We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Quantum Chemistry and Chemometrics Applied to Conformational Analysis

Aline Thaís Bruni[1] and Vitor Barbanti Pereira Leite[2]
[1]Departamento de Química, Faculdade de Filosofia,
Ciências e Letras de Ribeirão Preto, Universidade de São Paulo
[2]Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas,
Universidade Estadual Paulista, São José do Rio Preto
Brazil

1. Introduction

1.1 Conformational analysis: Early history and Importance

Molecular structure plays a special role in science. Knowledge of the atomic arrangement is essential in order to be able to elucidate chemical properties and processes. The first advances in determining molecular structure occurred in nineteenth century. Around 1812, Jean-Baptiste Biot, a French physicist, discovered optical activity by observing polarized light shifting when crossing a quartz crystal. He observed that the light was displaced to the right in some cases and to the left in others. The conclusion was that rotation of polarized light by quartz is an inherent property of the crystal. Interested in the phenomenon, Biot noticed in further studies that similar effects were found when polarized light passed through certain liquids such as natural oils (lemon extract and laurel), alcoholic solutions of camphor, some sugars and tartaric acid. (Drayer, 1993; Cintas, 2007; Gal, 2011) Biot's observations were very important in laying foundation for the concept of optical activity.

In 1948, Louis Pasteur discovered molecular chirality when studying a mixture of tartaric acid crystals.(Gal, 2007) He patiently performed the manual separation of tartarate enantiomer crystals (Cintas, 2007) and observed that each solution made with them was able to displace polarized light in one direction. He concluded that compounds with non-superimposable molecular asymmetry have identical chemical properties despite the inverse behavior related to polarized light. Pasteur argued that the optical activity of organic solutions is related to molecular geometry. This insight was far ahead of the organic structural theory of the time.(Drayer, 1993) Although Pasteur was the first to show a relationship between optical activity and molecular symmetry, he was not able to say exactly how a molecule could be right- or left-handed. The main advances in this idea occurred in 1874 when a theory of organic structure in three dimensions was independently and simultaneously developed by Jacobus Henricus van't Hoff in Holland, and Joseph Achille Le Bel in France. (Drayer, 1993; Cintas, 2007)

In 1865, August Kekulé proposed his theory of the benzene molecular structure and proposed that the carbon atom has valence 4.(Brush, 1999) His principal idea was that the carbon atom is

tetravalent and can form valence bonds with other carbon atoms yielding to chains. These carbon chains can sometimes have closed arrangements, forming rings. (Drayer, 1993)

Van't Hoff and Le Bel proposed that the four valences of the carbon atom were not planar, but directed into three-dimensional space. Van't Hoff specifically proposed that the spatial arrangement was tetrahedral. Later, he used the tetrahedron as a graphic representation of the valence arrangement around the carbon atom and also used this model to explain the physical property of optical activity.(Ramberg & Somsen, 2001) A compound containing a four different substituted carbon – described by Van't Hoff as asymmetric carbon - would be capable of existing in two distinctly different nonsuperimposable forms. Finally, he stated that the asymmetric carbon atom was the cause of molecular asymmetry and optical activity.(Drayer, 1993)

Le Bel, in turn, also published his stereochemical ideas in 1874, but with a different approach to the problem from that presented by Van't Hoff. His hypothesis was not based on the tetrahedral model for the carbon atom and the fixed valences between the atoms. His investigation was into the asymmetry as a whole, without evaluating the individual atoms. The full system was considered in his evaluation, and his interpretation could be inserted into the field that is currently understood as molecular asymmetry. He mentions the tetrahedral carbon atom only in special cases, and not as a general principle. Many molecules confirm Le Bel's concepts of molecular asymmetry. Allenes, spiranes, and biphenyls are some examples of asymmetric molecules that do not contain any asymmetric carbons.

Van't Hoff's and Le Bel's different approaches can be explained by the origin of their formation. Van't Hoff, based on Kekulé tetrahedron models, suggested the concept of the asymmetric carbon atom. On the other hand, Le Bel based his investigations on Pasteur's considerations of the connections between optical rotation and molecular structure.(Drayer, 1993)

The historical development of conformational search does not end here and has many other important aspects and particularities. Our goal was just to give a basic outline of the initial concepts and how they influence current conformational understanding. Despite the historical progress in conformational studies, the advances in structure determination has been relatively recent and have been made possible by the development of analytical instruments and computational tools. Early structural studies were applied only to small molecules or substructures that could be expressed in terms of a few settings.(Allen et al., 2010)

Currently, a great evolution is occurring in mechanisms for determining and understanding molecular structures. The relationship between geometry and energy is experimentally measurable and gives an idea of the balance between energy factors involved in each structure. (Pietropaolo et al., 2011) Reactivity and other properties are directly linked to the conformational arrangement of molecules.(Hunger & Huttner, 1999) Every chemical property must be understood according to its molecular structure and atomic connections. (Pietropaolo et al., 2011) Indeed, knowledge of structural arrangement is important since it underlies studies in chemical reactions and other molecular behaviors. There are experimental techniques for the structural determination, such as X-ray, magnetic resonance, infrared, mass spectroscopy and others.

In this chapter we will discuss theoretical methods for molecular conformational determination. The field that concerns ways to mimic the behavior of molecules and molecular systems is molecular modeling. It seeks a simplified or idealized description of

molecular systems, making it possible to produce three-dimensional representations that provide insights into their behavior. As computer tools have enjoyed a spectacular increase in last decades, theoretical methods are invariably associated with computer modeling. This has become a powerful tool for evaluating molecular structure, from which special chemical information about molecular behavior can be inferred. (Pietropaolo et al., 2011)

2. Statement of the problem

For the theoretical and computational determination of molecular properties it is necessary to previously determine the minimum energy structure of the system being studied. A central issue is to probe the equilibrium configuration of the molecular system. The way that energy varies with the coordinates is usually referred to as the potential energy surface. At the atomic level, the interaction energy between atoms is essentially ruled by quantum mechanics, which provides the basic elements and methods used in molecular modeling. However, the potential energy surface can be addressed with different degrees of approximation, *i.e.*, *ab initio*, effective potentials or even more coarse-grained potentials. Irrespective of the details with which the system is considered, one usually faces the problem of a highly dimensional system with the occurrence of multiple minima. Low energy minima play an important role in determining molecular properties, and the determination of these minima conformational states is a non-trivial task, usually referred to as energy minimization method for exploring the energy surface.

If four or more atoms are connected in chain by single bonds we can suppose that there is considerable flexibility in the molecule. The existence of hindered rotation about a single bond is one of the fundamental concepts in conformational analysis.(Mo & Gao, 2007) The understanding of the connections between the atoms is related to the internal coordinate parameters, *i.e.*, bond length, bond angle and dihedral angle, and is essential in designing molecular models.
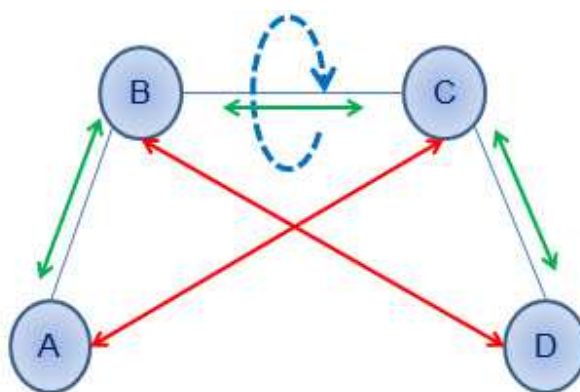

Fig. 1. Model for a generic molecule with four different atoms.

For instance, let us consider a molecule composed of four different atoms which are single-bond linked (Figure 1). The green arrows represent the bond stretch and the average value is the bond length; the red arrows correspond to the angle formed by three sequential atoms, *i.e.*, the angle bond; the curved blue arrow indicates the free rotation around the only single bond able to perform changes in the molecule conformation, as shown in Figure 2:
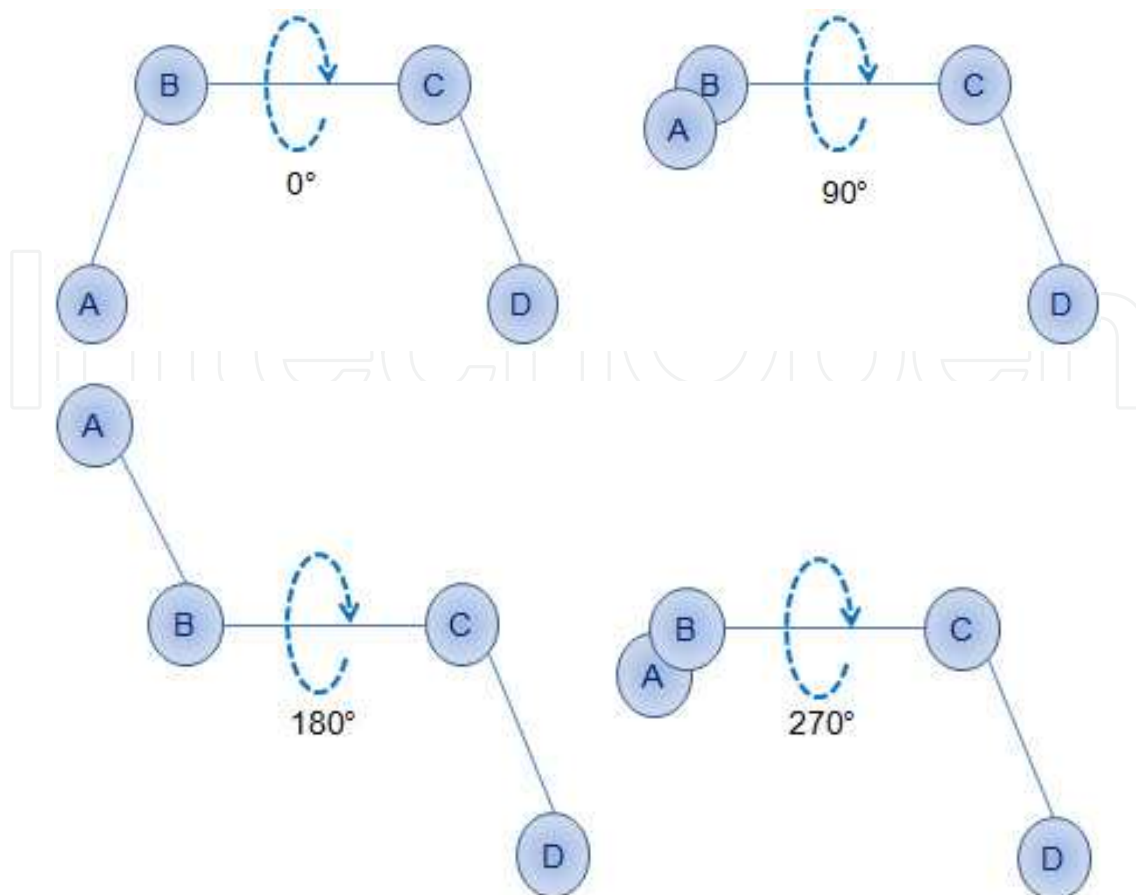
Fig. 2. 90° rotation around the single bond.

In other words, different conformations are obtained when a dihedral angle is rotated. A dihedral angle is that composed by the planes formed by the sequence of three atoms (Figure 3):
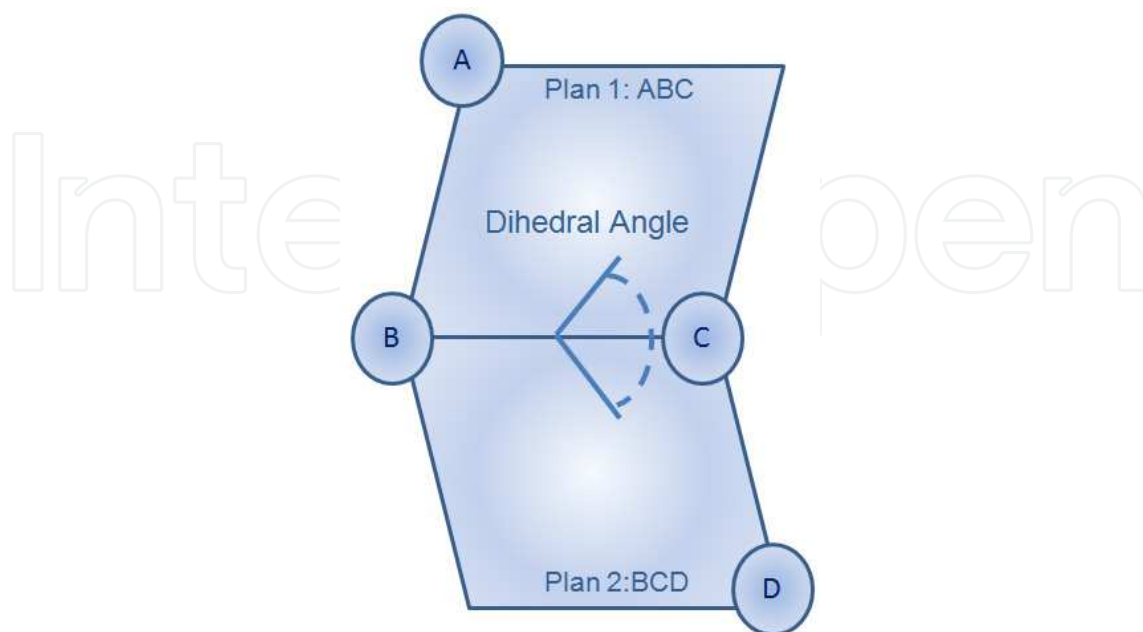

Fig. 3. Dihedral angle representation for the molecule ABCD.

The main questions on molecular modeling are concerned with a good way of finding the global minimum energy structure. Important information is also concerned with the behavior of the conformational space. It is not only necessary to know which is the global minimum, but also the whole shape of the potential energy surface (PES). The main characteristic of this conformational phase space is that it is exponentially large, and a computationally hard problem. (Fraenkel, 1993) One problem that illustrates this difficulty is the protein folding, in which one searches for the global energy minimum structure associated with its functional conformation. If a method can be used to describe the relevant potential energy surface of a given molecule, it can also accurately elucidate its behavior against many interest situations.

Many techniques have been presented, and it is not the goal of this work to make a deep study on them. A good discussion of practical methods is given by Leach.(Leach, 2001) We intend to give a brief idea of the most popular techniques used for investigating the molecular structure computationally. For conformational sampling, one can imagine a hierarchy of methods with different computational costs.(Seabra et al., 2009) However, there is no sovereign truth about what is the best method for performing a conformational analysis. Each situation must be evaluated. The best method is one that has the best fit to the problem studied; in practical terms it will provide the answer as quickly as possible, using the least amount of computer resources.

3. Stochastic methods of conformational analysis

The literature reports many methods for trying to solve multiconformational problems, and most of them are based on stochastic approaches. Put simply, stochastic methods work with random variables, such as initial conformations for the search or the steps probing the configuration phase space. The simple criterion for establishing a minimum energy conformation is that the first derivatives of the energy E with respect to each variable (x_i) is zero and the second derivatives are all positive:

$$\frac{\partial E}{\partial x_i} = 0 \quad \text{and} \quad \frac{\partial^2 E}{\partial x_i^2} > 0.$$

The algorithms that search for minimum energy states can be classified into two groups: those which use derivatives of the energy with respect to the coordinates, and those which do not. The most used derivative minimization methods are the steepest descent, line search in one dimension and conjugate gradient methods.(Leach, 2001) These algorithms are very useful for conducting local (restricted) searches of minima, or downhill searches to the nearest minimum, since they are not able to overcome energy barriers. They are often used in combination with other stochastic methods.

In the remainder of this section we discuss examples of stochastic methods. Havel, Kuntz and Crippen described distance geometry algorithms in conformational analysis.(Havel et al., 1983b) Given the impossibility of examining all possible conformations, they introduced a method which is capable of finding global optima without considering all possible solutions by means of combinatorial optimization. The method is known as *brunch and bound* and involves logical tests that allow whole classes of solutions to be eliminated without

examining them one by one. The method converts a set of distance ranges (or "bounds") into a set of Cartesian coordinates that are consistent with those bounds. (Spellmeyer et al., 1997) The efficiency of a *branch and bound* algorithm depends on how effective these tests are compared to the time required to perform them. (Havel et al., 1983a; Havel et al., 1983b) In another study, Havel *et al* presented the basic theorems of distance geometry in Euclidean space. They proposed new algorithms and described refinements to the existing ones. All these algorithms were similar because they utilize geometric principles in order to interpret structural relationships. (Havel et al., 1983b)

According to Leach and Smellie,(Leach & Smellie, 1992) distance geometry is a method for searching conformational space in which a structure is initially formulated in terms of interatomic distances. Any molecular system can be described as the set of minimum and maximum interatomic distances between all pairs of atoms in the molecule. The complete conformational space of the molecule is contained within this space. In distance geometry, a matrix is defined as the set of minimum and maximum distances, and then used to create a series of conformers that are consistent with those distances.(Spellmeyer et al., 1997)

Another tool for performing conformational searches is the genetic algorithm, a stochastic method first introduced by Holland in 1975. Genetic algorithm (GA) is a method applied to solve problems using a natural evolution process simulation. It is a stochastic method developed in analogy to Darwin's theory of evolution in order to perform the optimization.(Brodmeier & Pretsch, 1994; Lucasius, 1993; Nair & Goodman, 1998)

Genetic algorithm is commonly used for studying a large-scale space of possible solutions. The goal is to identify the best solutions within that space without the need to evaluate all possibilities.(Yanmaz et al., 2011) The GA is the optimization of a large number of possible solutions using a randomly generated population. When applied to conformational analysis, the population of interest consists of different conformations. The biological evolution of this population is simulated. A population of trial solutions is iteratively manipulated by a series of genetic operators to satisfy an objective function. The adjustment is calculated, and a new population is generated according to operators, such as selective reproduction, recombination and mutation. The process is repeated until the minimum energy structures are obtained.(Lucasius & Kateman, 1994; Beckers et al., 1996; Beckers et al., 1997)

Artificial Neural Networks (ANN) are another example of stochastic methods used in conformational analysis. This method is based on concepts of the behavior of the human brain. Although artificial neural networks are primitive compared to their biological counterparts, they exhibit some interesting properties which make them useful as multivariate tools in various fields of research. During the last decade, ANN have been successfully applied in non-linear modeling, classification, signal processing and process control.(Derks & Buydens, 1996) The properties of a molecule are intimately linked to the conformations that it adopts and so an understanding of the conformational space is important in rationalizing and predicting its behavior.(Jordan et al., 1995)

Among the most popular stochastic methods for covering the conformational space are Monte Carlo (MC) and Molecular Dynamics (MD). They are similar in the sense that both procedures include the same representation of molecules and use classical force fields for the potential energy terms, under periodic boundary conditions. The main purpose of these methods is to sample the phase space and to use the force fields ability to represent the conformational space

near minima and connecting transition structures.(Jorgensen & TiradoRives, 1996; Grouleff & Jensen, 2011) However, large differences are found in sampling and configuring space available to the system. For MC, a new configuration is generated by selecting a random molecule or part of it, rotating it, translating it, and performing an internal structural variation. These changes do not necessarily need to follow a realistic physical trajectory. The acceptance of the new configuration is, however, determined by the Metropolis sampling algorithm. The sampling criterion is set in a way that enhances the likelihood of probing low energy conformations. Application over enough configurations yields properly Boltzmann-weighted averages for structure and thermodynamic properties. For MD, given a set of initial conditions (position and velocities of all atoms), new configurations are generated by application of Newton's equations of motion, so that the new atomic positions and velocities of all atoms are determined simultaneously over a small time step. In both cases, the force field controls the total energy (MC) and forces (MD), which determines the evolution of the systems. (Jorgensen & TiradoRives, 1996)

Examples of problems related to large systems are the interaction between drug and the receptor, and protein behavior and folding. Molecular docking procedures are capable of predicting the three-dimensional structure of macromolecular complexes and their binding affinity. The information required is simple and corresponds to the structures of the receptor and ligand and the presumable interfacing region between them. Besides the simplicity of these docking procedures, they have low computational costs. However, molecular plasticity and solvation effects are not, or are only approximately, taken into account in these approaches. Free energy simulations may be then used to investigate the molecular association process and to predict binding affinity. (Biarnés et al., 2011)

It is important to realize that sometimes the probing of PES addresses singular questions, which involve association of several methods, also called hybrid methods. A particular well-known tailored one is the quantum mechanics/molecular dynamics approach, also known as QM/MM approach. This is a molecular simulation method that combines the strength of both QM (high accuracy in specific regions) and MD fast calculations (in not so crucial regions), in such a way that it efficiently allows the study of chemical processes in solution and in proteins.

When stochastic methods are used to find minimum energy conformations, asymptotic states in restricted regions of the phase space are probed. This means that there is no end point in the search, and the convergence cannot be assured.

4. Systematic search in conformational analysis

As seen before, stochastic techniques use different heuristics to randomly cover the conformational space. These algorithms apply a perturbation to the initial conformer and minimum energy conformation is associated with the lowest energy state that is found through out this procedure. They provide a sampling of energy minima structures and the shape of the PES is obtained in an indirect way.

Beyond the stochastic methods there are procedures that do not work with random choice to cover conformational space. These classes of methods are described as deterministic and are capable of searching the conformational map in a systematic way, providing a direct

knowledge of PES shape. These searches divide conformational space into quantized units and apply algorithms to search this discrete space or define a set of heuristic rules that are used to drive the search.(Smellie et al., 2003)

Systematic methods are those that explore all conformational space at some fixed degree of resolution. To perform the systematic search, a molecule must be numerically described by its atoms' internal coordinates. The internal coordinates are bond length, angle bond and dihedral (torsion) angle. For a given initial structure the systematic conformational search is conducted by regular variation in dihedral angles (Figure 2).

Although a systematic search can obtain the morphology of a molecule's energetic behavior directly, this method is not feasible for evaluating complex systems. (Beusen et al., 1996) Systematic search is most usefully applied for molecules with few degrees of freedom.(Li Manni et al., 2009)

According to literature (Beusen et al., 1996), to cover the PES corresponding to the conformational space, different molecular structures must be systematically generated by rotating the torsion angles around the single bonds between 0° and 360°. The number of conformations is given by:

$$\text{Number of conformations} = s^N \tag{1}$$

where N is the number of free rotation angles, and s is the number of defining steps according to the angle increment:

$$s = \frac{360°}{\theta_i} \tag{2}$$

with θi being the dihedral increment of angle i.

An examination of equation (1) reveals that the number of conformations generated will exponentially increase in proportion to the number of bonds with free rotation in the molecule under study. A problem arises if the number of steps is large, *i.e.*, when a very refined surface is required by small angle increments. This problematic behavior of the systematic study of PES, described as combinatorial explosion, is the major restriction involved in this kind of search. Figures 4 and 5 illustrate how combinatorial explosion works. In Figure 4, we have a representation of the system growth where many single bonds can be rotated.

The combinatorial explosion problem is represented by Figure 5. The number of branches to be considered is shown by the ramification achieved according the number of angles (A, B, C, D…) and will depend on the dihedral increment chosen. Due to the problem involved in combinatorial explosion, systematic search becomes nonviable for studying large molecules, since the number of degrees of freedom increases. A useful strategy for reducing the dimensionality of the conformational space is to perform systematic conformational searches on small portions of the molecule (either as isolated fragments or *in situ*). Using these optimal parts, one builds the conformation of the whole molecule with only limited additional searching of the relative conformations of the fragments. Approaches that incorporate this principle are known as ''build-up'' methods. (Beusen et al., 1996; Izgorodina et al., 2007) There are some strategies for overcoming the combinatorial explosion. We will focus our discussion on procedures that involve chemometrical approaches.
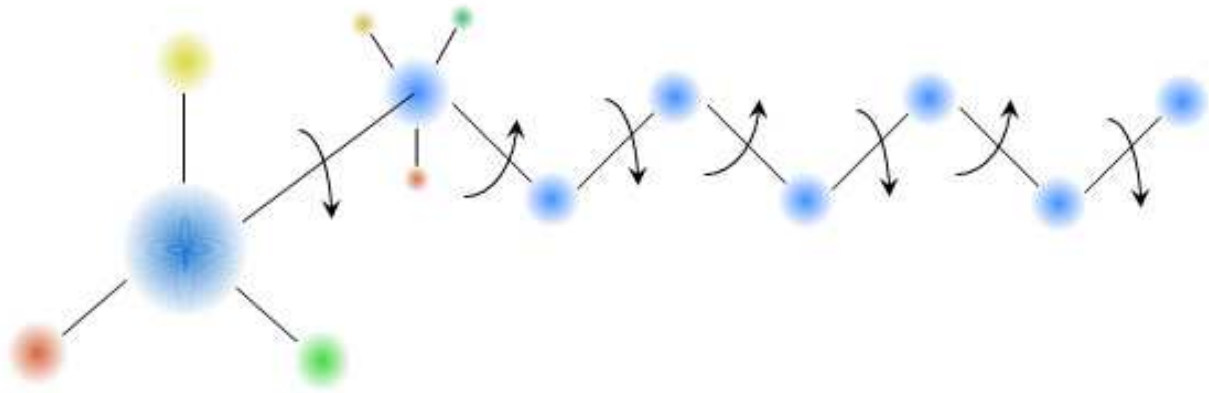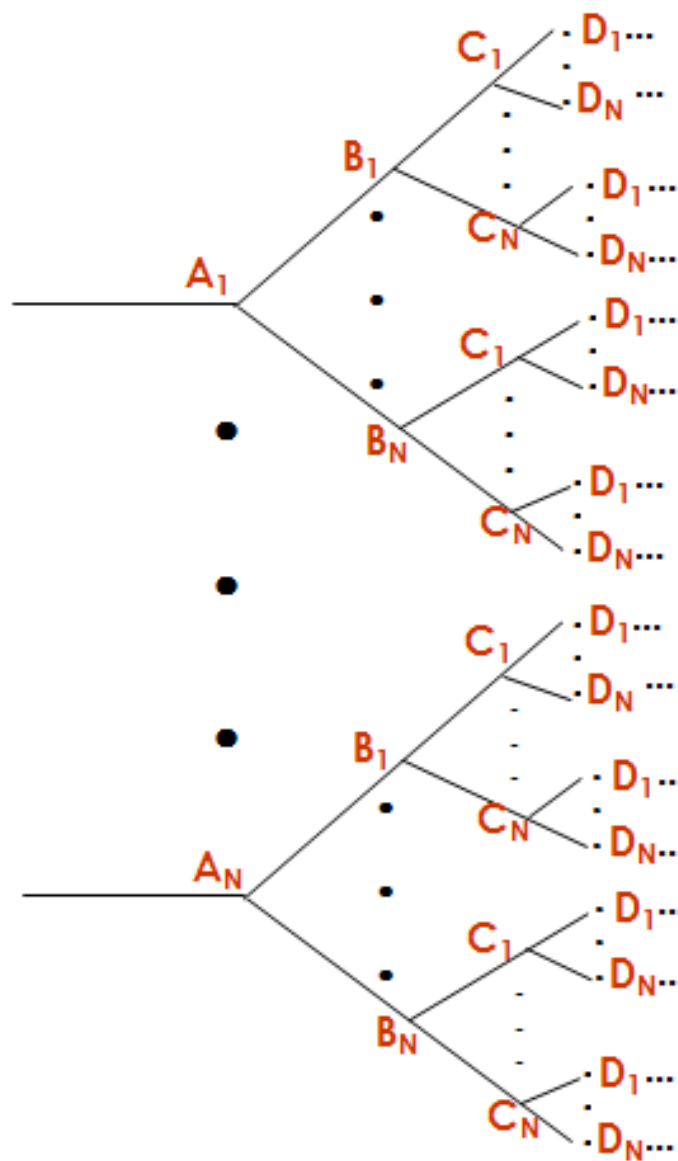
Fig. 4. A general structure with many single bonds.

Fig. 5. Branches generated by the dihedral angles A, B, C and D.

5. Chemometrics and structure determination

A conformational search, independently of the method chosen, usually involves large amounts of data. Sometimes, data achieved from a given methodology must be explored by an additional technique. According to Geladi (2003) *"data exploration means taking a look at the data to find interesting phenomena, often without prior expectations. As a result, outliers, clustering of objects and gradients between clusters may be detected."*(Geladi, 2003)

Chemometrics has been used extensively in recent years for exploring chemical problems by means of computer tools and statistical observations. The literature presents many definitions for chemometrics. For our purposes, this field of knowledge is better defined as a combination of two definitions found in the literature:

a. According to Wold (1995), chemometrics can be understood as a way *"to get chemically relevant information out of measured chemical data, how to represent and display this information, and how to get such information into data"*;(Wold, 1995)
b. For Beeb (1998), chemometrics corresponds to *"the entire process whereby data (e.g., numbers in a table) are transformed into information used for decision making."* (Beebe, 1998)

The above definitions indicate that chemometrics offers a broad approach to chemical measurement sciences. It is not restricted to the actual experimental analysis but also considers what happens before and after it. (Massart et al., 2004) It is the goal of chemometrics to extract the information from the data. (Ramos et al., 1986) Chemometrical approaches have been applied to conformational analysis for handling special difficulties of large amounts of data generated both by stochastic and by systematic searches.

Among the various chemometrical techniques, Principal Component Analysis (PCA) is the most commonly used for conformational problems. In many ways, it forms the basis for multivariate data analysis**.** PCA is a multivariate method of analysis whose main concern is to reduce the dimensions needed to portray accurately the characteristics of a large dimensional data matrix.(Beebe, 1998; Wold et al., 1987) This mathematical procedure consists of eliminating a large number of correlated variables without changing the characteristics of the original data-set that contribute most to its variance.

For an easy graphical representation, consider a two-dimensional set of variables as shown in Figure 6 (a).

PCA can be performed on the original variables as shown in Figure 6 (b) and new axes, called Principal Components, arise to account for the maximum variation. A subsequent rotation (Figure 6(c)) is made on these new PC axes in order to rewrite the original variables in terms of this new axes-system. Each PC is constructed as a linear combination of variables:

$$P_i = \sum_{j=1}^{v} C_{i,j}x_j \qquad (3)$$

where P_i is the ith principal component and $c_{i,j}$ is the coefficient of the variable $x_{i,j}$. (Leach, 2001) There are v such variables. The first principal component PC1 is chosen in order to maximize the data variance of the axis. The second and subsequent ones are chosen to be orthogonal to each other and account for the maximum variance in the data not yet

described by previous principal components. A variety of algorithms can be used to calculate the principal components. The most commonly employed approach is singular value decomposition SVD. (Golub & Loan, 1996)
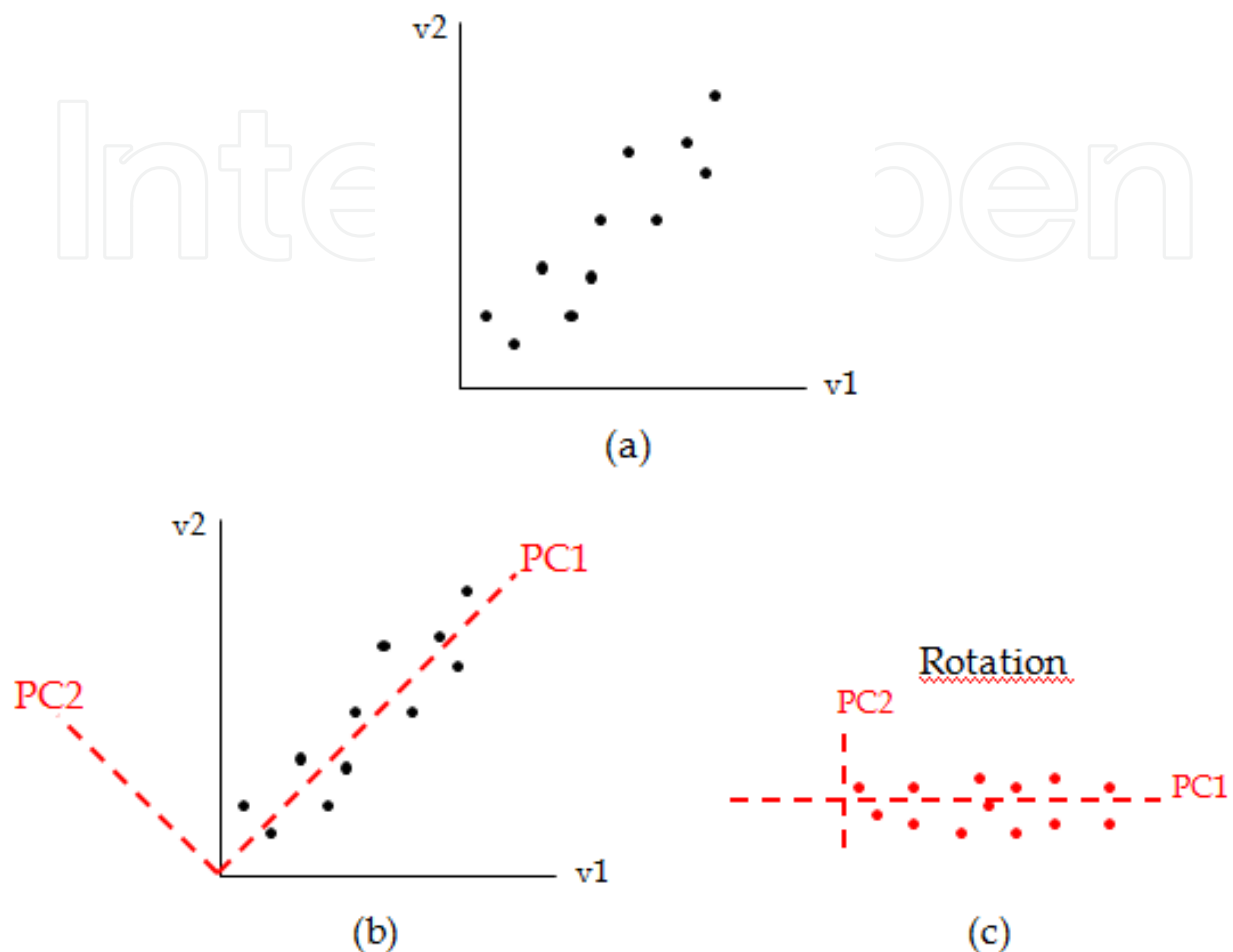

Fig. 6. PCA procedure: (a) original data set; (b) PCA on original data set and (c) Variables according to the new PC coordinates.(Beebe, 1998)

A matrix of arbitrary size can be decomposed into the product of three matrices in such a way that:

$$\mathbf{X} = \mathbf{USV}_t \qquad (4)$$

where $\mathbf{U}$ and $\mathbf{V}$ are square orthogonal matrices. The matrix $\mathbf{U}$ (whose columns are the eigenvectors of $\mathbf{XX}_t$) contains the coordinates of samples along the PC axes. The $\mathbf{V}$ matrix (which contains the eigenvectors of the correlation matrix $\mathbf{X}_t\mathbf{X}$) contains the information about how the original variables were used to make the new axes[c_i,j coefficients in eq. (3)]. The $\mathbf{S}$ matrix is a diagonal matrix that contains the eigenvalues of the correlation matrix (standard deviations) or singular values of each of the new PCs. The diagonalization of symmetric matrices (such as $\mathbf{XX}_t$ and $\mathbf{X}_t\mathbf{X}$) and SVD are fundamental problems in linear algebra (Golub & Loan, 1996), for which computationally efficient software has been developed and can be used on a routine basis (Hanselma et al., 1997) for very large-size matrices.

In chemistry, PCA was introduced by Malinowski around 1960 under the name Principal Factor Analysis, and further developed after 1970.(Malinowski, 2003) Principal Component Analysis can be used for crystallographic structure data; in its general form, conformational analysis is applied to multivariate numerical problems. (Allen et al., 2010) Many studies report on the use of PCA for handling Molecular Dynamics data. Among them, we highlight the application that uses PCA for mapping potential energy surfaces, by the quantitative visualization of a macromolecular energy funnel. (Becker, 1998) Other examples where PCA can be applied in molecular structure determination can also be found in recent studies. (Das et al., 2011; Araujo-Andrade et al., 2010; Kiralj et al. 2007; Oblinsky et al., 2009; Silva et al., 2011)

6. Pairs of dihedral angles-systematic analysis

There is a variety of theoretical methods that are capable of locating minimum energy structures in the potential energy surface. The problem of stochastic methods is that there is no natural end point for the conformational search. In some cases, only a small subset of conformational space is explored and the convergence of the system is not guaranteed. Only Systematic Conformational Analysis maps the conformational space completely. We stress the principal difficulty inherent in this method is the combinatorial explosion. In a previous study (Bruni et al., 2002), a new methodology was introduced for controlling the combinatorial explosion through a systematic reduction in the size of the system by means of chemometrics.

This method consists of a small systematic conformational analysis, in which the conformational space is studied by rotation of the important free rotation in pairs, described as Pairs of Dihedral Angles-Systematic Analysis – PDA-SA. The main objective is to reduce the dimension of the investigated system. The idea is to address the conformational space in small portions, evaluating PES in combinations of angles in pairs. If the problem of combinatorial explosion is controlled, the conformational space can be sufficiently refined in the regions of minimum energy, taking care to minimize the information lost. The energy surfaces are obtained for each pair of angles and the number of conformations is given by Equation 5:

$$\text{Number of conformations} = s^2 \frac{N(N\text{-}1)}{2} \qquad (5)$$

where s and N have the same meaning as in Equation 1. The number of conformations, in this case, is given by the combinatory arrangement of the N dihedrals in pairs.

The main observation of the comparison between equations (1) and (5) is that the number of conformations as given by Eq. (1) increases exponentially with the number of bonds with free rotation, while from Eq. (5), the number of studied conformations increases quadratically with *N*. As the number of free rotation angles increases, the difference in the number of conformers generated by these two equations becomes more evident.

The computational procedure for PDA-SA can be organized in five basic steps:

1. **Molecular Building:** The interest molecule must be defined in terms of its internal coordinates: bond length, angle bond and dihedral angles. There are many softwares able to define this molecular initial structure. A quantum chemistry optimization is required at this step in order to adjust internal parameters. The best method must be chosen according to the system under study.

2. **Dihedral Pair Rotation:** The PDA-SA conformational search begins and the combination of the existing pairs of angles is taking account. Sometimes it is only possible to choose a dihedral increment with a less refined value. A rough PES is obtained in this case. The matrix to be analyzed consists of energy values from potential surfaces for angle combinations, and they are grouped according to Figure 7 for N angles. Appendix A shows the energy values for omprazole basic structure. The idea is to perform a cyclical permutation on the data, and this matrix form ensures that no information about the total PES is lost. The energy values obtained for each angle rotation as a function of the others allow the conformational space to be completely mapped. The major advantage is that the shape of these small portions can be visually observed, since we have a 3-D fitting. (see Figures 9 and 10)

3. **PCA application on data matrix:** After the energy matrix statement, PCA is performed on the data. The regions with minimum energy points on the grid search can be easily selected. The number of selected regions will depend on the nature of the studied system.

4. **Refinement with a short dihedral increment:** The regions initially obtained in step 3 can be refined with a small angle increment. It is important to emphasize that this step is not obligatory, since a small dihedral increment can be used in step 3, depending on the studied system. However, previous experience in this methodology (Bruni et al., 2002; Bruni & Ferreira, 2008) shows that this is the easiest procedure, *i.e.*, firstly make rotations with a large dihedral increment and subsequently refine the minimum energy regions selected by PCA with small dihedral increments.

5. **Optimization of the final structure:** the procedure described above provides angle values for the conformational search with a good level of accuracy. When these values are combined, we obtain all the possible minimum structures. Those structures constrained by the angle values obtained by PCA analysis are submitted to final optimization and the resulting structures are considered to be those of minimum energy.

In the study that introduced this method, the approach was successfully tested in the analysis of omeprazole and its derivatives, in which the results were in agreement with the experimental ones.(Bruni et al., 2002)

In a second study, the technique was used to find minimum energy conformations of omeprazole derivative molecules in a QSAR study. (Bruni & Ferreira, 2002) It was shown that conformational analysis is crucial when establishing SAR/QSAR models using theoretically calculated descriptors, and they are strongly dependent on the details of molecular structure. Though all minima conformation have similar energetic values, some calculated properties are very sensitive to the structural variation, which is understandable since electronic properties are intrinsically dependent on molecular conformation.(Bruni & Ferreira, 2002)

Omeprazole's racemization barrier and decomposition reaction was also studied. Quantum chemistry coupled to PDA-SA chemometric method was used to find all omeprazole minimum energy structures. To obtain the racemization barriers it was essential that the starting structure was in a global energy minimum. In that work, for all the studied structures, there was no change in the values of the racemization barriers, which confirmed the identification of the most stable structures for omeprazole.(Bruni & Ferreira, 2008)

$$
\begin{array}{c|ccccc}
i\backslash j & 1 & 2 & 3 & \ldots & N \\
\hline
1 & - & E_{12} & E_{13} & \ldots & E_{1N} \\
2 & E_{21} & - & E_{23} & \ldots & E_{2N} \\
3 & E_{31} & E_{32} & - & \ldots & E_{3N} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
N & E_{N1} & E_{N2} & E_{N3} & \ldots & -
\end{array}
$$

Fig. 7. Matrix scheme for N angles: the discrete energy values for each rotation angle must be evaluated. E_{ij} are the energy matrices with elements $E_{ij}{}^{km}$, in which k and m are the angle increment indices for the angles i and j, respectively.

This approach is straightforward and in principle would have no size limits for its application. However, it presents limitation due to some initial condition dependence. Given a system with N degrees of freedom, for each pair of angles there are $N-2$ parameters that can interfere in the method. For example, in Figure 4 the potential energy surface for first and last dihedral angle combinations depends on the dihedral angles conformation between them. When the dihedral angles are too far from each other along the chain of atoms, the method may not become feasible. In this case the method may need to be repeated with different initial conditions to improve the sampling of the configurational phase space, and moreover we cannot be sure that we have reached the global minimum. When the correlations between the pairs of angles do not depend strongly on these initial conditions the method is very useful. Such system corresponds to small molecules, not so flexible, in which there are few large potential basins, such as omeprazol and its derivatives.(Bruni et al., 2002; Bruni & Ferreira, 2002; Bruni & Ferreira, 2008)

The limit of validity for this method is under investigation. We are applying this method to study the IAN peptide, which is a tetrapeptide isobutyryl-(ala)3-NH-methyl. (Nascimento et al., 2009) This is the smallest polypeptide that can have secondary-like structure (an helix) (Becker & Karplus, 1997) and it has 11 free rotation bonds. For a flexible system, such as this, the initial condition dependence in the calculation of the minimum energy conformations is expected to increase with the size of the system. Since the system is more flexible and expected to be more rugged, we partially overcome this problem by using small angle increments steps, in order to probe all local minima of the system and compare them.

7. Numerical results

7.1 Study of basic structure for omeprazole and derivatives

Initially, the basic structure of omeprazole and derivatives was evaluated. This structure has three bonds with free rotation. To validate the proposed methodology, two different approaches were performed. In the first approach, pairs of angles were taken account ((1,2), (1,3) e (2,3) in Fig. 8) and the number of conformations is given according to Equation 5. The resulting matrix analyzed was composed by the energy values from the potential energy surface for each angle combination (see matrix example in Figure 7). A matrix with discrete energy values for the basic structures with 30° angle increment in Equation 2 is showed in Appendix A.
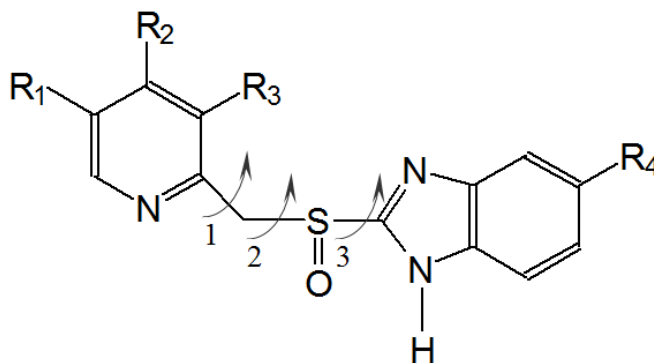
Fig. 8. Basic structure for omeprazole and derivatives.

Three PES were obtained for a 30° dihedral increment and are showed in Figures 9 and 10. Figure 9 shows the original energy values and Figure 10 shows the same surfaces, but with a 0,12 hartrees cut off for better visualization. PCA was performed on autoscaled original data and the results are shown in Figure 11. 64% of the whole information is cumulated in first and second Factors (or Principal Components-PCs). The convergence of the points for one region is observed. Figure 12 shows the PCA results for the leveled data in 0,12 hartrees. Factor 1 and Factor 2 now cumulate 73% of the entire information.
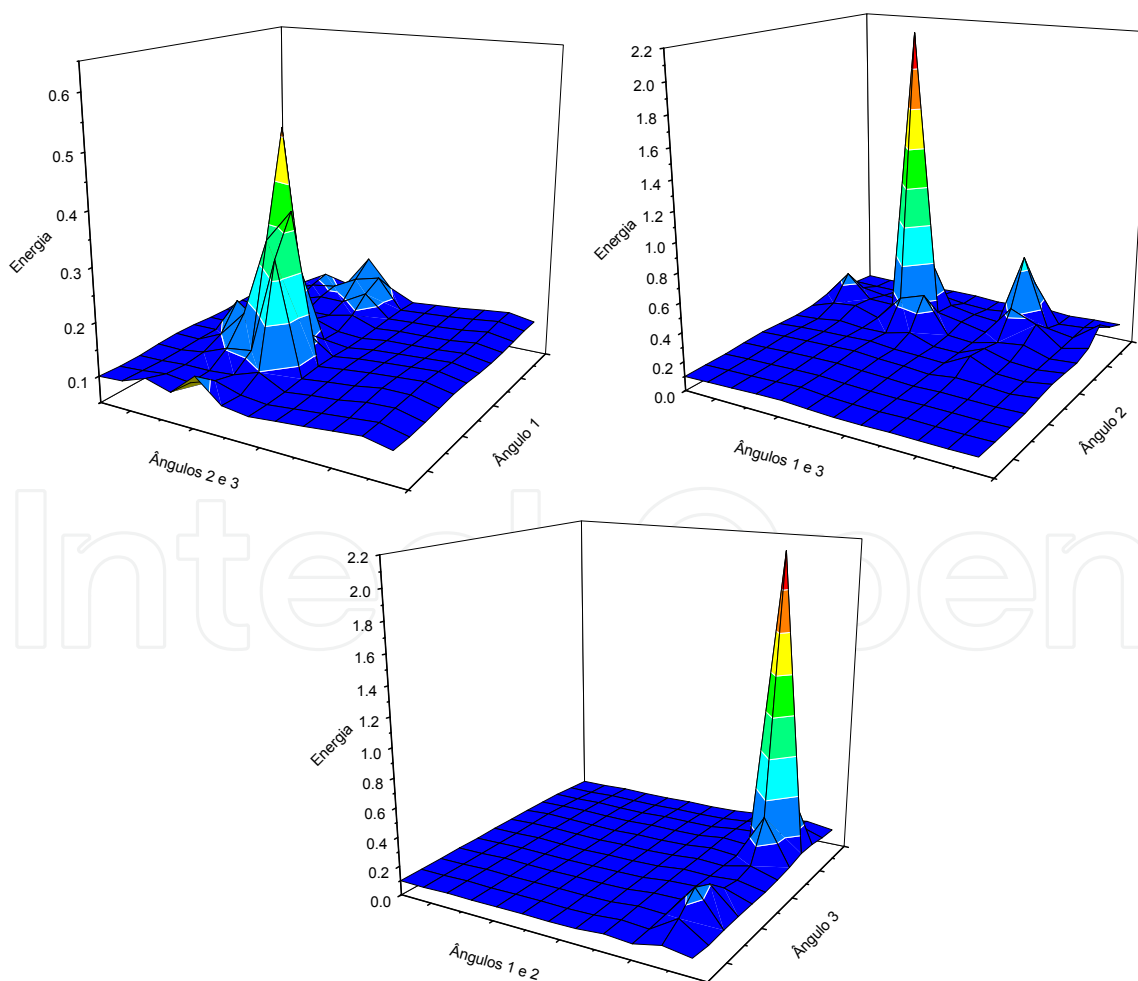

Fig. 9. Orignal PES obtained from PDA-SA method for structure from Fig. 8.
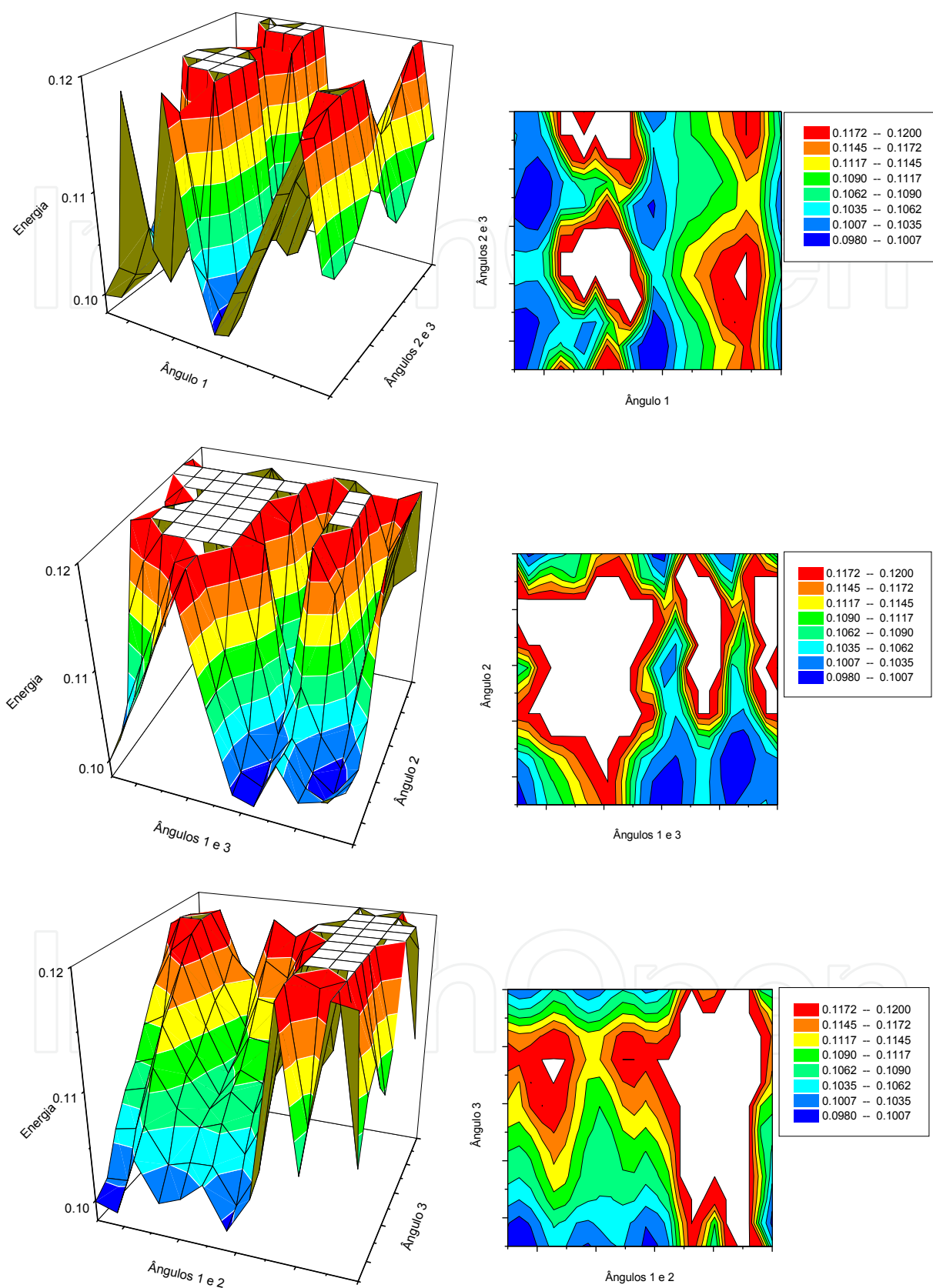
Fig. 10. PES obtained from PDA-SA method for structure from Fig. 8, with a 0,12 hartress cutoff.
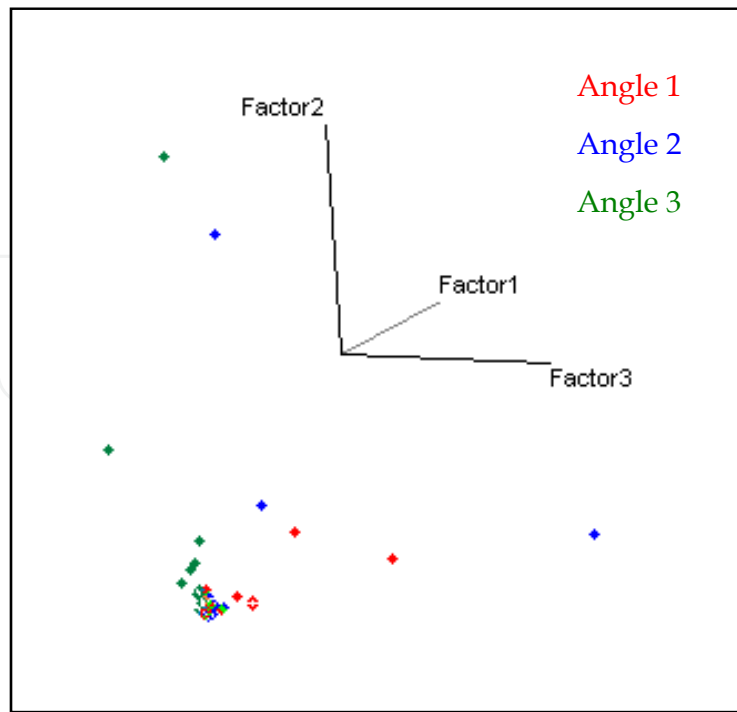
Fig. 11. PCA for data from original PES (Fig.9).
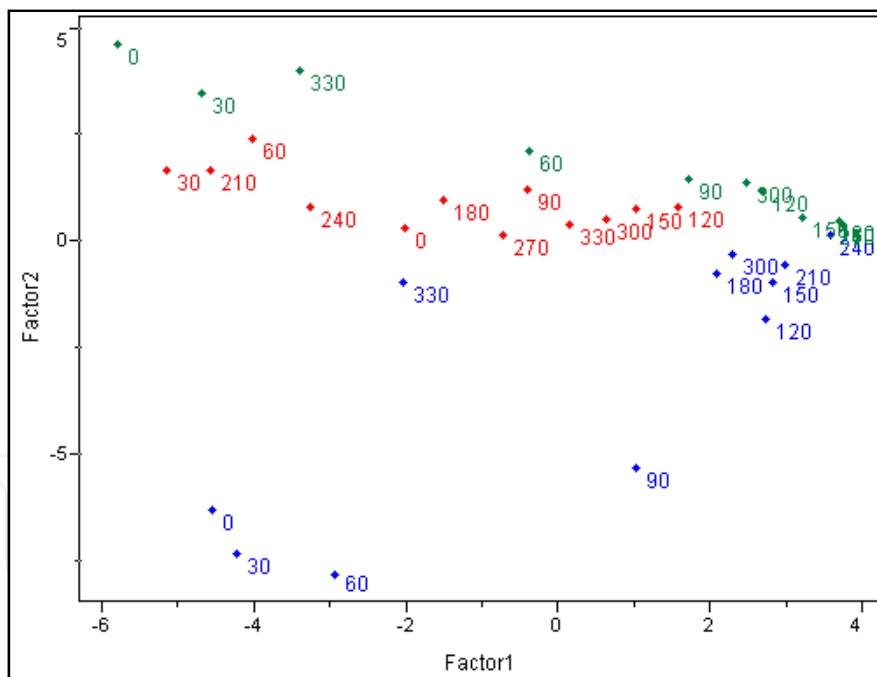
Fig. 12. Principal Component Analysis forPES from Fig. 10, with a 0,12 cutoff.

Factor 1 accounts to the minimum region in each case and Factor 2 accounts for the energy range for the different combinations. Table 1 shows the selected minimum energy for each angle. The first column shows that two different regions were chosen for Angle 1 and only one region for Angles 2 and 3. Second column shows the rotation over the initial angle value (third column) resulting in the fourth column.

Angle	Rotation	Initial Value	Value obtained by PCA
1 (a)	0° - 60°	48,48°	48,48° - 108,48°
1 (b)	180° - 240°	48,48°	228,48° - 288,48°
2	0° - 60°	209,79°	209,79° - 269,79°
3	330° - 30°	289,09°	259,09° - 319,09°

Table 1. Regions separated by PCA

Once minima energy regions were defined, a small angle increment (5°) was used on them. Results for PCA are in Figure 13. In all cases a parabolic behavior was observed. When data variation decreases, curves are more easily observed and the minimum point is detectable. The amount of information accounted for both first and second PC´s (Factors) is around 90%. Table 2 shows the final values for each angle. When these values are combined, two different geometries were obtained with similar energy values (Table 3). These conformations are shown in Figure 14.
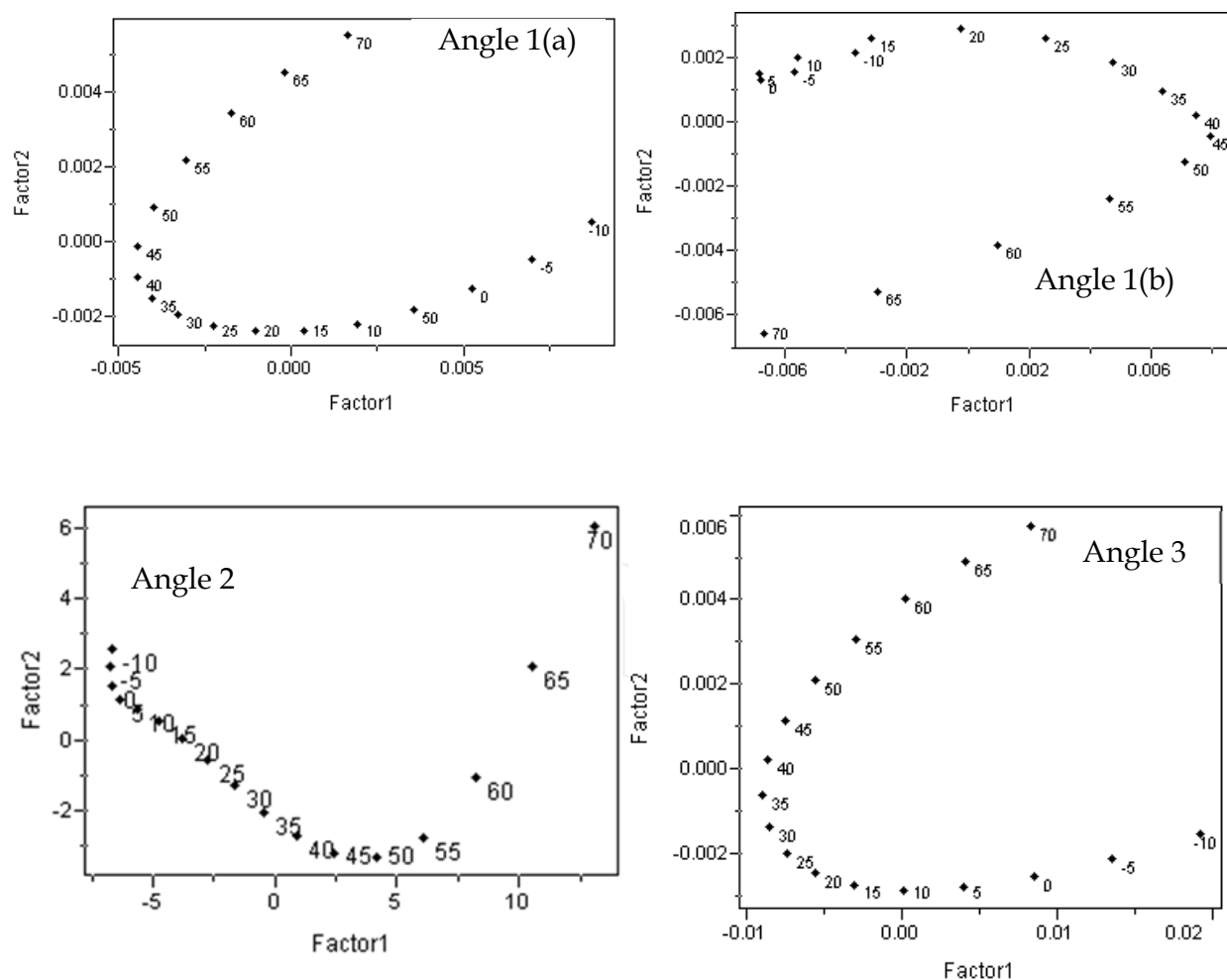

Fig. 13. PCA results for 5° angle increment refinement.

Angle	Rotation	Initial Value	Value obtained by PCA
1 (a)	45°	48,48° - 108,48°	93,48°
1 (b)	45°	228,48° - 288,48°	273,48°
2	45°	209,79° - 269,79°	254,79°
3	35°	259,09° - 319,09°	294,09°

Table 2. Regions obtained through PCA

Conformation	Angle	Obtained value	$\Delta H_{f\,(PM3)}$/kcal mol^{-1}	$E_{e(6-31G^{**})}$/hartree
A	1	92.29		
	2	107.10	54.96	-1134.33
	3	298.32		
B	1	266.45		
	2	175.70	54.71	-1134.35
	3	296.45		

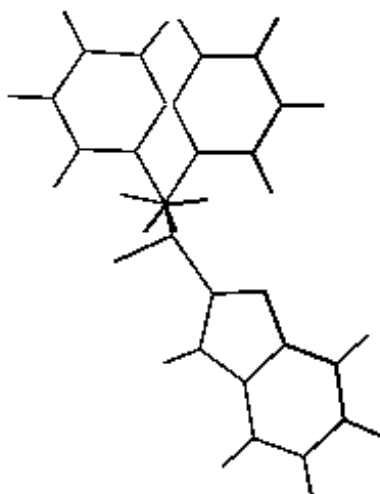Table 3. Minimum conformation characteristics (basic structure)

Fig. 14. Optimized superposed conformations for structure from Fig. 8.

In the second approach, the conformational analysis was made according to Equation 1 and took into account all possible conformations. PCA was performed on data matrix and minimum energy regions were selected. The next step a lower dihedral increment of 5° was used to refine those selected regions. PCA was performed again, and the same structures and energy, shown in Table 3, were obtained. This indicates that the two approaches are equivalent. The details of this complete systematic search can be found in (Bruni et al., 2002).

7.2 IAN preliminary studies

IAN (isobutyryl-Ala3-NH-methyl) tetrapeptide has also been studied to validate PDA-SA methodology. IAN has 11 consecutives dihedrals and its main characteristic is to be the shorter peptide able to make a complete helix turn. Figure 15 shows the IAN 2D structure

(Becker, 1998). Red arrows indicate the ψ, Φ e ω dihedrals. The dihedral angles ψ, ω and Φ are related to the rotations of single bonds between atoms in the main chain C (i)-C, OC-NH and N-C(i+1), respectively, where C (i) is the ith alpha carbon of the polypeptide chain. Angles ψ and Φ are connected to two arrays of functional protein chain: alpha-helix or beta-sheet.
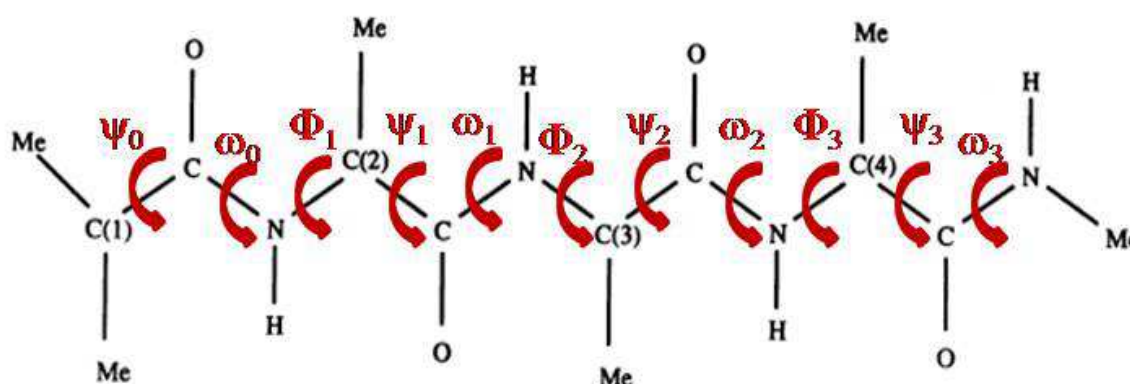

Fig. 15. 2D IAN peptide structure.

Ten random different starting conformations were studied. Table 4 shows the angles and energy values corresponding to these initial conformations. The red values indicate dihedrals that were changed in comparison to initial conformation number 1. The starting conformation 2 is close to an alpha-helix. Energy values correspond to single point AM1 semi-empirical calculation, in kcal mol^{-1}.

Number	Energy	ψ_0	ω_0	ϕ_0	ψ_1	ω_1	ϕ_1	ψ_2	ω_2	ϕ_2	ω_3	ϕ_3
1	-179.66	79.45	-169.01	-64.73	-44.75	171.78	-84.62	44.47	179.30	-144.22	-59.37	178.07
2	-122.24	**-60.55**	**179.00**	-64.73	**-64.75**	**-180.00**	**-64.62**	**-65.55**	**180.00**	**-64.22**	-59.37	178.07
3	-156.02	79.45	**170.99**	-64.73	**-34.75**	171.78	-84.62	**74.47**	179.30	**-124.22**	-59.37	178.07
4	195.46	79.45	-169.01	**-14.73**	-44.75	171.78	**-134.62**	44.47	179.30	-144.22	**-39.37**	178.07
5	-80.48	**49.45**	-169.01	-64.73	-44.75	**151.78**	-84.62	44.47	179.30	-144.22	**-39.37**	178.07
6	-169.24	79.45	**-149.01**	-64.73	-44.75	171.78	-84.62	44.47	**159.30**	-144.22	**-79.37**	178.07
7	-149.11	**59.45**	-169.01	**-84.73**	-44.75	**-168.22**	-84.62	**74.47**	179.30	-144.22	-59.37	178.07
8	-147.67	79.45	-169.01	-64.73	**-24.75**	171.78	**-104.62**	44.47	**-160.70**	-144.22	**-39.37**	178.07
9	-29.05	**109.45**	-169.01	-64.73	-44.75	171.78	-84.62	**24.47**	179.30	**-174.22**	-59.37	178.07
10	-167.76	79.45	-169.01	**-44.73**	-44.75	**-178.22**	-84.62	44.47	179.30	-144.22	-59.37	178.07

Table 4. Energy(kcal mol^{-1}) and dihedrals values (degrees) for each starting IAN structure.

IAN was analyzed using the PDA-SA procedure. The eleven dihedral angles provide 55 different conformations according to all possible combinations. Conformational analysis was performed with a 20° increment. PCA was carried out and Figure 16 shows that all points converge to specific regions of the phase space. Each selected region for each angle was refined with a 5° angle increment. PCA was performed again and the final structures characteristics are shown in Table 5.
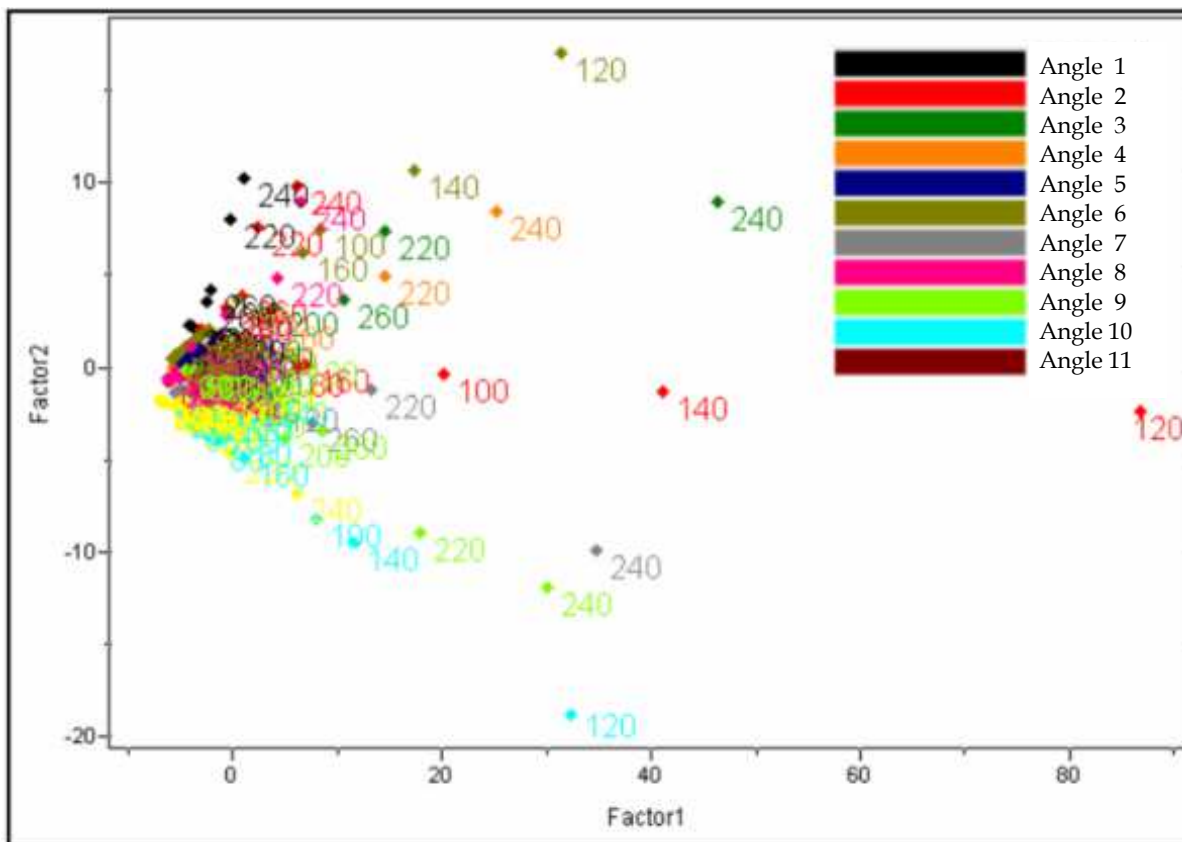
Fig. 16. PCA results for IAN peptide.

Table 5 shows the obtained energy values for the final structures and they indicate that some correspond to identical conformations. Three different groups were identified. Figure 17 shows the group that corresponds to structures 1, 5, 6, 9 and 10 superposed (blue ones in Table 5). Structure 9 shows a slightly different value on ψ_0 but it does not change the energy value. These five structures have two stabilizing hydrogen bonds which are indicated by the red circle and the resulting conformations for them resemble a beta-sheet.

Number	Energy	ψ_0	ω_0	ϕ_0	ψ_1	ω_1	ϕ_1	ψ_2	ω_2	ϕ_2	ω_3	ϕ_3
1	- 181.498	70.50	176.87	85.57	66.93	-177.14	-85.01	67.39	179.60	-112.24	-46.78	177.92
2	- 180.70	-61.83	-172.66	-78.26	25.05	157.24	-90.68	-29.19	177.15	-110.70	-45.10	178.62
3	- 179.661	79.45	-168.99	-64.80	-44.71	171.76	-84.61	44.32	179.26	143.99	-59.33	178.07
4	-179.276	72.93	-174.15	-104.24	-32.54	-179.9	-82.66	70.38	-179.69	-115.57	-51.50	179.42
5	-181.498	70.49	-176.89	-85.56	66.89	-177.17	-84.96	67.52	179.62	-112.34	-46.80	177.93
6	-181.498	70.49	-176.88	-85.57	66.93	-177.15	-85.01	67.40	179.61	-112.26	-46.78	177.92
7	-179.276	72.93	-174.15	-104.24	-32.52	-179.9	-82.65	70.39	-179.69	-115.58	-51.50	179.44
8	-179.276	72.94	-174.14	-104.27	-32.45	-179.89	-82.66	70.40	-179.67	-115.55	-51.48	179.36
9	-182.385	-62.43	-177.31	-84.72	67.46	-177.22	-84.95	67.38	179.58	-112.28	-46.85	177.94
10	-181.498	70.49	-176.89	-85.56	66.93	-177.15	-85.00	67.43	179.62	-112.32	-46.82	177.91

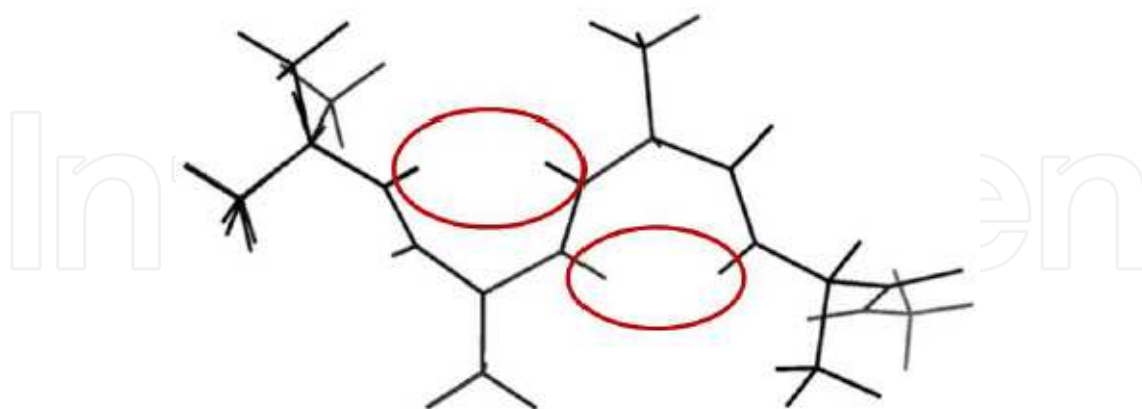Table 5. Energy (kcal mol^{-1}) and dihedrals values (degrees) for each obtained IAN structure.

Fig. 17. Superposed final conformations for 1, 5, 6, 9 and 10 structures (blue ones in Table 5).

The second group is composed by final conformations 4, 7 and 8 (green ones in Table 5). The superposed conformations can be observed in Figure 18. These conformations are more open and have only one hydrogen-bond (red circle in Fig.18). The last group, the black ones in Table 5, are superposed in Figure 19. The resulting structures show an alpha-helix like behavior, with two stabilizing hydrogen bond (red circles, Fig. 19).
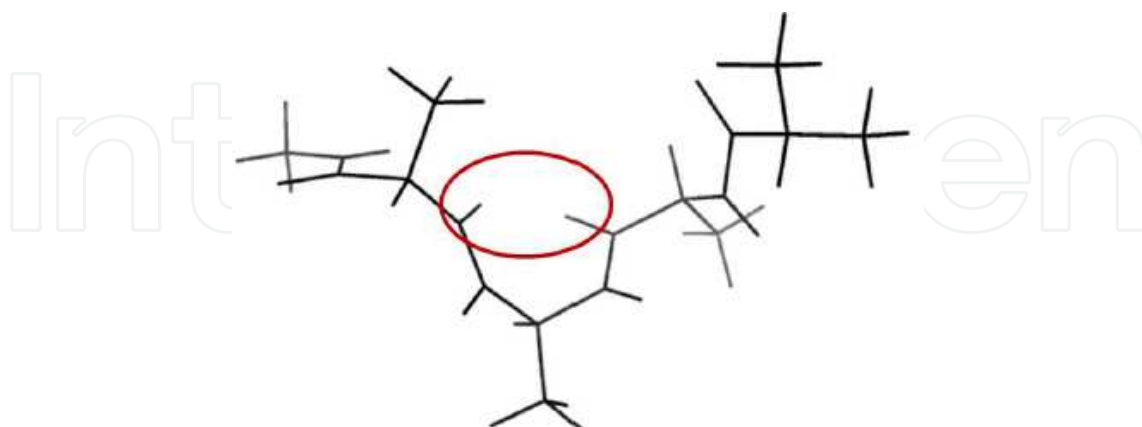

Fig. 18. Superposed final conformations for 4,7and 8 structures (green ones in Table 5).
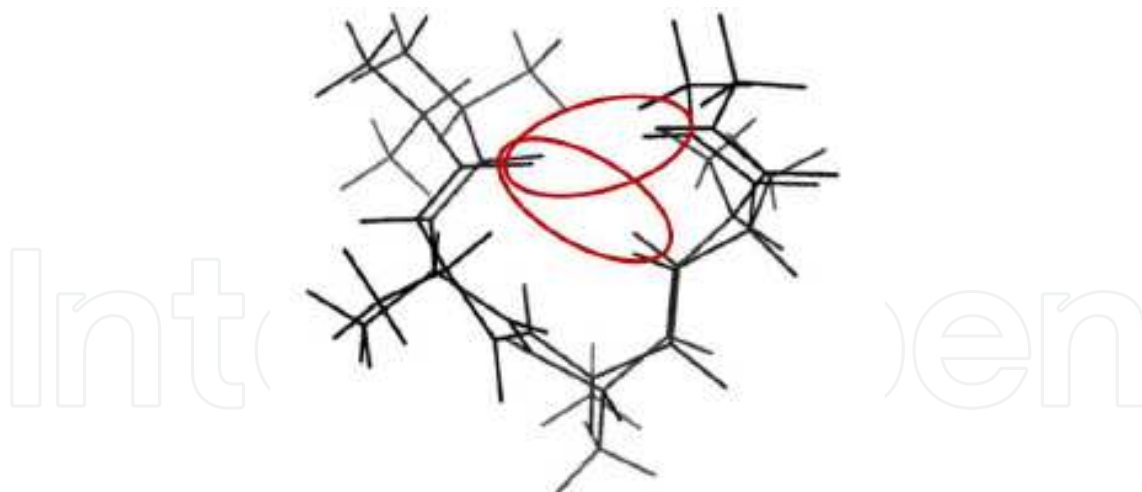
Fig. 19. Superposed final conformations for structures 2 and 3 (black ones in Table 5).

Results presented for IAN peptide are partial and were only performed for one minimum region for each starting structure. Other minimum energy regions of this system are being investigated. A gradual increase in the size of the chain is also been explored.

8. Conclusion

The arrangement of atoms in a molecule or its structure determination has intrigued scientists through history. However only with recent experimental and computational advances the discussions on this theme became more effective and elucidative. The nature of PES is intrinsically multidimensional, usually has a very complex landscape. The global minima search, like the one encountered in the protein folding problem, is a NP-hard problem. This means that this task belongs to a large set of computational problems, assumed to be very hard ("conditionally intractable") (Fraenkel, 1993). The search for its relevant minima in molecular modeling has motivated the development of methods with very specific applications, as discussed in this chapter. For each particular problem one finds a variety of methods that allows feasible solutions, and most likely a combination of methods provides the optimum solution.

In this chapter, we discussed some aspects of conformational search that controls the combinatorial explosion. In particular, Principal Component Analysis was associated with a systematic search method to find structures with low energy in PES. The methodology can be useful to handle small- and medium-size molecules. The maximum size which the method can efficiently handle is being investigated (Nascimento et al., 2009). Due to the PCA dimension reduction, the method's efficiency is highly increased, allowing it to be of practical use in the study of more complex molecules.

9. Acknowledgment

10. Apendix A

Matrix with the discrete values for each rotation angle and its corresponding energy value for the first rotation for basic structure in Figure 8. Labels in bold were not used in PCA analysis, they are shown to help the matrix notation and visualization.

Table part 1

Angle	Rotation	2												3												1											
		0	30	60	90	120	150	180	210	240	270	300	330	0	30	60	90	120	150	180	210	240	270	300	330	0	30	60	90	120	150	180	210	240	270	300	330
1	0	0.0998	0.100	0.103	0.111	0.119	0.113	0.106	0.118	0.146	0.133	0.106	0.099	0.0998	0.100	0.104	0.107	0.108	0.111	0.114	0.115	0.118	0.116	0.108	0.102	0.099	0.098	0.099	0.102	0.104	0.103	0.101	0.100	0.101	0.103	0.103	0.102
	30	0.0984	0.099	0.101	0.105	0.106	0.103	0.102	0.106	0.117	0.111	0.103	0.099	0.1006	0.101	0.105	0.108	0.110	0.113	0.115	0.118	0.123	0.115	0.108	0.103	0.100	0.099	0.099	0.101	0.104	0.103	0.100	0.099	0.099	0.101	0.102	0.102
	60	0.0992	0.099	0.100	0.102	0.103	0.104	0.103	0.103	0.107	0.115	0.123	0.108	0.1032	0.104	0.108	0.111	0.113	0.116	0.118	0.120	0.120	0.117	0.111	0.105	0.103	0.101	0.100	0.101	0.104	0.104	0.103	0.106	0.107	0.115	0.123	0.108
	90	0.1020	0.101	0.101	0.103	0.107	0.115	0.126	0.108	0.132	0.201	0.310	0.151	0.1115	0.113	0.117	0.120	0.122	0.125	0.127	0.128	0.141	0.138	0.127	0.121	0.107	0.108	0.111	0.113	0.115	0.117	0.119	0.120	0.122	0.117	0.113	0.112
	120	0.1046	0.104	0.104	0.107	0.124	0.196	0.174	0.252	0.322	0.634	0.383	0.116	0.1194	0.121	0.126	0.129	0.132	0.134	0.136	0.138	0.141	0.138	0.127	0.121	0.111	0.113	0.116	0.120	0.124	0.130	0.134	0.132	0.134	0.125	0.127	0.115
	150	0.1037	0.103	0.104	0.109	0.173	0.186	0.136	0.166	0.516	0.459	0.137	0.105	0.1136	0.116	0.122	0.127	0.132	0.131	0.132	0.134	0.139	0.150	0.127	0.114	0.110	0.113	0.116	0.125	0.136	0.131	0.125	0.132	0.134	0.121	0.108	0.110
	180	0.1010	0.100	0.100	0.103	0.116	0.121	0.104	0.112	0.243	0.145	0.107	0.100	0.1062	0.109	0.119	0.134	0.141	0.133	0.125	0.123	0.127	0.150	0.171	0.116	0.108	0.111	0.117	0.119	0.136	0.132	0.123	0.147	0.133	0.123	0.115	0.109
	210	0.1003	0.099	0.098	0.100	0.104	0.104	0.107	0.107	0.119	0.114	0.118	0.108	0.1182	0.133	0.172	0.238	0.189	0.147	0.133	0.434	0.319	1.241	0.376	0.140	0.114	0.123	0.135	0.147	0.145	0.252	0.166	0.319	2.168	0.178	0.123	0.113
	240	0.1019	0.099	0.098	0.099	0.102	0.105	0.109	0.114	0.120	0.150	0.146	0.117	0.1467	0.133	0.166	0.279	0.174	0.146	0.167	0.178	0.533	0.552	0.181	0.135	0.115	0.118	0.123	0.129	0.322	0.516	0.243	0.319	2.168	0.533	0.146	0.115
	270	0.1033	0.101	0.100	0.101	0.107	0.118	0.121	0.121	0.152	0.249	0.163	0.111	0.1467	0.133	0.166	0.310	0.383	0.459	0.243	0.178	0.533	0.368	0.167	0.120	0.116	0.123	0.120	0.125	0.634	0.516	0.516	0.150	0.552	0.368	0.249	0.533
	300	0.1036	0.102	0.102	0.106	0.121	0.132	0.127	0.126	0.180	0.216	0.126	0.103	0.1196	0.106	0.111	0.113	0.110	0.137	0.145	0.123	0.146	0.140	0.163	0.117	0.108	0.115	0.117	0.120	0.138	0.129	0.171	0.139	0.181	0.167	0.146	0.167
	330	0.1023	0.102	0.103	0.112	0.128	0.128	0.117	0.113	0.115	0.216	0.126	0.103	0.1047	0.099	0.101	0.104	0.105	0.105	0.109	0.113	0.115	0.116	0.111	0.103	0.102	0.103	0.105	0.113	0.121	0.114	0.116	0.140	0.135	0.120	0.108	0.107

Table part 2

	Angle	Rotation	3												1												2											
		0	30	60	90	120	150	180	210	240	270	300	330	0	30	60	90	120	150	180	210	240	270	300	330	0	30	60	90	120	150	180	210	240	270	300	330
1	0	0.099	0.100	0.104	0.107	0.108	0.111	0.114	0.115	0.118	0.116	0.108	0.102	0.099	0.098	0.103	0.102	0.104	0.103	0.101	0.100	0.101	0.103	0.103	0.099	0.099	0.100	0.104	0.111	0.119	0.113	0.106	0.118	0.146	0.146	0.108	0.104
	30	0.098	0.099	0.103	0.106	0.108	0.110	0.113	0.114	0.118	0.115	0.107	0.101	0.098	0.099	0.103	0.101	0.104	0.103	0.100	0.099	0.099	0.101	0.102	0.099	0.100	0.101	0.104	0.113	0.121	0.116	0.109	0.133	0.225	0.133	0.106	0.099
	60	0.099	0.101	0.104	0.107	0.109	0.112	0.117	0.125	0.119	0.115	0.107	0.101	0.103	0.101	0.104	0.107	0.109	0.112	0.117	0.119	0.125	0.119	0.115	0.107	0.104	0.105	0.108	0.111	0.113	0.116	0.118	0.120	0.120	0.166	0.111	0.101
	90	0.102	0.103	0.107	0.110	0.112	0.116	0.120	0.119	0.120	0.117	0.109	0.103	0.111	0.105	0.107	0.110	0.113	0.116	0.110	0.120	0.134	0.167	0.113	0.104	0.107	0.108	0.111	0.120	0.129	0.127	0.134	0.238	0.279	0.167	0.113	0.104
	120	0.104	0.106	0.110	0.113	0.115	0.117	0.119	0.120	0.122	0.119	0.111	0.106	0.104	0.106	0.110	0.113	0.115	0.117	0.119	0.120	0.132	0.134	0.136	0.121	0.108	0.110	0.113	0.120	0.132	0.125	0.127	0.189	0.174	0.140	0.110	0.105
	150	0.103	0.103	0.107	0.109	0.110	0.112	0.113	0.114	0.116	0.114	0.107	0.102	0.103	0.103	0.104	0.109	0.173	0.186	0.116	0.104	0.105	0.118	0.120	0.114	0.111	0.113	0.117	0.127	0.130	0.131	0.133	0.147	0.146	0.121	0.108	0.107
	180	0.101	0.103	0.104	0.106	0.107	0.109	0.110	0.111	0.114	0.112	0.106	0.101	0.101	0.100	0.103	0.103	0.124	0.173	0.116	0.119	0.107	0.108	0.109	0.116	0.114	0.115	0.118	0.127	0.130	0.125	0.133	0.167	0.126	0.111	0.109	0.102
	210	0.101	0.104	0.106	0.107	0.107	0.108	0.109	0.110	0.114	0.113	0.109	0.102	0.114	0.113	0.117	0.119	0.125	0.132	0.166	0.147	0.109	0.110	0.111	0.109	0.115	0.118	0.119	0.120	0.141	0.139	0.133	0.126	0.111	0.113	0.110	0.101
	240	0.103	0.104	0.106	0.108	0.107	0.108	0.110	0.113	0.117	0.115	0.111	0.104	0.115	0.118	0.123	0.129	0.134	0.139	0.243	0.434	0.533	0.552	0.376	0.135	0.118	0.123	0.125	0.129	0.140	0.178	0.126	0.140	0.146	0.115	0.104	0.102
	270	0.103	0.104	0.106	0.108	0.109	0.111	0.114	0.117	0.120	0.118	0.111	0.106	0.116	0.118	0.123	0.125	0.138	0.139	0.533	0.319	2.168	0.368	0.368	0.117	0.117	0.120	0.115	0.117	0.110	0.118	0.111	0.123	0.113	0.109	0.108	0.104
	300	0.102	0.103	0.106	0.108	0.110	0.112	0.115	0.113	0.120	0.118	0.111	0.105	0.108	0.108	0.117	0.120	0.127	0.129	0.150	0.134	0.181	0.167	0.140	0.112	0.108	0.115	0.107	0.109	0.111	0.120	0.107	0.106	0.106	0.106	0.111	0.106
	330	0.102	0.101	0.101	0.103	0.116	0.105	0.100	0.101	0.108	0.117	0.111	0.103	0.102	0.103	0.105	0.113	0.121	0.114	0.116	0.140	0.135	0.120	0.108	0.107	0.102	0.101	0.103	0.103	0.116	0.105	0.100	0.101	0.108	0.107	0.111	0.105

11. References

Allen, F. H., Galek, P. T. a, & Wood, P. a. (2010). Energy matters! *Crystallography Reviews*, *16*(3), 169-195. doi:10.1080/08893110903476919

Araujo-Andrade, C., Lopes, S., Fausto, R., & Gómez-Zavaglia, A. (2010). Conformational study of arbutin by quantum chemical calculations and multivariate analysis. *Journal of Molecular Structure*, *975*(1-3), 100-109. doi:10.1016/j.molstruc.2010.04.002

Becker, O M. (1998). Principal coordinate maps of molecular potential energy surfaces. *Journal of Computational Chemistry*, *19*(11), 1255-1267. 605 Third Ave, New York, NY 10158-0012 Usa: John Wiley & Sons Inc. doi:10.1002/(SICI)1096-987X(199808)19:11<1255::AID-JCC5>3.3.CO;2-H

Becker, Oren M., & Karplus, M. (1997). The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *The Journal of Chemical Physics*, *106*(4), 1495. doi:10.1063/1.473299

Beckers, M. L. M., Derks, E. P. P. A., Melssen, W. J., & Buydens, L. M. C. (1996). Pergamon oo!n-8485(%~6-0. *Science*, *20*(4), 449-457.

Beckers, M. L., Buydens, L. M., Pikkemaat, J. a, & Altona, C. (1997). Application of a genetic algorithm in the conformational analysis of methylene-acetal-linked thymine dimers in DNA: comparison with distance geometry calculations. *Journal of biomolecular NMR*, *9*(1), 25-34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9081542

Beebe, K. R. (1998). *Chemometrics: A Practical Guide* (p. 360). Wiley-Blackwell. Retrieved from http://www.amazon.co.uk/Chemometrics-Practical-Wiley-Interscience-Laboratory-Automation/dp/0471124516/ref=sr_1_1?ie=UTF8&qid=1320084260&sr=8-1

Beusen, D. D., Shands, E. F. B., Karasek, S. F., Marshall, G. R., & Dammkoehler, R. A. (1996). Systematic search in conformational analysis. *Theochem-Journal of Molecular Structure*, *370*(2-3), 157-171. Po Box 211, 1000 Ae Amsterdam, Netherlands: Elsevier Science Bv.

Biarnés, X., Bongarzone, S., Vargiu, A. V., Carloni, P., & Ruggerone, P. (2011). Molecular motions in drug design: the coming age of the metadynamics method. *Journal of computer-aided molecular design*, *25*(5), 395-402. doi:10.1007/s10822-011-9415-3

Brodmeier, T., & Pretsch, E. (1994). Application of genetic algorithms in molecular modeling. *Journal of Computational Chemistry*, *15*(6), 588-595. doi:10.1002/jcc.540150604

Bruni, A. T., Leite, V. B. P., & Ferreira, M. M. C. (2002). Conformational analysis: A new approach by means of chemometrics. *Journal Of Computational Chemistry*, *23*(2), 222-236. Commerce Place, 350 Main St, Malden 02148, MA USA: Wiley-Blackwell. doi:10.1002/jcc.10004

Bruni, A. T., & Ferreira, M. M. C. (2008). Theoretical study of omeprazole behavior: Racemization barrier and decomposition reaction. *International Journal of Quantum Chemistry*, *108*(6), 1097-1106. doi:10.1002/qua.21597

Bruni, A. T., & Ferreira, M. M. C. (2002). Omeprazole and analogue compounds: a QSAR study of activity againstHelicobacter pylori using theoretical descriptors. *Journal of Chemometrics*, *16*(8-10), 510-520. doi:10.1002/cem.737

Brush, S. G. (1999). Dynamics of Theory Change in Chemistry : Part 1 . The Benzene Problem 1865 – 1945. *Science*, *30*(1), 21-79.

Cintas, P. (2007). Tracing the origins and evolution of chirality and handedness in chemical language. *Angewandte Chemie (International ed. in English)*, *46*(22), 4016-24. doi:10.1002/anie.200603714

Das, G., Gentile, F., Coluccio, M. L., Perri, a M., Nicastri, a, Mecarini, F., Cojoc, G., et al. (2011). Principal component analysis based methodology to distinguish protein SERS spectra. *Journal of Molecular Structure*, *993*(1-3), 500-505. Elsevier B.V. doi:10.1016/j.molstruc.2010.12.044

Derks, E. P. P. A., & Buydens, L. M. C. (1996). E. P. P. A DERKS,* M. L. M. BECKER& W. J. MELSSEN and L. M. C. BUYDENS, *20*(4), 439-448.

Drayer, D. (1993). The Early History of Stereochemistry: From the Discovery of Molecular Asymmetry and the First Resolution of a Racemate by Pasteur to the Asymmetrical Chiral Carbon of van't Hoff and Le Bel. *Clinical Pharmacology-New York-Marcel Dekker Incorporated-*, *18*(3), 1–1. Marcel Dekker Ag. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Early+History+of+Stereochemistry+:+From+the+Discovery+of+Molecular+Asymmetry+and+the+First+Resolution+of+A+Racemate+By+Pasteur+To+The+Asymmetrical+Chiral+Carbon+Of+Van+?+T+Hoff+And+Le+Bel+*#0

Fraenkel, a S. (1993). Complexity of protein folding. *Bulletin of mathematical biology*, *55*(6), 1199-210. Retrieved from
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3042729&tool=pmcentrez&rendertype=abstract

Gal, J. (2007). Review Article Carl Friedrich Naumann and the Introduction of Enantio Terminology : A Review and Analysis on the 150th Anniversary. *Chirality*, *98*(May 2006), 89-98. doi:10.1002/chir

Gal, J. (2011). Review Article Louis Pasteur , Language , and Molecular Chirality . I . Background and Dissymmetry. *Clinical Laboratory*, *16*(March 2010), 1-16. doi:10.1002/chir

Geladi, P. (2003). Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochimica Acta Part B Atomic Spectroscopy*, *58*(5), 767-782. doi:10.1016/S0584-8547(03)00037-5

Golub, G. H., & Loan, C. F. van V. (1996). *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition)* (p. 728). The Johns Hopkins University Press. Retrieved from http://www.amazon.com/Computations-Hopkins-Studies-Mathematical-Sciences/dp/0801854148

Grouleff, J., & Jensen, F. (2011). Searching Peptide Conformational Space. *Journal of Chemical Theory and Computation*, 1783-1790.

Hanselman, D., Littlefield, B., Inc., M., & Mathworks. (1997). *The Student Edition of Matlab Version 5 User's Guide* (p. 429). Prentice Hall College Div. Retrieved from http://www.amazon.com/Student-Matlab-Version-Users-Guide/dp/0132725509

Havel, T. F., Crippen, G. M., Kuntz, I. D., & Blaney, J. M. (1983). The combinatorial distance geometry method for the calculation of molecular conformation. II. Sample problems and computational statistics. *Journal of theoretical biology*, *104*(3), 383-400. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6197591

Havel, T. F., Kuntz, I. D., & Crippen, G. M. (1983). The combinatorial distance geometry method for the calculation of molecular conformation. I. A new approach to an old problem. *Journal of theoretical biology*, *104*(3), 359-81. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/6656266

Hunger, J., & Huttner, G. (1999). Optimization and analysis of force field parameters by combination of genetic algorithms and neural networks. *Journal of Computational Chemistry*, *20*(4), 455-471. doi:10.1002/(SICI)1096-987X(199903)20:4<455::AID-JCC6>3.0.CO;2-1

Izgorodina, E. I., Lin, C. Y., & Coote, M. L. (2007). Energy-directed tree search: an efficient systematic algorithm for finding the lowest energy conformation of molecules. *Physical chemistry chemical physics : PCCP*, *9*(20), 2507-16. doi:10.1039/b700938k

Jordan, S. N., Leach, A. R., & Bradshaw, J. (1995). The Application of Neural Networks in Conformational Analysis. 1. Prediction of Minimum and Maximum Interatomic Distances. *Journal of Chemical Information and Modeling*, *35*(3), 640-650. doi:10.1021/ci00025a035

Jorgensen, W. L., & TiradoRives, J. (1996). Monte Carlo vs molecular dynamics for conformational sampling. *Journal of Physical Chemistry*, *100*(34), 14508-14513. 1155 16th St, Nw, Washington, Dc 20036: Amer Chemical Soc. doi:10.1021/jp960880x

Kiralj, R., Ferreira, M. C., Donate, P. M., & Silva, R. (2007). Combined Computational , Database Mining , NMR , and Chemometric Approaches. *Analysis*, 6316-6333.

Lucasius, C. B., & Kateman, G. (1994). Understanding and Using Genetic Algorithms.2. Representation, Configuration and Hybridization. *Chemometrics and Intelligent Laboratory Systems*, *25*(2), 99-145. Po Box 211, 1000 Ae Amsterdam, Netherlands: Elsevier Science Bv. doi:10.1016/0169-7439(94)85038-0

Leach, A. (2001). *Molecular Modelling: Principles and Applications (2nd Edition)*. Prentice Hall. Retrieved from http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0582382106

Leach, A. R., & Smellie, A. S. (1992). A combined model-building and distance-geometry approach to automated conformational analysis and search. *Journal of Chemical Information and Modeling*, *32*(4), 379-385. doi:10.1021/ci00008a019

Lucasius, C. (1993). Understanding and using genetic algorithms Part 1. Concepts, properties and context. *Chemometrics and Intelligent Laboratory Systems*, *19*(1), 1-33. doi:10.1016/0169-7439(93)80079-W

Malinowski, E. R. (2003). *Factor Analysis in Chemistry. Technometrics* (Vol. 45, pp. 180-181). Wiley. doi:10.1198/tech.2003.s145

Li Manni, G., Barone, G., Duca, D., & Murzin, D. Y. (2009). Systematic conformational search analysis of the SRR and RRR epimers of 7-hydroxymatairesinol. *Journal of Physical Organic Chemistry*, (June 2009), n/a-n/a. doi:10.1002/poc.1595

Massart, D. L., Heyden, Y. V., & Brussel, V. U. (2004). What Can Chemometrics Do for Separation Science ? *Europe*, *17*(9).

Mo, Y., & Gao, J. (2007). Theoretical analysis of the rotational barrier of ethane. *Accounts of chemical research*, *40*(2), 113-9. doi:10.1021/ar068073w

Nair, N., & Goodman, J. M. (1998). Genetic Algorithms in Conformational Analysis. *Journal of Chemical Information and Modeling*, *38*(2), 317-320. doi:10.1021/ci970433u

Nascimento, R. R., Bruni, A. T. , & Leite, V. B. P. (2009). Estudo conformacional do peptídeo IAN e seus fragmentos pelo método de análise sistemática reduzida. *07/10/09*. Retrieved November 1, 2011, from http://www.athena.biblioteca.unesp.br/exlibris/bd/brp/33004153068P9/2009/nascimento_rr_me_sjrp_parcial.pdf

Oblinsky, D. G., Vanschouwen, B. M. B., Gordon, H. L., & Rothstein, S. M. (2009). Procrustean rotation in concert with principal component analysis of molecular dynamics trajectories: Quantifying global and local differences between conformational samples. *The Journal of chemical physics*, *131*(22), 225102. doi:10.1063/1.3268625

Pietropaolo, A., Branduardi, D., Bonomi, M., & Parrinello, M. (2011). A Chirality-Based Metrics for Free-Energy Calculations in Biomolecular Systems. *Journal of Computational Chemistry*. doi:10.1002/jcc

Ramberg, P. J., & Somsen, G. J. (2001). Annals of Science The Young J . H . van ′ t Hoff : The Background to the Publication of his 1874 Pamphlet on the Tetrahedral Carbon Atom , Together with a New English Translation. *Annals of Science*, (September 2011), 51-74.

Ramos, L. S., Beebe, K. R., Carey, W. P., M, E. S., Erickson, B. C., Wilson, B. E., Wangen, L. E., et al. (1986). L. Scott Ramos, Kenneth R. Beebe, W. Patrick Carey, Eugenio Sfinchez M., Brice C. Erickson, Bruce E. Wilson, Lawrence E. Wangen,' and Bruce R. Kowalski* Laboratory for Chemometrics, Department. *Education*, (300), 31-49.

Seabra, G. D. M., Walker, R. C., & Roitberg, A. E. (2009). Are current semiempirical methods better than force fields? A study from the thermodynamics perspective. *The journal of physical chemistry. A*, *113*(43), 11938-48. doi:10.1021/jp903474v

Silva, D.-A., Domínguez-Ramírez, L., Rojo-Domínguez, A., & Sosa-Peinado, A. (2011). Conformational dynamics of L-lysine, L-arginine, L-ornithine binding protein reveals ligand-dependent plasticity. *Proteins*, *79*(7), 2097-108. doi:10.1002/prot.23030

Smellie, A., Stanton, R., Henne, R., & Teig, S. (2003). Conformational analysis by intersection: Conan. *Journal of Computational Chemistry*, *24*(1), 10-20. 111 River St, Hoboken, Nj 07030 Usa: John Wiley & Sons Inc. doi:10.1002/jcc.10175

Spellmeyer, D. C., Wong, a K., Bower, M. J., & Blaney, J. M. (1997). Conformational analysis using distance geometry methods. *Journal of molecular graphics & modelling*, *15*(1), 18-36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9346820

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal Component Analysis. *Chemometrics And Intelligent Laboratory Systems*, *2*(1-3), 37-52. Po Box 211, 1000 Ae Amsterdam, Netherlands: Elsevier Science Bv. doi:10.1016/0169-7439(87)80084-9

Wold, S. (1995). Chemometrics; what do we mean with it, and what do we want from it? *Chemometrics and Intelligent Laboratory Systems*, *30*(1), 109-115. doi:10.1016/0169-7439(95)00042-9

Yanmaz, E., Sarıpınar, E., Şahin, K., Geçen, N., & Çopur, F. (2011). 4D-QSAR analysis and pharmacophore modeling: electron conformational-genetic algorithm approach for penicillins. *Bioorganic & medicinal chemistry*, *19*(7), 2199-210. doi:10.1016/j.bmc.2011.02.035
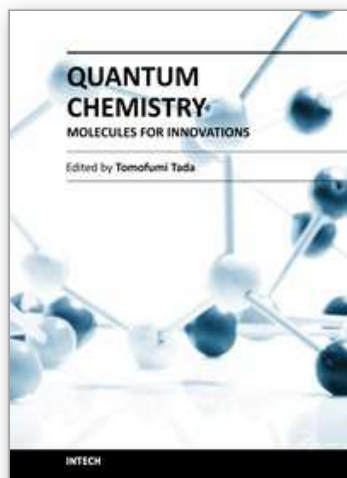
Quantum Chemistry - Molecules for Innovations

Edited by Dr. Tomofumi Tada

Molecules, small structures composed of atoms, are essential substances for lives. However, we didn't have the clear answer to the following questions until the 1920s: why molecules can exist in stable as rigid networks between atoms, and why molecules can change into different types of molecules. The most important event for solving the puzzles is the discovery of the quantum mechanics. Quantum mechanics is the theory for small particles such as electrons and nuclei, and was applied to hydrogen molecule by Heitler and London at 1927. The pioneering work led to the clear explanation of the chemical bonding between the hydrogen atoms. This is the beginning of the quantum chemistry. Since then, quantum chemistry has been an important theory for the understanding of molecular properties such as stability, reactivity, and applicability for devices. This book is devoted for the theoretical foundations and innovative applications in quantum chemistry.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Aline Thaís Bruni and Vitor Barbanti Pereira Leite (2012). Quantum Chemistry and Chemometrics Applied to Conformational Analysis, Quantum Chemistry - Molecules for Innovations, Dr. Tomofumi Tada (Ed.), ISBN: 978-953-51-0372-1, InTech, Available from: http://www.intechopen.com/books/quantum-chemistry-molecules-for-innovations/quantum-chemistry-and-chemometrics-applied-to-conformational-analysis

INTECH
open science | open minds