compounds studied, extended Fourier expansions, augmented with Gaussian functions, were needed to achieve accuracy within a few kJ/mol. The present dissertation aims to take the first steps of a bottom up solution for protein folding by finding the analytic functions for the Ramachandran type PES of small peptide residues.

Double rotor Potential energy surfaces (PES) were generated and analytic functions were fitted for the grids for propane [2], n-pentane [3], glycine diamide [4], alanine diamide [4] and valine diamid [4]. The accuracy of the fit improved with the complexity of the fitted functions. The less extensive functional fit led to $R^2 = 0.911$ and RMSE= 4.059 while the more extensive fitted function resulted in $R^2 = 0.998$ and RMSE = 0.344.

$$E(\varphi,\psi) = a_0 + (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi) \qquad [2]$$

where a_0 is the constant term in the series, ω is the fundamental frequency of the signal and m is the number of the terms.

$$E(\varphi,\psi) = a_0 + \sum_{m=1}^{4}\ (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi) \qquad [3]$$

where a_0 is the constant term in the series, ω is the fundamental frequency of the signal and m is the number of the terms.

$$E(\varphi,\psi) = a_0 + \sum_{m=1}^{9} A_m e^{\left(-\left(\frac{c_\varphi (\varphi-\varphi_{0m})^2}{2\sigma_{\varphi m}^2} + \frac{c_\psi (\psi-\psi_{0m})^2}{2\sigma_{\psi m}^2}\right)\right)}$$

$$+ \sum_{m=1}^{6} (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi)$$

$$+ \sum_{m=1}^{6} f_1 \cos(m\omega\varphi + m\omega\psi) f_2 \cos(m\omega\varphi - m\omega\psi) + f_3 \cos(m\omega\varphi + m\omega\psi f_4 \sin(m\omega\varphi - m\omega\psi))$$

$$+ f_5 \sin(m\omega\varphi + m\omega\psi) f_6 \cos(m\omega\varphi - m\omega\psi) + f_7 \sin(m\omega\varphi + m\omega\psi) f_8 \sin(m\omega\varphi - m\omega\psi)$$

[4]

where a_0 is the constant term in the series, ω is the fundamental frequency of the signal, m is the number of the terms, A is the amplitude, φ_0 and ψ_0 are define the center and the σ_φ and σ_ψ are the φ and ψ extension of the ellipsoid.

The locations of the critical points were determined for both the hydrocarbons and the amino acids studied from the first and second derivatives of the fitted functions. The pathways of the glycine diamide and alanine diamide were analyzed and the network of the conformational subspace were described by the location of the critical points. Peptide folding occur along such reaction coordinates. The number of pathways increase dramatically with the increasing molecular size. Protein folding is therefore a consequence of the choice of the lowest energy path on an innumerable network of lines.

Beyond the relatively simple example, case study was performed to estimate how many of the minima of the opiorphin pentapeptide maybe of low enough energy to have any biological significance. It turned out that at least 3000 conformer were of relatively low energy (0-80 kJ/mol). In the case of opiorphin not the surface fitting but the determination of the critical points required computational resources nearing the limits of computational capabilities. Nevertheless, with the aid of the preliminary study it was possible to determine of the global minimum of the pentapeptide

Összefoglaló

A potenciális energiafelületek fontos szerepet töltenek be a kémiában. Információt hordoznak a molekula szerkezetéről, termodinamikai tulajdonságairól, valamint a kémiai reaktivitásról. A teljes potenciális energiafelület két altérből áll; a reakciós altérből és a konformációs altérből. A potenciális energiafelületekkel az összes kémiai változás leírható, ezért jelen disszertáció a konformációs potenciális energiafelületekkel foglalkozik.

A disszertáció távlati célja egyrészt az, hogy olyan matematikai függvényt állítsunk elő, amely leírja a fehérjefelcsavarodás teljes konformációs terét, ez pedig közelebb vihet minket a probléma megoldásához. Másrészt, hogy ezzel a matematikai közelítéssel a teljes konformációs hálózatot felrajzoljuk, amely egy útvonal a minimumok és az átmeneti állapotok között. Ezt célszerű az alapoktól, kis vegyületek leírásával kezdeni, melyekből később következtetni lehet nagyobb molekulák leírásához használható függvényekre.

A disszertáció célkitűzései ezért a következők:

➤ elfogadható pontosság mellett létrehozni egy olyan függvényt szerves kismolekulákra, melynek egyetlen független változója, azaz torziós szöge van.

➤ ezt az egyetlen független változós függvényt kiterjeszteni olyan szerves molekulákra, amelyeknek két torziós szögük van, és peptidkötést is tartalmaznak.

➤ ezek a függvények legyenek elemezhetők, segítségükkel meg lehessen határozni a minimumokat, a maximumokat és az átmeneti állapotokat is.

A relatív energia (ΔE) a molekulák belső forgatásából származik, melyet kvantumkémiai módszerrel, B3LYP sűrűségfunkcionál elmélet segítségével számítottunk, 6-31G(d) bázisfüggvényekkel gázfázisban, a Gaussian09 szoftvercsomag segítségével.

A görbék, felületek és hiperfelületek 15°-os felbontásban [-π, π] intervallumon készültek a molekulák ϕ és ψ szerinti torziós szögei alapján.

A molekulák belső rotációjának leírása véges Fourier-soron [1] alapszik, mely [-π, π] intervallumon periodikus függvény, ezért képes leírni a potenciális energiagörbék és -felületek periodikusságát. Az eredmények azt mutatták, hogy azok a molekulák, melyek nem tartalmaznak magános elektronpárokat, leírhatók egytagú trigonometrikus

függvénnyel. Ezzel ellentétben, ha a molekulák magános elektronpárokat tartalmaznak 2 vagy 3 tagú páros Fourier-sor szükséges a potenciális energiagörbe leírásához.

A Fourier-sor általános formája

$$\Delta E(\varphi) = a_0 + \sum_{n=1}^{m} \left(a_n \cos n\omega\varphi + b_n \sin n\omega\varphi \right) \tag{1}$$

ahol a_0 a sorozat konstans tagja, m a tagok száma, és ω a jel alapfrekvenciája. Ezek alapján megállapítható, hogy egy egyszerűsített Fourier-sor képes reprodukálni a konformációs változásból származó adatokat, azaz néhány tagú analitikus függvénnyel minden belső rotáció leírható.

A párosával elhelyezkedő funkciós csoportok belső rotációból származó potenciális energiafelületeinek tanulmányozása elektronszerkezet-számításokkal történt. A vizsgált vegyületek közül az első csoport a központi CH_2 egységhez csatolt telített forgó csoportokat, a második csoport pedig amid (-CONH-) kötéseket tartalmazott. A legtöbb tanulmányozott vegyület esetén szükség volt egy kiterjesztett Fourier-sorra az illesztéshez, valamint Gauss-függvényre is, hogy az eltérés néhány kJ/mol-nál kevesebb legyen. A disszertáció egyik hosszú távú célja, hogy olyan matematikai függvényt állítsunk elő, amely leírja a fehérjefelcsavarodás teljes konformációs terét, melyet célszerű az alapoktól, kis vegyületek leírásával kezdeni, melyekből később következtetni lehet nagyobb molekulák leírásához használható függvényekre.

A propán [2], pentán [3], glicin diamid [4], alanin diamid [4] és a valin diamid [4] potenciális energiafelületére analitikus függvényeket illesztettünk. Az illesztés pontosságát nagymértékben befolyásolta a függvény bonyolultsága. A legegyszerűbb függvény esetén a pontosság $R^2=0.911$ (RMSE=4.059), ugyanerre a molekulára (pentán) egy sokkal összetettebb függvény esetén a pontosság $R^2=0.998$ (RMSE=0.344).

$$E(\varphi,\psi) = a_0 + (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi) \tag{2}$$

ahol a_0 a sorozat konstans tagja, m a tagok sorszáma, és ω a jel alapfrekvenciája.

$$E(\varphi,\psi) = a_0 + \sum_{m=1}^{4} (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi) \tag{3}$$

ahol a_0 a sorozat konstans tagja, m a tagok sorszáma, és ω a jel alapfrekvenciája.

Az összetett függvény általános formája:

$$E(\varphi,\psi) = a_0 + \sum_{m=1}^{9} A_m e^{\left(-\left(\frac{c_\varphi\,(\varphi-\varphi_{0m})^2}{2\sigma_{\varphi m}^2} + \frac{c_\psi\,(\psi-\psi_{0m})^2}{2\sigma_{\psi m}^2}\right)\right)}$$

$$+ \sum_{m=1}^{6} (a_1\cos m\omega\varphi + b_1\cos m\omega\psi + a_2\sin m\omega\varphi + b_2\sin m\omega\psi)$$

$$+ \sum_{m=1}^{6} f_1\cos(m\omega\varphi + m\omega\psi)f_2\cos(m\omega\varphi - m\omega\psi) + f_3\cos(m\omega\varphi + m\omega\psi\, f_4\sin(m\omega\varphi - m\omega\psi)$$

$$+ f_5\sin(m\omega\varphi + m\omega\psi)f_6\cos(m\omega\varphi - m\omega\psi) + f_7\sin(m\omega\varphi + m\omega\psi)f_8\sin(m\omega\varphi - m\omega\psi)$$

[4]

ahol a_0 a sorozat konstans tagja, m a tagok sorszáma, ω a jel alapfrekvenciája, A az amplitúdó, φ_0 és ψ_0 határozzák meg a középpontját, és σ_φ és σ_ψ a φ és ψ szerinti kiterjedését az ellipszoidnak.

Mindkét szénhidrogén és az aminosavak esetén az illesztett függvények elemezhetők a matematikai analízis módszereivel, a kritikus pontok első és második deriváltak számolásával meghatározhatók. Glicin diamid és alanin diamid esetén a konformációs hálózatot elemeztük, mely felrajzolható a minimumok és az átmeneti állapotok összekötésével. Ebből megállapítható, hogy a konformációs utak száma nagymértékben nő a növekvő molekulamérettel. Ezek alapján arra jutottunk, hogy a fehérjefelcsavarodás számtalan alacsony energiájú hálózati útvonalból történő választás következménye.

Egyszerű példaként készítettünk egy esettanulmányt, hogy egy kisebb fehérje esetén, mint az opiorphin, mennyi azoknak a szerkezetek a száma, amely kellően alacsony ahhoz, hogy biológiai jelentősége is legyen. Az esettanulmány alapján az utolsó 3000 stabil szerkezet bizonyult elegendően alacsony energiájúnak (0–80 kJ/mol). Az opiorphin esetén már a kritikus pontok meghatározása is elérte a mai számítási kapacitás felső korlátait. Ezen előzetes tanulmány eredményeként elmondható, hogy egy pentapeptid esetén meg lehet határozni a globális minimumot.

Supplementary Materials

Table S1 Parameters of fitted functions (I-II molecules)

Propane					
$E_a(\varphi,\psi) = a_0 + (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi)$]					
a_0	ω				
1.32E+01	0.052				
a_1	a_2				
7.01E+00	-1.61E-04				
b_1	b_2				
7.01E+00	-1.17E-03				
Propane					
$E_a(\varphi,\psi) = a_0 + \sum_{m=1}^{6} (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi)$					
a_0	ω				
1.33E+01	0.052				
a_1	a_2	a_3	a_4	a_5	a_6
-2.95E-02	-2.12E-04	3.07E-02	7.48E-05	6.98E+00	-4.62E-04
a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
-1.73E-02	1.54E-04	2.21E-02	-3.67E-04	1.09E-01	2.34E-04
b_1	b_2	b_3	b_4	b_5	b_6
-2.88E-02	-1.43E-04	1.99E-02	1.59E-02	6.98E+00	-3.01E-04
b_7	b_8	b_9	b_{10}	b_{11}	b_{12}
-1.62E-02	1.28E-04	2.30E-02	-1.99E-04	1.10E-01	8.69E-05
Pentane					
$E_a(\varphi,\psi) = a_0 + \sum_{m=1}^{6} (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi)$					
a_0	ω				
2.23E+01	0.052				
a_1	a_2	a_3	a_4	a_5	a_6
7.98E+00	2.51E-08	3.41E+00	-7.33E-09	7.86E+00	-6.71E-09
a_7	a_8				
-3.81E-01	-5.27E-09				
b_1	b_2	b_3	b_4	b_5	b_6
7.93E+00	3.09E-05	1.73E+00	1.86E+00	7.81E+00	9.28E-05
b_7	b_8				
-3.32E-01	-1.24E-04				

Pentane

$$E(\varphi,\psi) = a_0 + \sum_{m=1}^{9} A_m e^{\left(-\left(\frac{c_\varphi (\varphi - \varphi_{0m})^2}{2\sigma_{\varphi m}^2} + \frac{c_\psi (\psi - \psi_{0m})^2}{2\sigma_{\psi m}^2}\right)\right)}$$

$$+ \sum_{m=1}^{6} (a_1 \cos m\omega\varphi + b_1 \cos m\omega\psi + a_2 \sin m\omega\varphi + b_2 \sin m\omega\psi)$$

$$+ \sum_{m=1}^{6} f_1 \cos(m\omega\varphi + m\omega\psi) f_2 \cos(m\omega\varphi - m\omega\psi) + f_3 \cos(m\omega\varphi + m\omega\psi \, f_4 \sin(m\omega\varphi - m\omega\psi)$$

$$+ f_5 \sin(m\omega\varphi + m\omega\psi) f_6 \cos(m\omega\varphi - m\omega\psi) + f_7 \sin(m\omega\varphi + m\omega\psi) f_8 \sin(m\omega\varphi - m\omega\psi)$$

a_0	c_φ	c_ψ	ω		
1.97E+01	-0.5	-0.5	0.0175		
A_1	A_2	A_3	A_4	A_5	A_6
-6.95E+02	1.05E+00	9.35E-02	4.57E-01	8.75E-01	5.17E-01
A_7	A_8	A_9			
9.43E-01	5.89E-01	7.07E+02			
φ_{01}	φ_{02}	φ_{03}	φ_{04}	φ_{05}	φ_{06}
3.76E+01	1.38E+00	5.64E-01	6.44E-01	6.48E-01	6.95E-01
φ_{07}	φ_{08}	φ_{09}			
6.37E-01	1.10E+00	3.64E+01			
ψ_{01}	ψ_{02}	ψ_{03}	ψ_{04}	ψ_{05}	ψ_{06}
1.89E+01	-1.30E+00	-1.67E+00	6.07E-01	4.50E-01	4.59E-01
ψ_{07}	ψ_{08}	ψ_{09}			
6.71E-01	8.33E-01	1.79E+01			
$\sigma_{\varphi 1}$	$\sigma_{\varphi 2}$	$\sigma_{\varphi 3}$	$\sigma_{\varphi 4}$	$\sigma_{\varphi 5}$	$\sigma_{\varphi 6}$
2.25E+01	2.78E+00	7.61E-01	2.88E-01	5.76E-01	1.04E-01
$\sigma_{\varphi 7}$	$\sigma_{\varphi 8}$	$\sigma_{\varphi 9}$			
4.72E-01	1.53E-02	5.10E+01			
$\sigma_{\psi 1}$	$\sigma_{\psi 2}$	$\sigma_{\psi 3}$	$\sigma_{\psi 4}$	$\sigma_{\psi 5}$	$\sigma_{\psi 6}$
-4.06E+01	-2.12E-01	-2.90E-01	9.11E-02	3.58E-02	7.21E-01
$\sigma_{\psi 7}$	$\sigma_{\psi 8}$	$\sigma_{\psi 9}$			
1.10E-01	5.30E-01	4.09E+01			
a_1	a_2	a_3	a_4	a_5	a_6
4.79E+00	6.32E-02	2.20E+00	1.64E-01	7.58E+00	1.10E-01
a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
-4.01E-01	-2.27E-02	-2.74E-01	3.99E-02	8.11E-02	1.34E-02
b_1	b_2	b_3	b_4	b_5	b_6
3.90E+00	6.67E-01	1.11E+00	1.43E+00	7.09E+00	5.50E-01
b_7	b_8	b_9	b_{10}	b_{11}	b_{12}
-7.44E-01	1.92E-01	-3.16E-01	1.01E-01	1.05E-01	5.29E-02

Table S2 Parameters of fitted functions (III-V molecules)

N-Acetyl-Glycine N-methylamide					
a₀	**c_φ**	**c_ψ**	**ω**		
5.92E+01	-0.5	-0.5	0.0175		
A₁	**A₂**	**A₃**	**A₄**	**A₅**	**A₆**
5.57E-01	4.40E-01	-1.33E+02	-1.64E+00	-1.45E+00	-4.48E+01
A₇	**A₈**	**A₉**			
-4.66E+01	6.71E-01	6.56E-01			
φ₀₁	**φ₀₂**	**φ₀₃**	**φ₀₄**	**φ₀₅**	**φ₀₆**
7.15E-01	3.84E-01	1.80E+00	4.49E+00	4.05E+00	4.22E+01
φ₀₇	**φ₀₈**	**φ₀₉**			
-4.33E+01	8.24E-01	6.15E-01			
ψ₀₁	**ψ₀₂**	**ψ₀₃**	**ψ₀₄**	**ψ₀₅**	**ψ₀₆**
8.25E-01	5.76E-01	-6.26E+00	1.24E+00	1.85E+00	-6.86E+01
ψ₀₇	**ψ₀₈**	**ψ₀₉**			
6.90E+01	8.06E-01	1.04E+00			
σ_φ1	**σ_φ2**	**σ_φ3**	**σ_φ4**	**σ_φ5**	**σ_φ6**
-2.99E-02	3.74E-01	-2.40E+01	4.27E+00	1.78E+00	6.65E+01
σ_φ7	**σ_φ8**	**σ_φ9**			
-6.24E+01	1.28E-01	-6.97E-02			
σ_ψ1	**σ_ψ2**	**σ_ψ3**	**σ_ψ4**	**σ_ψ5**	**σ_ψ6**
-4.67E-01	8.75E-02	-2.15E+02	1.95E-01	8.14E-02	5.74E+01
σ_ψ7	**σ_ψ8**	**σ_ψ9**			
-6.11E+01	9.57E-01	8.72E-01			
a₁	**a₂**	**a₃**	**a₄**	**a₅**	**a₆**
6.22E+01	4.31E-01	1.44E+01	-6.58E-01	2.17E+01	1.64E+00
a₇	**a₈**	**a₉**	**a₁₀**	**a₁₁**	**a₁₂**
5.06E+00	9.84E-01	4.62E+00	7.49E-01	1.48E+00	-8.66E-02
b₁	**b₂**	**b₃**	**b₄**	**b₅**	**b₆**
8.98E+00	-3.51E-01	-1.31E+00	-1.78E+00	2.82E-01	5.59E-02
b₇	**b₈**	**b₉**	**b₁₀**	**b₁₁**	**b₁₂**
3.65E-01	-7.79E-02	2.23E-01	-2.26E-02	-1.21E-01	-5.24E-02
f₁	**f₂**	**f₃**	**f₄**	**f₅**	**f₆**
-4.00E+00	-4.04E+00	8.90E-01	8.64E-01	1.64E+00	1.75E+00
f₇	**f₈**	**f₉**	**f₁₀**	**f₁₁**	**f₁₂**
-5.32E+00	5.29E+00	2.28E+00	2.32E+00	2.65E-01	9.52E-01
f₁₃	**f₁₄**	**f₁₅**	**f₁₆**	**f₁₇**	**f₁₈**
3.85E-01	5.65E-01	-2.20E+00	2.37E+00	3.40E-01	5.84E-01
f₁₉	**f₂₀**	**f₂₁**	**f₂₂**	**f₂₃**	**f₂₄**
3.85E-01	6.01E-01	4.44E-01	6.60E-01	-7.28E-02	9.37E-02
f₂₅	**f₂₆**	**f₂₇**	**f₂₈**	**f₂₉**	**f₃₀**
4.67E-01	-4.46E-01	1.41E+00	-2.78E-02	-9.72E-02	1.25E+00
f₃₁	**f₃₂**	**f₃₃**	**f₃₄**	**f₃₅**	**f₃₆**
9.70E-01	1.23E-02	-9.63E-01	-4.67E-02	-5.76E-02	1.65E+00
f₃₇	**f₃₈**	**f₃₉**	**f₄₀**	**f₄₁**	**f₄₂**
-3.63E-01	3.18E-01	-6.04E-01	2.92E-01	1.00E+00	1.04E-01
f₄₃	**f₄₄**	**f₄₅**	**f₄₆**	**f₄₇**	**f₄₈**
9.69E-01	1.30E+00	-4.22E-01	-5.36E-01	1.69E-01	-3.49E-01

N-Acetyl-Alanine N-methylamide					
a_0	c_φ	c_ψ	ω		
7.34E+01	-0.5	-0.5	0.0175		
A_1	A_2	A_3	A_4	A_5	A_6
5.49E-01	7.28E-01	5.76E-01	-3.07E+00	-1.01E+00	1.25E+03
A_7	A_8	A_9			
5.21E-01	3.72E-01	-1.33E+03			
φ_{01}	φ_{02}	φ_{03}	φ_{04}	φ_{05}	φ_{06}
9.97E-01	5.54E-01	5.35E-01	-3.32E+00	1.36E+00	2.06E+01
φ_{07}	φ_{08}	φ_{09}			
7.10E-02	8.88E-01	1.76E+01			
ψ_{01}	ψ_{02}	ψ_{03}	ψ_{04}	ψ_{05}	ψ_{06}
4.36E-01	8.27E-01	3.95E-01	-1.73E+00	1.94E+00	7.75E+01
ψ_{07}	ψ_{08}	ψ_{09}			
9.31E-01	1.91E-01	7.17E+01			
$\sigma_{\varphi 1}$	$\sigma_{\varphi 2}$	$\sigma_{\varphi 3}$	$\sigma_{\varphi 4}$	$\sigma_{\varphi 5}$	$\sigma_{\varphi 6}$
9.51E-01	5.40E-01	6.66E-02	3.35E-01	-1.89E-01	-6.05E+01
$\sigma_{\varphi 7}$	$\sigma_{\varphi 8}$	$\sigma_{\varphi 9}$			
1.34E-01	8.43E-02	-6.18E+01			
$\sigma_{\psi 1}$	$\sigma_{\psi 2}$	$\sigma_{\psi 3}$	$\sigma_{\psi 4}$	$\sigma_{\psi 5}$	$\sigma_{\psi 6}$
1.17E-02	9.54E-02	6.31E-01	1.27E+00	-9.07E-01	1.15E+02
$\sigma_{\psi 7}$	$\sigma_{\psi 8}$	$\sigma_{\psi 9}$			
6.05E-02	1.64E-01	1.19E+02			
a_1	a_2	a_3	a_4	a_5	a_6
5.25E+01	-5.95E+00	4.28E+00	-1.03E+01	3.56E+00	-5.49E+00
a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
2.73E-01	-6.23E-02	4.17E-01	4.89E-01	2.18E-01	2.34E-02
b_1	b_2	b_3	b_4	b_5	b_6
1.33E+01	-4.94E+00	-2.96E+00	-3.66E+00	9.47E-01	1.08E+00
b_7	b_8	b_9	b_{10}	b_{11}	b_{12}
-6.29E-01	8.00E-01	5.44E-01	2.13E-01	-4.52E-01	-2.82E-01
f_1	f_2	f_3	f_4	f_5	f_6
1.17E+00	1.11E+00	-2.71E+00	2.63E+00	-2.50E-01	-1.99E-01
f_7	f_8	f_9	f_{10}	f_{11}	f_{12}
-1.72E+00	1.76E+00	5.21E-01	-9.03E-02	2.09E-01	1.25E+00
f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}
-5.94E-01	5.92E-01	3.97E-01	1.00E-01	1.87E-01	-3.75E-01
f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}
1.19E-01	-1.06E+00	-5.31E-02	-3.58E-02	-6.14E-02	1.20E-01
f_{25}	f_{26}	f_{27}	f_{28}	f_{29}	f_{30}
-7.34E-01	4.29E-01	9.29E-01	-5.88E-03	-4.86E-01	5.33E-01
f_{31}	f_{32}	f_{33}	f_{34}	f_{35}	f_{36}
-4.47E-01	4.80E-01	1.19E+00	-2.69E-01	8.43E-01	2.00E-01
f_{37}	f_{38}	f_{39}	f_{40}	f_{41}	f_{42}
3.10E-01	-6.85E-01	8.48E-01	-2.69E-01	-6.92E-02	6.67E-01
f_{43}	f_{44}	f_{45}	f_{46}	f_{47}	f_{48}
1.43E+00	-2.43E+00	2.44E+00	2.43E+00	1.31E-01	5.41E-05

N-Acetyl-Valine N-methylamide					
a_0	c_φ	c_ψ	ω		
6.45E+01	-0.5	-0.5	0.0175		
A_1	A_2	A_3	A_4	A_5	A_6
-1.19E+03	1.76E+02	5.08E-01	2.55E-02	2.63E-01	1.05E+03
A_7	A_8	A_9			
9.82E+00	9.29E-01	-5.49E-02			
φ_{01}	φ_{02}	φ_{03}	φ_{04}	φ_{05}	φ_{06}
1.85E+01	3.46E+01	4.90E-01	3.42E-01	9.52E-01	1.97E+01
φ_{07}	φ_{08}	φ_{09}			
1.42E+00	7.38E-01	8.63E-01			
ψ_{01}	ψ_{02}	ψ_{03}	ψ_{04}	ψ_{05}	ψ_{06}
2.46E+01	1.04E+02	9.43E-01	5.33E-01	9.83E-01	1.79E+01
ψ_{07}	ψ_{08}	ψ_{09}			
1.07E+02	6.66E-01	2.37E+00			
$\sigma_{\varphi 1}$	$\sigma_{\varphi 2}$	$\sigma_{\varphi 3}$	$\sigma_{\varphi 4}$	$\sigma_{\varphi 5}$	$\sigma_{\varphi 6}$
2.75E+01	5.74E+01	4.32E-01	9.83E-01	1.06E-01	-6.11E+01
$\sigma_{\varphi 7}$	$\sigma_{\varphi 8}$	$\sigma_{\varphi 9}$			
1.95E-01	6.05E-02	1.40E-01			
$\sigma_{\psi 1}$	$\sigma_{\psi 2}$	$\sigma_{\psi 3}$	$\sigma_{\psi 4}$	$\sigma_{\psi 5}$	$\sigma_{\psi 6}$
7.58E+01	5.73E+01	1.55E-02	-9.47E-02	1.33E-01	6.75E+01
$\sigma_{\psi 7}$	$\sigma_{\psi 8}$	$\sigma_{\psi 9}$			
2.85E+02	3.99E-01	1.94E+00			
a_1	a_2	a_3	a_4	a_5	a_6
4.17E+01	8.04E+00	3.88E+00	5.35E+00	5.58E+00	1.02E+00
a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
3.69E-01	1.58E-01	-1.23E-01	4.11E-01	-1.27E-01	-3.16E-01
b_1	b_2	b_3	b_4	b_5	b_6
7.83E+00	9.41E+00	-9.29E+00	-9.29E-01	1.05E+00	1.52E-01
b_7	b_8	b_9	b_{10}	b_{11}	b_{12}
8.07E-01	3.30E-01	2.10E-01	-3.79E-02	-3.15E-01	-1.32E-02
f_1	f_2	f_3	f_4	f_5	f_6
1.24E+00	3.05E-01	1.48E+00	-6.24E+00	1.60E+00	2.44E+00
f_7	f_8	f_9	f_{10}	f_{11}	f_{12}
3.56E+00	3.62E+00	-6.09E-01	1.99E-01	1.24E+00	5.61E-01
f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}
-5.53E-01	4.44E-01	1.17E+00	-6.85E-01	-9.87E-01	-6.21E-02
f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}
-2.00E-01	1.33E-01	3.03E-01	7.07E-01	3.23E-01	-1.63E-01
f_{25}	f_{26}	f_{27}	f_{28}	f_{29}	f_{30}
5.21E-01	-5.42E-01	1.38E-01	-7.66E-01	3.11E-01	-6.76E-01
f_{31}	f_{32}	f_{33}	f_{34}	f_{35}	f_{36}
-7.75E-01	3.32E-01	3.87E-01	-6.11E-01	5.42E-01	4.07E-02
f_{37}	f_{38}	f_{39}	f_{40}	f_{41}	f_{42}
-2.57E-01	9.10E-01	-8.22E-02	7.12E-01	8.55E-02	1.27E+00
f_{43}	f_{44}	f_{45}	f_{46}	f_{47}	f_{48}
6.82E-01	1.45E+00	1.58E+00	2.09E+00	2.71E-01	9.68E-02

List of Publications

Anita Rágyanszki, Attila Surányi, Imre G. Csizmadia, András Kelemen, Svend J. Knak Jensen, Selma Yarligan Uysal, Béla Viskolcz. Fourier type potential energy function for conformational change of selected organic functional groups. *Chem. Phys. Lett.* **599 (2014) 169–174**

Impact Factor: 1.99

Anita Rágyanszki, Klára Z. Gerlei, Attila Surányi, Imre G. Csizmadia, András Kelemen, Svend J. Knak Jensen, Béla Viskolcz. Big Data reduction by fitting mathematical functions A search for appropriate functions to fit Ramachandran surfaces. *Chem. Phys. Lett.* **625 (2015) 91-97**

Impact Factor: 1.99

János J. Szórád, Eszter P. Faragó, **Anita Rágyanszki**, Franco A. Cimino Prado, Béla Fiser, Michael C. Owen, Balázs Jójárt, Claudio A. Morgado, Milán Szőri, Svend J. Knak Jensen, Imre G. Csizmadia, Béla Viskolcz. Conformation change of opiorphin derivates. A theoretical study of the radical initiated epimerization of opiorphin. *Chem. Phys. Lett.* **626 (2015) 29–38**

Impact Factor: 1.99

Published Papers: 3

Cumulative Impact Factor: 5.97

First Author in: 2

Conference Presentations

Oral presentations

Anita Rágyanszki, Attila Surányi, Imre G. Csizmadia, András Kelemen, Svend J. Knak Jensen, Béla Viskolcz. Analytical Functional Representation of Quantum Chemical energy curves and Surfaces. I. Innovation in Science Doctoral Student Conference. (Szeged, Hungary) May 2-3, 2014

Anita Rágyanszki, Attila Surányi, Imre G. Csizmadia, András Kelemen, Svend J. Knak Jensen, Béla Viskolcz. Functional Analyses of 2D Conformational Potential Energy Surfaces. KeMoMo–QSAR. (Szeged, Hungary) May 22-23, 2014

Anita Rágyanszki, Klára Z. Gerlei, Attila Surányi, András Kelemen, Svend J. Knak Jensen, Imre G. Csizmadia, Béla Viskolcz. Functional representation of Multidimensional Conformational Potential Energy Surfaces. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) June 17-20, 2014

Klára Z. Gerlei, **Anita Rágyanszki**, Tímea Tóth, Imre G. Csizmadia, Béla Viskolcz. 4D Analytical PEHS of L-Asparagine Diamide. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) June 17-20 2014

Scientific Posters

Eszter Pongráczné Faragó, **Anita Rágyanszki**, Franco A. Cimino P., János J. Szórád, Béla Fiser, Michael C. Owen, Balázs Jójárt, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. Radical Initiated Epimerization and Conformational Change of Opiorphin: A Theoretical Study. 3rd Visegrad Symposium on Structural Systems Biology (Smolenice Slovakia) 19– 22 June, 2013

János J. Szórád, Eszter P. Faragó, **Anita Rágyanszki**, Franco A. Cimino Prado, Béla Fiser, Michael C. Owen, Balázs Jójárt, Claudio A. Morgado, Milán Szőri, Svend J. Knak Jensen, Imre G. Csizmadia, Béla Viskolcz. A Theoretical Study of the Epimerization of Opiorphin Pentapeptide in the Presence of $OH^\bullet/H_2O_2$ Redox System. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014

Joe Bedard, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. A Computational ab initio Study on the Thermodynamic Stability of the 45 Constitutional $C_2H_4O_3$ Isomers. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014

Keyue Chen, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. A Thermodynamic Study of Glycerol and Its 191 Isomers Using Computational Ab Initio Methods. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

Zsolt Filákovics, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. A Theoretical Analysis on the Thermodynamic Stability of Various Constitutional Isomers of $C_4H_6O_2$. A case study of tetrahydrofuran derivatives. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

Kyungseop Lee, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. Theoretical stability assessment of isomers of C3H4O2 with thermodynamic properties. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

Li Jian Bin, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. An ab initio study on the thermodynamic properties of $C_3H_6O_3$ isomers and its potential in synthetic and metabolic processes. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

Ying Lu, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. A Theoretical Analysis on the Thermodynamic Stability of Various Constitutional Isomers of $C_3H_2O_3$. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

Hyemin Park, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. Ab initio Study on the Thermodynamic Stabilities of Isomers for $C_2H_6O_3$. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

Chaitnandanie Ramkissoon, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. Screening Constitutional Isomers of $C_3H_2O_2$ for thermodynamic properties obtained from Systematic Ab initio calculations. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

Zsófia Borbála Rózsa, Milán Szőri, János J. Szórád, Imre G. Csizmadia, Béla Viskolcz, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, Michael C. Owen, Béla Fiser. A theoretical analysis on the thermodynamic stability of various constitutional isomers of $C_4H_8O_2$. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

Dhiraj Sinha, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G.Csizmadia, Rudiger Ettrich, Béla Viskolcz. An ab initio Study of Thermochemical Properties of $C_4H_4O_2$ Isomers. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

Daniel Y. Wang, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. Systematic Ab initio Study of all Constitutional Isomers of $C_3H_8O_2$ Including Molecular Complexes. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

XuanTing Xu, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. Ab initio study of the thermodynamic stability of $C_3H_6O_2$ isomers. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

Peiqi Yu, Gabriella Kaszap, Klára Z. Gerlei, **Anita Rágyanszki**, János J. Szórád, Béla Fiser, Michael C. Owen, Milán Szőri, Imre G. Csizmadia, Béla Viskolcz. Ab initio study of the thermodynamic stability of $C_3H_4O_3$ isomers. 4th Visegrad Symposium on Structural Systems Biology (Nové Hrady Czech Republic) 17– 20 June, 2014.

List of Tables

List of Figures