Conformational Analysis: A New Approach by Means of Chemometrics

ALINE THAÍS BRUNI,[1] VITOR B. P. LEITE,[2] MÁRCIA M. C. FERREIRA[1]
[1]*Instituto de Química, Universidade Estadual de Campinas UNICAMP, Campinas, SP, 13083-970 Brazil*
[2]*Departamento de Física, IBILCE, Universidade Estadual Paulista, São José do Rio Preto, SP, 15054-000 Brazil*

Received 7 August 2000; Accepted 23 July 2001

Abstract: In conformational analysis, the systematic search method completely maps the space but suffers from the combinatorial explosion problem because the number of conformations increases exponentially with the number of free rotation angles. This study introduces a new methodology of conformational analysis that controls the combinatorial explosion. It is based on a dimensional reduction of the system through the use of principal component analysis. The results are exactly the same as those obtained for the complete search but, in this case, the number of conformations increases only quadratically with the number of free rotation angles. The method is applied to a series of three drugs: omeprazole, pantoprazole, lansoprazole—benzimidazoles that suppress gastric-acid secretion by means of H^+, K^+-ATPase enzyme inhibition.

 J Comput Chem 23: 222–236, 2002

Key words: principal component analysis; chemometrics; omeprazole; pantoprazole; lansoprazole; conformational analysis

Introduction

Experimental techniques are limited and are sometimes insufficient for the study of complex systems. Following recent computational advances, new methods have been applied to the study of compounds and reactions in several fields of science. In medicinal chemistry and pharmaceutical research, important issues are structure elucidation, conformational analysis, physico-chemical characterization, and biological activity determination.[1] These issues are helpful for investigating and elucidating how biological systems evolve and for determining the properties of a given drug. In these areas, methods of theoretical chemistry provide powerful tools for investigating and understanding, at a molecular level, the relationship between chemical structure and biological activity, and also for providing data for the design of new compounds.[2]

All chemical information is intimately tied to the three-dimensional atomic arrangement and to the electronic properties of specific sites of a given compound.[3] The natural way to begin the theoretical study of a given drug is through structural determination. The main goal of molecular structure determination is to provide a starting point for understanding the physical, chemical, and biological properties of matter.[4] Each different spatial arrangement of a molecule, known as a conformation, is defined by the arrangement of its atoms in space, which can be interconverted by rotation about single bonds.[5, 6] There are several ways to find the spatial arrangement of a molecule. Spectroscopic (microwave, Raman, Infrared, NMR) and diffraction techniques (X-ray, synchrotron, electron, neutron diffraction) are, among others, widely used experimental techniques for structural determination. In this study, we will only focus on the theoretical methods for three-dimensional arrangement determination.

Systems with many degrees of freedom have thermodynamic and dynamic properties determined by the nature of their potential energy surfaces. Analysis of molecular conformation space is used for locating stable structures of drug molecules. Potential energy surfaces (PES) can be characterized by their minima, which correspond to locally stable configurations, and by the saddle points or transition regions that connect the minima.[7–9] Theoretical calculations can be performed in different ways to find minimum energy structures, according to the methodology used. A variety of strategies have been described in recent years. They are capable of locating minimum energy structures on the conformational potential energy surfaces. The most common strategies are distance geometry, neural networks, genetic algorithm, simulation methods (Monte Carlo and Molecular Dynamics) and systematic analysis.

Correspondence to: M. M. C. Ferreira; e-mail: marcia@iqm.unicamp.br

Contract/grant sponsors: CNPq (to A.T.B.) and FAPESP (to M.M.C.F. and V.B.P.L.)

DOI 10.1002/jcc.1004

Distance geometry methods in conformational analysis convert a set of distance constraints into a set of cartesian coordinates. The distances are given according to bonded atoms, bond angles, and torsion angles for free and rigid angles, and any other known constraints on the system. These distances are taken into a matrix, and the minimum and maximum distances are entered as the lower and upper bounds, respectively.[10] Artificial neural networks are based on concepts inspired by the theories of the cell network of the human brain.[11] They have been applied to nonlinear problems and pattern recognition studies. In conformational analysis, neural networks have been used to predict the maximum and minimum distances between pairs of heteroatoms.[12] Another way to explore conformational space of molecules is by the use of genetic algorithms (GA). A GA is a large-scale optimization algorithm mimicking a biological evolution in a randomly generated population. In conformational analysis, this population is formed by a number of conformations. The adaptation is calculated, and a new population is generated according to operators (reproduction, crossover, and mutation). The process is repeated until it converges to a minimum energy structure.[13, 14] Monte Carlo (MC) and molecular dynamics (MD) are simulation methods used to find the minimum energy structure of a given system. The main difference between them is in the way the conformational space is sampled. In MC methods, the configuration is generated randomly by variations on the cartesian or internal coordinates. Each configuration is accepted or not, according to some algorithm, usually the Metropolis algorithm. In the Metropolis method, the average values for a studied structure are Boltzmann weighted.[15, 16] In MD methods, system configurations are given by integration of Newton's laws for motion over a small time step, and new atomic positions and velocities are determined.[15, 16] In some cases, a combination of these methods is used to perform a conformational search.[17] Simulated annealing is an example. In this method, the system is initially set at a high temperature, which is gradually lowered until a configurational minimum is achieved. At each thermal step, equilibrium is reached by MC or MD implementation in the program.[18] All methods described above are stochastic, in the sense that there is no natural endpoint for search. Sometimes only a subset of conformational space is explored, and system convergence is not guaranteed. Some of them may present difficulties depending on the characteristics of the investigated system. In general, if only the standard Metropolis Monte Carlo method is used, it fails in relation to flexible molecules due the small acceptance rate. There are exceptions in which this difficulty could be overcome. For example, the MC method has been successfully used to find minimum energy conformations for cycloalkanes using specific constraint conditions, but such conditions are used in particular cases, and cannot generally be extended to unrestricted systems.[19] A method that can be extended to both, restricted and unrestricted systems, proposed by Li and Scheraga, combines energy minimization and the Metropolis Monte Carlo method to study the multiple-minima problem in protein folding.[20]

Simulated annealing is another method that may have infinite temperature steps to be equilibrated, rendering the simulation impractical.

On the other hand, there are deterministic methods, which completely map the conformational space. These methods are known as a systematic search, in which, for a given starting geometry, the torsion angles are varied by regular increments.[21] However, it is sometimes impossible to use this method, due to the enormous combinatorial complexity of the problem. To perform a grid search in the conformational space, a series of conformations would be generated by systematically rotating the torsion angles around the single bonds between 0° and 360°. For each case, 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 discrete values for each rotation angle. This number is given by $360/\theta_i$ with θ_i being the dihedral increment of angle i. The number of conformations increases exponentially with the number of bonds that have free rotation. This combinatorial explosion is the major problem involved in a systematic search. There are some strategies for defeating the combinatorial explosion, for example, building molecules from aggregates, or by the use of distance constraint equations, etc. More details are given by Beusen et al.[21]

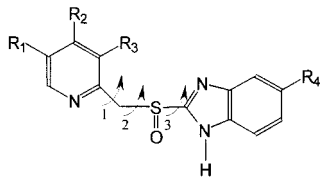
This study introduces a new methodology for control of combinatorial explosion in systematic searches. Our strategy minimizes computational time by reducing the system's dimensions. Quantum chemistry and chemometric methods were combined to find the best conformational structures, by identifying conformations, which correspond to minima on the potential energy surface.

Data analysis strategies have been published, using chemometrics for handling conformational problems.[22–25] In medicinal chemistry, chemometrics is widely applied in quantitative structure-activity relationship (QSAR) studies.[26–28] One of its applications is the mapping of potential energy surfaces, by the quantitative visualization of a macromolecular energy funnel.[29, 30] Quantitative QSAR studies can also be performed by a combination of the methods described earlier.[31–33]

The PCA-reduced search introduced in this study is a systematic conformational analysis. The dihedral increment to be taken is not less than usually used in the literature for a complete systematic search, which is believed to be sufficient to avoid any gross variation in between. Another advantage of the proposed methodology is that because one of the principal components refers to the surface rugosity, it can also be used as a validation criterion. In this way, one can be sure that the potential energy surface is completely explored. If the combinatorial explosion problem is controlled, the grid can be sufficiently refined on the minimum energy regions, as it is shown in this article. Thus, no information about minimum energy configuration is lost.

Systems Studied

The proposed method was used to investigate the conformational analysis of three drugs: omeprazole, lansoprazole, and pantoprazole (Fig. 1). These drugs are substituted benzimidazoles, which suppress gastric-acid secretion by means of H^+, K^+-ATPase enzyme inhibition.[34–36] There are several pharmacokinetic and metabolical studies about these molecules as well as their interaction with other drugs. However, there have been few stereochemical



Substituent	Omeprazole	Pantoprazole	Lansoprazole
R_1	CH_3	-	-
R_2	OCH_3	OCH_3	OCH_2CF_3
R_3	CH_3	OCH_3	CH_3
R_4	OCH_3	OCF_2H	-

Figure 1. The basic structure for omeprazole, pantoprazole, and lansoprazole molecules.

investigations,[36, 37] and no conformational theoretical studies of them in the literature.

Methodology

Energy surfaces were obtained for each pair of angles, indicated by arrows in Figure 1. The number of conformations is given by:

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

where s is the same as defined in eq. (1).

One can observe that the number of conformations as given by eq. (1) increases exponentially with the number of bonds that have free rotation, while from eq. (2), the number of studied conformations increases quadratically with N. As the number of free rotating angles increases, the difference in the number of conformations between these two equations becomes more evident. After first calculating the energy surface for each pair of angles, the principal component analysis (PCA) technique was used to find the lowest energy conformations for each molecule, in accordance with eq. (2).

Principal Component Analysis is a mathematical technique used to reduce the dimensions needed to portray accurately the characteristics of data matrices.[38, 39] By means of this method the original matrix is represented by a set of new variables, called *principal components*. Each PC is constructed as a linear combination of variables:

$$p_i = \sum_{j=1}^{v} c_{i,j} x_j, \tag{3}$$

where p_i is the ith principal component and $c_{i,j}$ is the coefficient of the variable $x_{i,j}$.[5] There are v such variables. The first principal component PC1 is chosen in such a way that the new axis p_l has the direction that maximizes the variance of the data along 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 the singular value decomposition SVD.[40] A matrix of arbitrary size can be decomposed into the product of three matrices in such a way that:

$$\mathbf{X} = \mathbf{USV}^t, \tag{4}$$

where **U** and **V** are square orthogonal matrices. The matrix **U** (whose columns are the eigenvectors of $\mathbf{XX}^t$) contains the coordinates of samples along the PC axes. The **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 axis [$c_{i,j}$ coefficients in eq. (3)]. The **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,[40] for which computationally efficient software has been developed and can be used on a routine basis[41] for very large-size matrices.

Procedure Details

All the initial structures were constructed using the Spartan software.[42] The PM3[43] semiempirical method from the Gaussian 98 program[44] was used to carry out the calculations. After being constructed, the molecules were preoptimized, and the rotation barriers were calculated.

In this study, molecules shown in Figure 1 have at least three bonds with free rotation. To introduce and validate the proposed methodology, the basic structure (corresponding to the common structure for the compounds, with the exception of the substituents, Fig. 1) was studied initially. Two distinct approaches were used to perform conformational systematic analysis and compared to each other. One uses the new methodology, and the other performs an extensive conformational search.

In the first one, rotations were performed in pairs of angles [(1,2), (1,3), and (2,3) in Fig. 1], and the number of conformations (points) follows eq. (2). The matrix to be analyzed consists of energy values from potential surfaces for angle combinations, and they are grouped according to Scheme 1. 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 two others, allow the conformational space to be completely mapped. An example of such matrix, with the discrete values for each rotation angle is presented by Scheme 2 (the complete matrix can be found in the Supplementary Material, item A).

In the second approach the analysis was performed through an extensive search, in which the number of conformations (points) follows eq. (1).

$$\begin{bmatrix} [E](1,2) & [E](1,3) \\ [E](2,3) & [E](2,1) \\ [E](3,1) & [E](3,2) \end{bmatrix}$$

SCHEME 1. Schematic representation of cyclical permutation on energy data matrix. The numbers correspond to the dihedral angles according to Figure 1. $[E](i, j)$s are symmetric matrices.

Angle									
		2				3			
	Rotation	0	30	...	330	0	30	...	330
1	0 30 ⋮ 330	Energy values				Energy values			
		3				1			
		0	30	...	330	0	30	...	330
2	0 30 ⋮ 330	Energy values				Energy values			
		1				2			
		0	30	...	330	0	30	...	330
3	0 30 ⋮ 330	Energy values				Energy values			

SCHEME 2. Matrix example with the discrete values for each rotation angle.

PCA was performed on the data matrix, according to the first approach. The regions with minimum energy points on the grid search could be easily selected. The regions containing these selected points were subsequently refined, and PCA was performed again on these refined data. Further details of this methodology will be discussed in the next section.

By comparing both approaches, it may be observed that, in the second case, the number of points is larger than in the first one. However, the results will show that the selected regions are exactly the same.

All data matrices were constructed within MATLAB.[41] Principal component analysis was implemented by PIROUETTE.[45]

Results and Discussion

Basic Structure

Principal Component Analysis: The First Rotation

Three potential energy surfaces are generated for each angle studied. The results obtained are the energy values for the first dihedral pair rotations, with a 30° increment. The data matrix set up according to Scheme 1 contains all these energy surface values. Figure 2 shows these surfaces for the basic structure, where the angle coordinates indicate the energy involved in each angle rotation as a function of the two others.

The energy data matrix had their maximum pike values cutoff for a better visualization of the surfaces. For this basic structure, the cutoff value of 0.12 hartrees was chosen. The resulting surfaces and the respective contour plots are shown in Figure 3.

From Figures 2 and 3, it can be seen that the potential surface for angle 2 presents a major number of points with high energy, and it is the roughest one. The potential surface for angle 1 is not so rough, and there are more energy points with lower values. Finally, for angle 3, the potential surface has only a few points with high energy. However, some energy points also have high values.

The data matrix was autoscaled prior to principal component analysis (variables have a mean of zero and a variance of one). Initially, the analysis was performed for the matrix containing nonleveled data, and the result may be observed in Figure 4a. In this case, 64% of the original information is accumulated in the first and second principal components (or Factors). It is easy to note that the three curves converge to a unique region. Figure 4b shows the results corresponding to the matrix containing leveled data. In this case, the original information accumulated in both, first and second principal components, increases to 73%.

The first principal component (or Factor 1) is related to the energy gradient (Fig. 4b). An increase of energy is observed from left to right. The second principal component (or Factor 2) is related to the surface behavior, i.e., to the surface rugosity. Smoother surfaces show more positive score values in the second principal component.

These results can be readily understood through Figure 4b. One can observe that points related to the rotation of angle 2 are located on the right for the first principal component, while for angles 1 and 3 they are more to the left, indicating, in general, their lower energies. In the second principal component, angle 2 surface, having the highest rugosity, is separated from the others and located on the lower side of the diagram. The surface related to angle 1 is in the middle of the diagram, and finally we have the angle 3 surface, which is least rough, and so it is in the upper side of PCA diagram. The general behavior of each surface is easily understood from the contour diagrams shown in Figure 3.

A careful analysis of PCA scores in Figure 4 allows us to identify and select the regions of minimum energy. This was done using the following procedure:

1. For each angle, (α), the point (P_α) located at the extreme left side on the first principal component was selected, i.e., the lowest value point in this component.
2. Next, the closest points to P_α (with respect to PC1 axis) were taken into consideration. (a) When the selected points are adjacent to P_α ($\pm 30°$), then only this interval is taken for further refinement. (b) The points not chosen by (a) but also present in the left extreme, should also be taken into consideration. In this case, two intervals should be defined and submitted to further refinement.
3. To define the same matrix dimension in all cases, a 60° range was chosen for each separated region.
4. To ensure the capture of the minimum, an additional 10° was taken into account in the extremes of the 60° interval.

The chosen intervals for the basic structure are in Table 1. The first column shows how many regions were selected for each angle according to the PCA results in Figure 4. Two regions were chosen for angle 1 (a and b), while only one was selected for angles 2 and 3. The second column has the corresponding rotation range found from the initial value for each dihedral (see Fig. 4). The third column presents the initial values, and the fourth one has the values obtained from the sum performed on initial values according to the respective rotation selected in column 1.

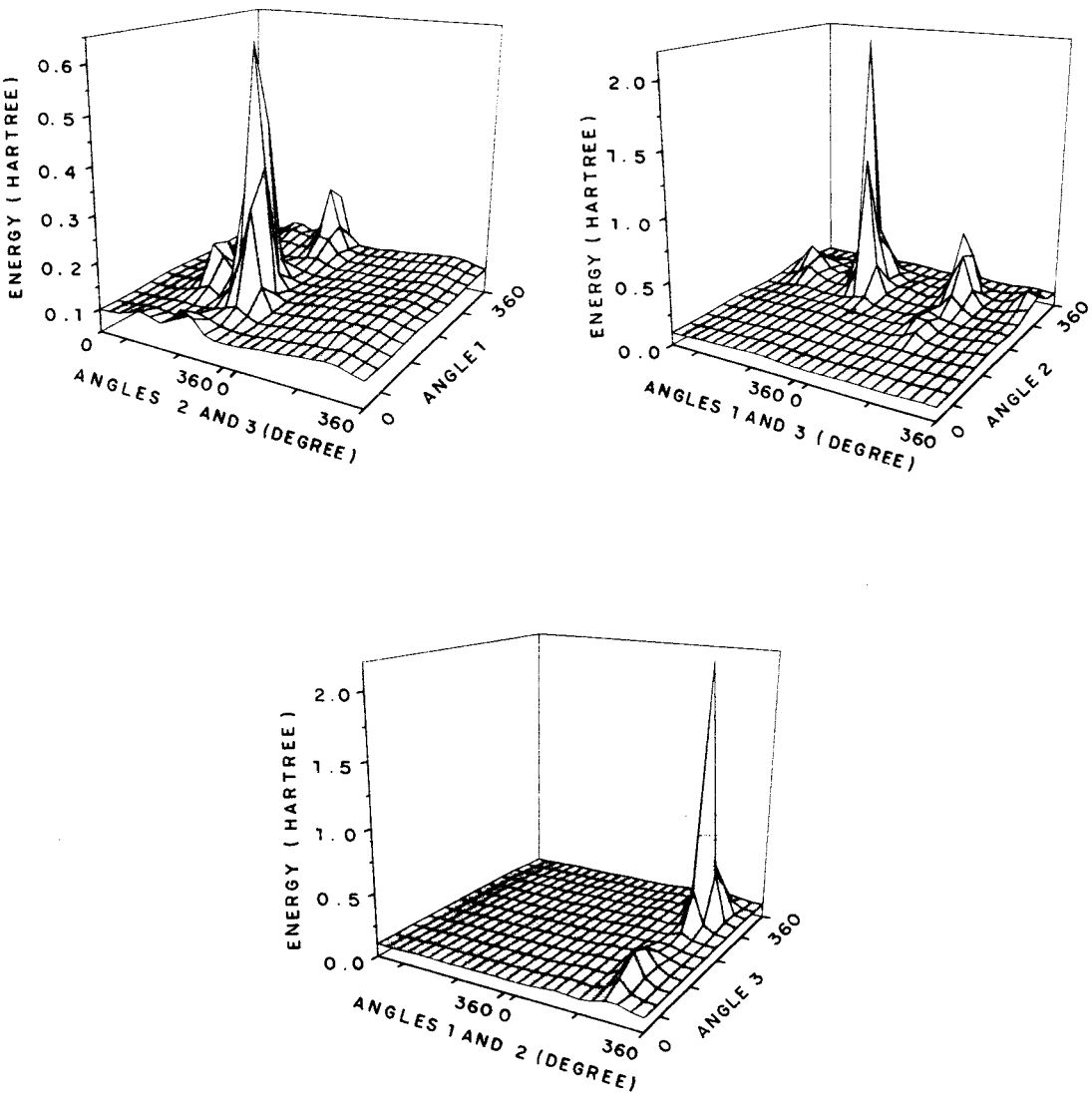


Figure 2. Potential energy surfaces for each angle in the basic structure.

Principal Component Analysis: Refinement

After the definition of the regions containing minimum energy structures, the refinement was performed on their corresponding energy surfaces according to Scheme 1 but using now an increment of 5°.

However, special attention must be paid to the new matrix dimension for the analysis in the refinement step by PCA. Because different regions were chosen for each angle, PCA must also be performed separately for each one (it can be analyzed in only one step when the dimension of data energy matrix is the same for each angle). That can be better explained by observing Scheme 3 where two minimum regions were selected for angle 1 (a and b) and only one for angles 2 and 3.

PCA score plots for the basic structure refinement are shown in Figure 5. It is interesting to note that all the plots obtained from PCA have a parabolic shape, as expected. With the decrease in the dihedral angle range, the first and second principal components describe the potential energy curves very clearly and efficiently. These curves converge to a minimum, which corresponds exactly to the minimum energy values. Table 2 shows the values corresponding to the rotations representing the minimum on each parabola, the 60° intervals obtained from the first rotation angle, and the final minimum energy dihedral angle. By combining these angles, two geometries were obtained, which correspond to the electronic structures of minimum energy. These geometries were submitted to final optimization by a semiempirical PM3 method leading to two final conformations.

The energy values for each minimum obtained are set out in Table 3, which also contains the dihedral angle obtained for each conformation. A single-point *ab initio* calculation was per-

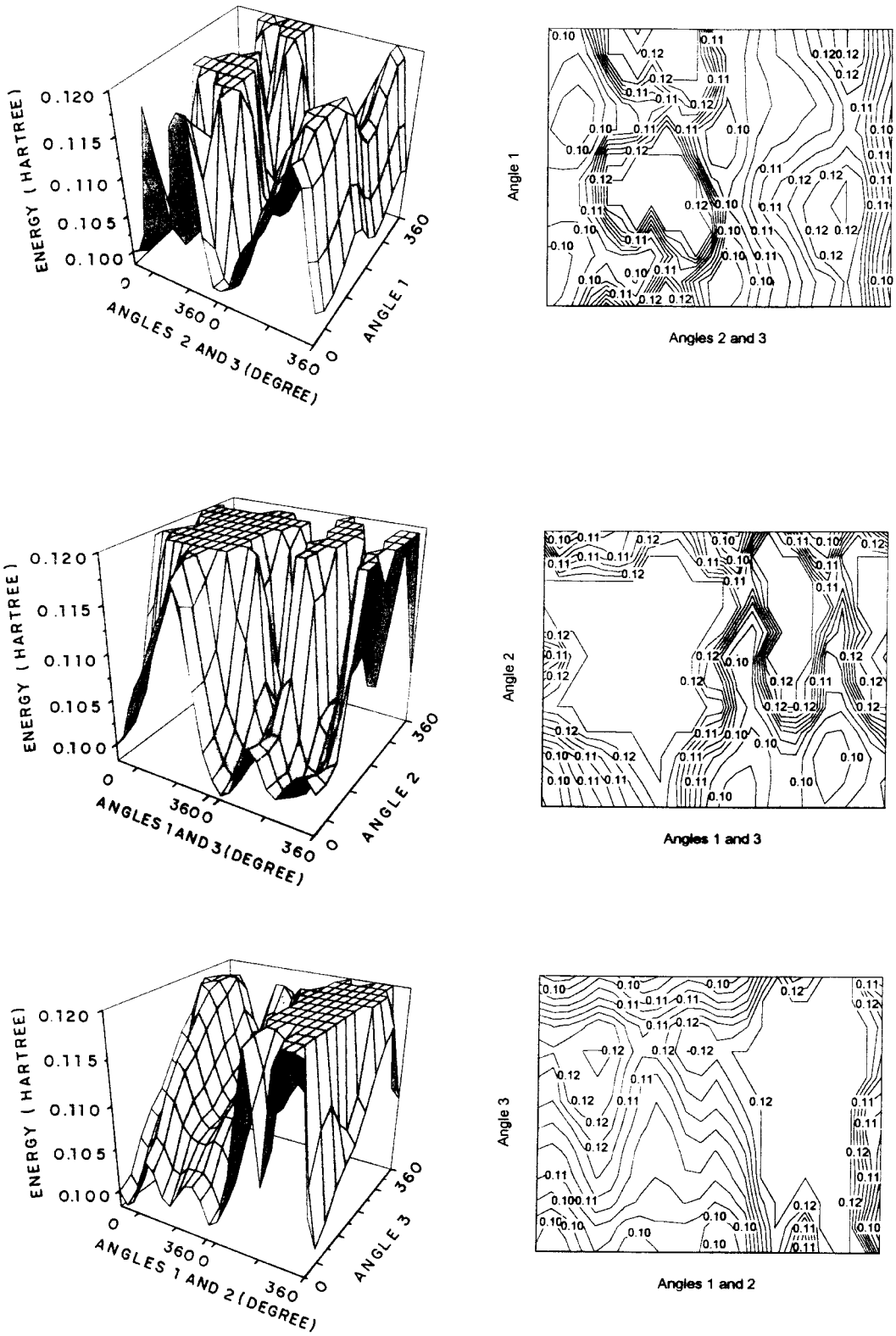


Figure 3. Potential energy surfaces for each angle in the basic structure, with energy values cutoff at 0.12 hartrees.

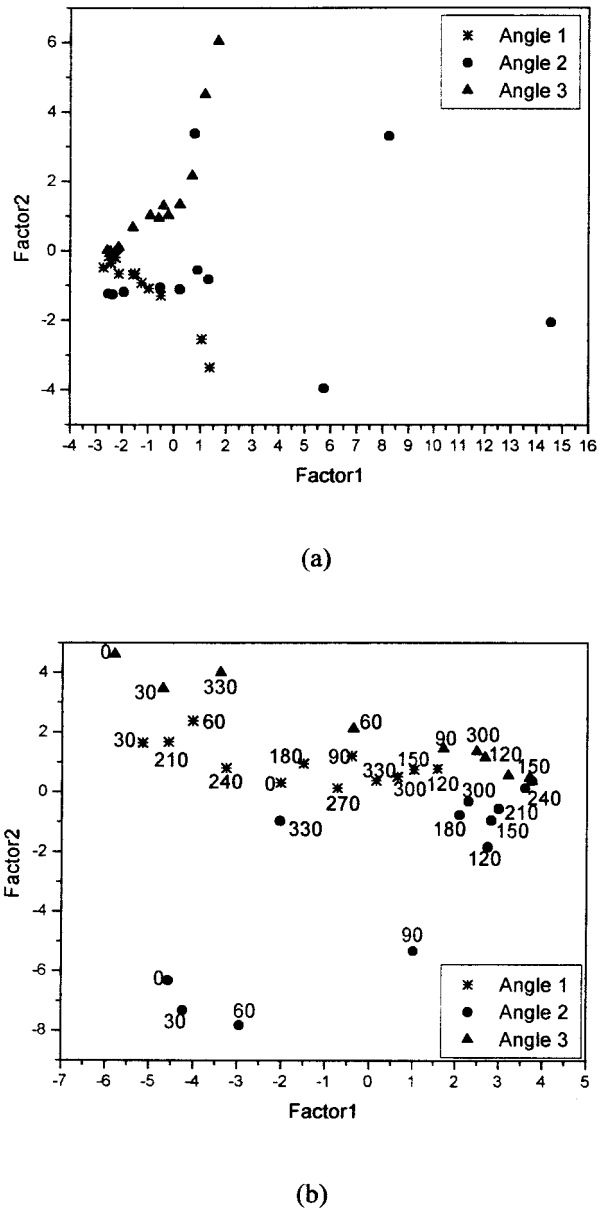


Figure 4. PCA results for the basic structure: (a) results for all potential energy surfaces data organized according to Scheme 1; (b) results for all potential energy surface values cutoff at 0.12 hartrees and organized according to Scheme 1.

formed at Hartree–Fock level using a 6-31G** basis set[46] on the PM3 optimized structures. The energy values are practically the same for both conformations. The optimized conformations for each electronic minimum found are presented in Figure 6.

A complete systematic search was also carried out for the basic structure, and the results are in perfect agreement with those obtained by our proposed methodology. Although not presented here, these results are available as Supplementary Material (item B).

Table 1. Regions Obtained by PCA for Each Angle (Basic Structure).

Angle	Corresponding Rotation	Initial Value	PCA Selected Region
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°

The same procedure was then applied to obtain global minimum energy for omeprazole, pantoprazole, and lansoprazole molecules. The results are summarized as follow.

Omeprazole

Principal Component Analysis: The First Rotation

According to the same procedure described above to select regions containing minimum energy structures, two regions for angles 1 and 3, and only one for angle 2 were found. Table 4 contains the related numerical results for omeprazole dihedral rotation angles with a 30° increment.

Principal Component Analysis: Refinement

After defining regions of minimum energy, the refinement was carried out, now with the smaller increment of 5°. Numerical data are presented in Table 5 for each angle.

By combining the dihedral angles, four minimum energy structures could be obtained when the final geometry optimization was performed. The corresponding angles and energy values are presented in Table 6, and the corresponding optimized geometries are shown in Figure 7.

It may be observed that the conformations obtained have very similar energy values. Another observation can be made comparing the dihedral angle values obtained after optimization (Table 6) with those from the combination of PCA values (Table 5). In every conformation, it can be seen that the angle 2 value presents a great variation from its predefined value in Table 5. This occurs because this angle is spatially located between the two others (Fig. 1), and its final value depends on how the angles 1 and 3 change after the optimization.

Table 7 shows the angle values for the X-ray strucure.[47] This structure was chosen for comparison with the optimized geometries because the R factor is 0.057 and the goodness of fit is 0.70. The estimated standard deviations for bond lengths and bond angles are

$$\begin{bmatrix} [E](1(a),2) & [E](1(a),3) \\ [E](1(b),2) & [E](1(b),3) \end{bmatrix}$$

(a)

$$\begin{bmatrix} [E](2,1(a)) & [E](2,1(b)) & [E](2,3) \\ [E](3,1(a)) & [E](3,1(b)) & [E](3,2) \end{bmatrix}$$

(b)

SCHEME 3. Schematic representation of cyclical permutation in energy data refinement: (a) Matrix representation for angle 1 data, where two different regions were selected on the grid search; (b) Matrix representation for angles 2 and 3 data, where a single region was selected on the grid search for each one.

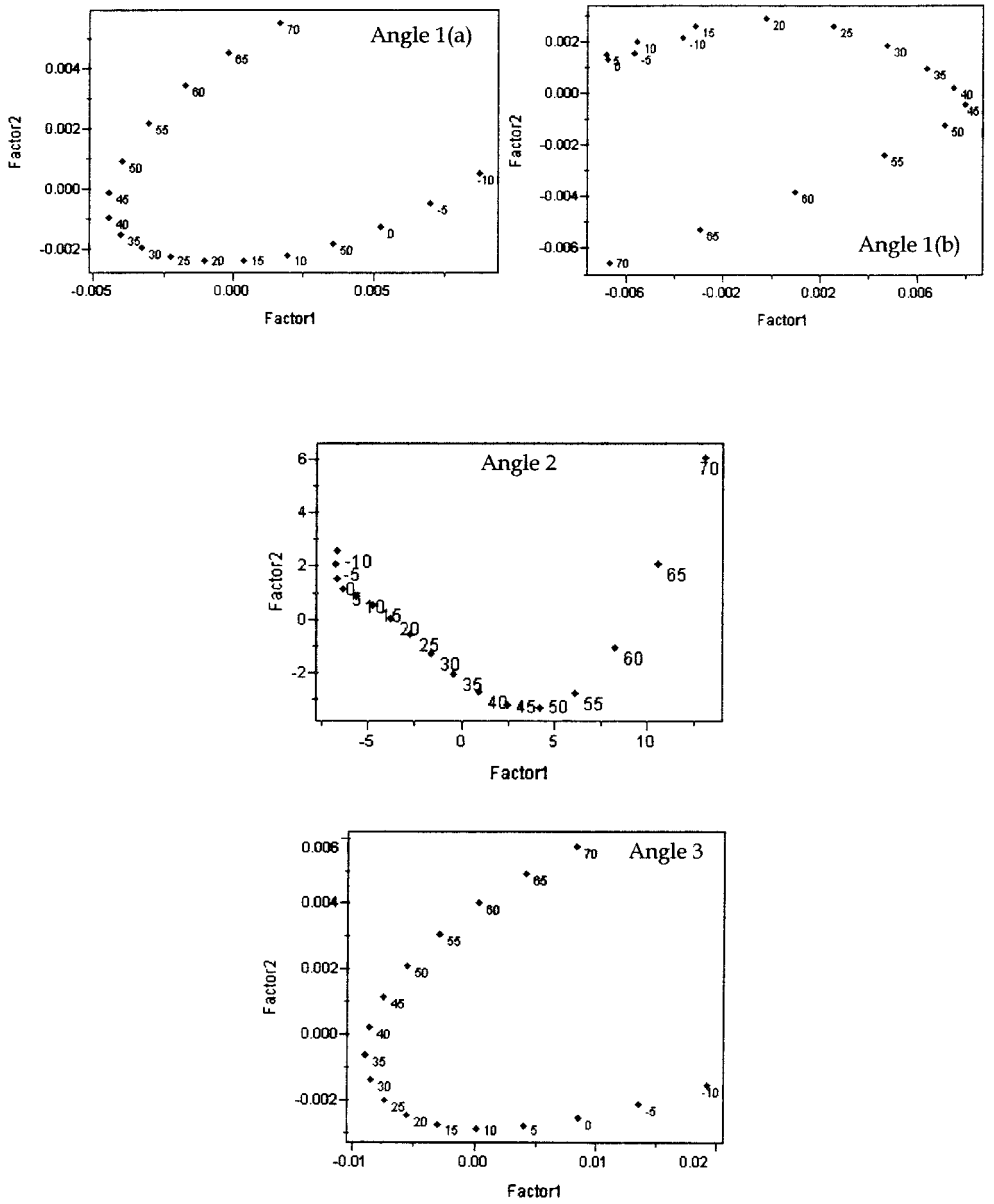


Figure 5. PCA results for selected regions in the basic structure refinement. The data are organized according to Scheme 3.

Table 2. Regions Obtained by PCA for Refinement (Basic Structure).

Angle	Corresponding Rotation	Initial Region	PCA Obtained Value
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 3. Characteristics of Minimum Conformations (Basic Structure).

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

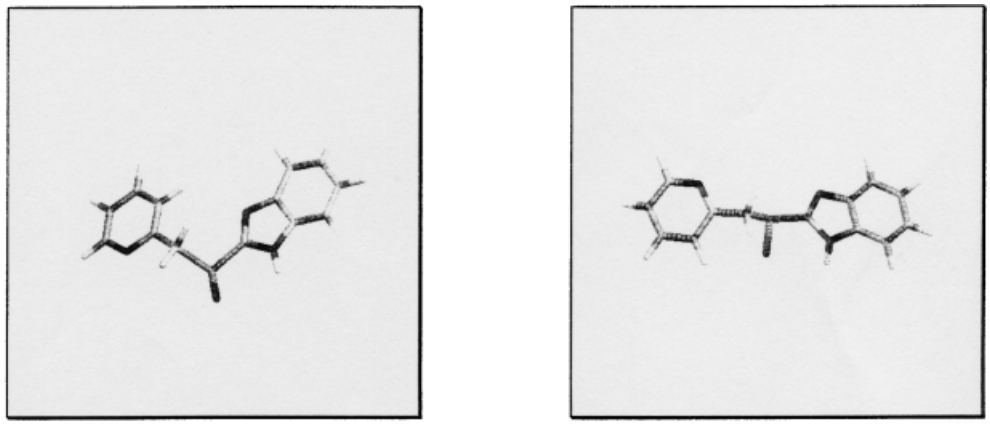


Figure 6. Optimized conformations for the basic structure.

0.004–0.010 Å and 0.02–0.04°, respectively. By simple observations of the angles from Tables 6 and 7 it cannot be seen how close the found and X-ray conformations are.

For better visualization, the X-ray structure was individually compared to the other conformations, and the results are presented in Figure 8. These compounds have an optical center located on the sulfur of the sulfoxide group. Figure 8b shows that conformation B is practically the mirror image of X-ray structure, i.e., one can say that this conformation is the X-ray structure enantiomer. On the other hand, conformation C (Fig. 8c) is practically superimposed on the to X-ray structure, and can be considered as being the same. Consequently, it is expected that conformation B will be the enantiomer of conformation C. These two conformations are compared in Figure 8e, confirming our expectations. In conclusion, it may be considered that these structures are the same, because the enantiomer properties are exactly the same. Thus, three minima conformations for omeprazole have been obtained.

Pantoprazole

Principal Component Analysis: The First Rotation

For the pantoprazole molecule, two regions for angles 1 and 3 were selected, while only one was chosen for angle 2 when the dihedral increment was 30°. Table 8 shows the PCA results.

Principal Component Analysis: Refinement

In this step, the regions selected above were refined with a dihedral angle increment of 5°. The results are summarized in Table 9. By combining the dihedral angle (Table 9), four minimum energy conformations were obtained. The related angles and respective energy values obtained after PM3 geometry optimization and *ab initio* calculation are presented in Table 10.

By examining the obtained data, some interesting conclusions can be drawn. Conformations C and D can be considered to be practically the same structure, because the dihedral angles obtained after optimization are very close to each other. The corresponding geometries are presented in Figure 9. Another interesting observation can be made from Figure 10, where both C (or D) and B conformations are compared. As in the case of omeprazole, these molecules form an enantiomer pair. Considering that conformations C and D are equal, and both are enantiomers of conformation B, it can be concluded that these three conformations correspond to a unique structure, assuming that enantiomers have the same physico-chemical properties. In this case, two conformations of minimum energy were obtained for pantoprazole. Unfortunately, there is no X-ray structure available in this case, and no comparisons can be made.

Table 4. Regions Obtained by PCA for Each Angle (Omeprazole).

Angle	Corresponding Rotation	Initial Value	PCA Selected Region
1(a)	180–240°	68.94°	248.94–308.94°
1(b)	300–0°	68.94°	8.94–68.94°
2	60–20°	37.22°	97.22–157.22°
3(a)	300–0°	27.98°	213.98–273.98°
3(b)	120–180°	27.98°	33.98–93.98°

Table 5. Regions Obtained by PCA for Refinement (Omeprazole).

Angle	Corresponding Rotation	Initial Region	PCA Obtained Value
1(a)	45°	248.94–308.94°	293.94°
1(b)	5°	8.94–68.94°	13.94°
2	40°	97.22–157.22°	137.22°
3(a)	35°	218.98–278.98°	248.98°
3(b)	25°	33.98–93.98°	58.98°

Table 6. Characteristics of Minima Conformations (Omeprazole).

Conformation	Angle	Obtained Value	$\Delta H_{f(PM3)}$ (kcal mol^{-1})	$E_{e(6\text{-}31G^{**})}$ (hartrees)
A	1	12.88°	−33.49	−1440.18
	2	151.48°		
	3	297.75°		
B	1	79.19°	−35.94	−1440.18
	2	181.75°		
	3	62.02°		
C	1	281.84°	−36.19	−1440.19
	2	171.89°		
	3	298.19°		
D	1	284.59°	−36.39	−1440.18
	2	81.88°		
	3	74.70°		

Table 7. X-ray Structure Characteristics.

Conformation	Angle	Angle Value
X-ray	1	326.34°
	2	179.08°
	3	238.68°

Lansoprazole

Principal Component Analysis: The First Rotation

While for the previous studies three dihedral angles were simultaneously studied, now four dihedral angles must be taken into account, because lansoprazole is far more complex. One more substituent (R_2 in Fig. 1) must be rotated in this molecule, due to the fact that it also causes steric effects. Thus, now there are four angles to be varied in pairs. Two regions for angles 1 and 3 were obtained. For angle 2 only one region was obtained while for angle 4 a broad

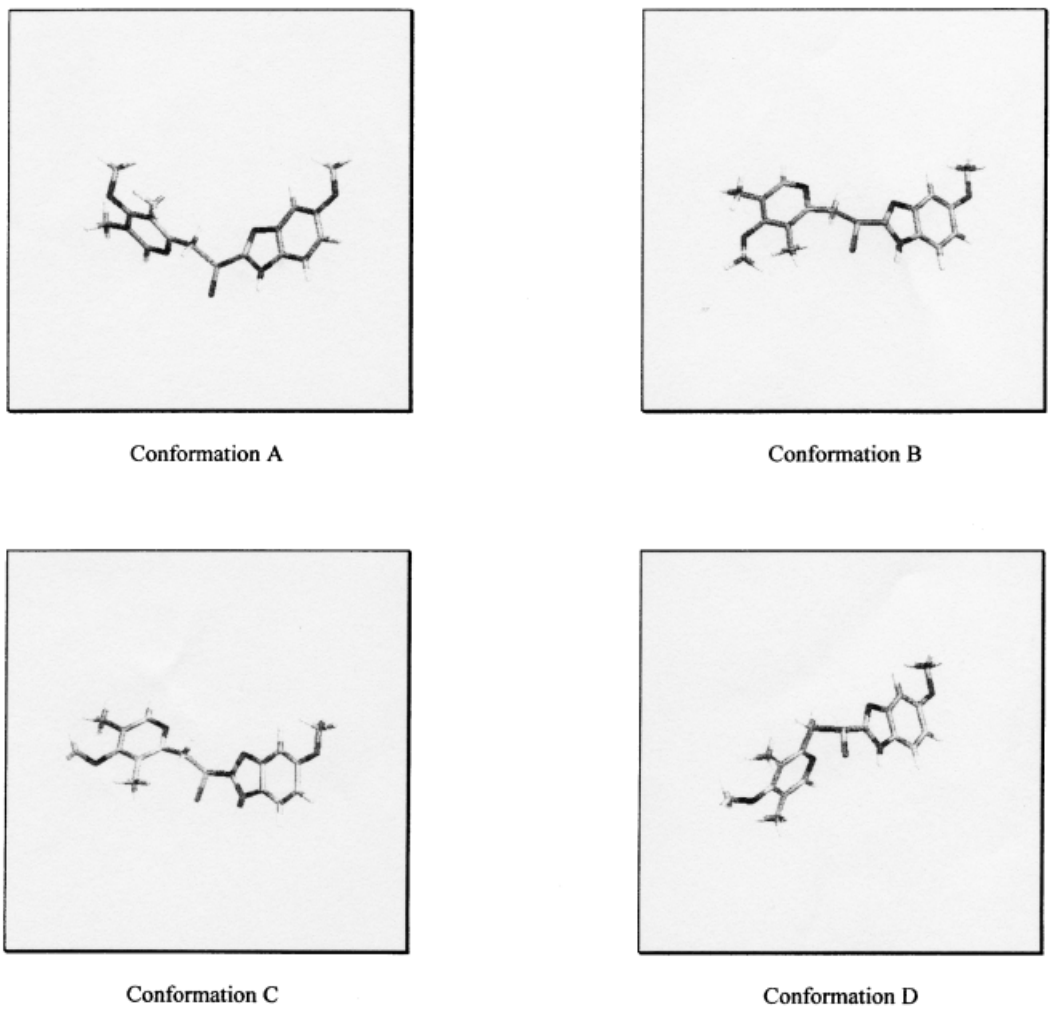


Figure 7. Optimized conformations for the omeprazole molecule.

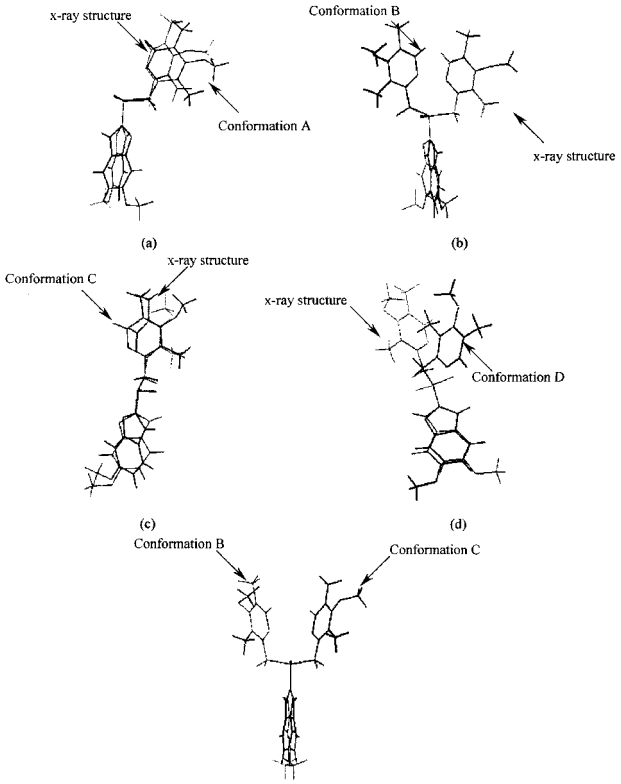


Figure 8. Omeprazole-optimized conformations and X-ray structure comparison: (a) superimposed conformation A and X-ray structure; (b) superimposed conformation B and X-ray structure; (c) superimposed conformation C and X-ray structure; (d) superimposed conformation D and X-ray structure; (e) superimposed conformations B and C.

region between 330° and 60° was selected in the first rotation. The results obtained for lansoprazole are summarized in Table 11.

One can observe that, for angle 4 relative to lansoprazole molecule, the separated region is larger than that 60° range defined earlier. To solve the problem, a central value was chosen that ranged from 345° to 45°. To perform the calculations, as mentioned above, an additional 10° over the extremes was taken account. Thus, the total variation can be considered as 335° and 55°, covering practically the entire interval for this angle.

Table 8. Regions Obtained by PCA for Each Angle (Pantoprazole).

Angle	Corresponding Rotation	Initial Value	PCA Selected Region
1(a)	300–0°	84.93°	24.93–84.93°
1 (b)	150–210°	84.93°	234.93–294.93°
2	60–120°	52.48°	112.48–172.48°
3(a)	330–30°	266.84°	236.84–296.84°
3(b)	90–150°	266.84°	356.84–56.84°

Table 9. Regions Obtained by PCA for Refinement (Pantoprazole).

Angle	Corresponding Rotation	Initial Region	PCA Obtained Value
1(a)	30°	24.93–84.93°	54.93°
1(b)	15°	234.93–294.93°	249.93°
2	35°	112.48–172.48°	147.48°
3(a)	25°	236.84–296.84°	261.84°
3(b)	60°	356.84–56.84°	46.84°

Principal Component Analysis: Refinement

Table 12 shows the numerical results obtained for lansoprazole refinement study, and four minimum energy conformations were found. The corresponding optimized structures for each of them are in Figure 11.

Angles and energy results are presented in Table 13. There is no X-ray structure reported for this molecule. However, as in the cases of omeprazole and pantoprazole, conformations B and C are optical isomers, as can be seen from the comparison made in Figure 12.

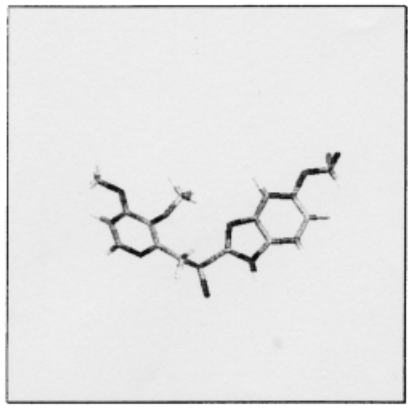
General Considerations

A few general conclusions can be drawn by observing the overall results. Despite the fact that positive values for heat of formation were obtained for the basic structure energy when using the semiempirical PM3 method, this does not mean that the conformations are not in a state of minimum energy, but shows that the basic structure is not stable without the substituents (experimentally, this basic structure has not been reported). This structure was only used to introduce the proposed methodology.

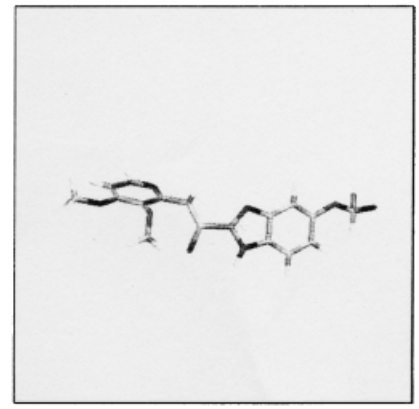
When comparing the conformations obtained for each molecule, it can be seen that all B and C conformations for the substituted compounds (omeprazole, pantoprazole, and lansoprazole)

Table 10. Characteristics of Minimum Conformations (Pantoprazole).

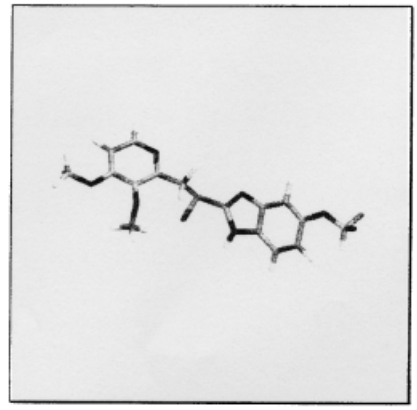
Conformation	Angle	Obtained Value	$\Delta H_{f(PM3)}$ (kcal mol^{-1})	$E_{e(6\text{-}31G^{**})}$ (hartrees)
A	1	49.37°	−157.67	−1673.69
	2	119.22°		
	3	295.36°		
B	1	68.93°	−158.57	−1673.70
	2	177.83°		
	3	62.82°		
C	1	291.09°	−158.57	−1673.70
	2	182.18°		
	3	297.05°		
D	1	290.96°	−158.74	−1673.70
	2	182.23°		
	3	297.22°		



Conformation A



Conformation B



Conformation C (D)

Figure 9. Optimized conformations for the pantoprazole molecule.

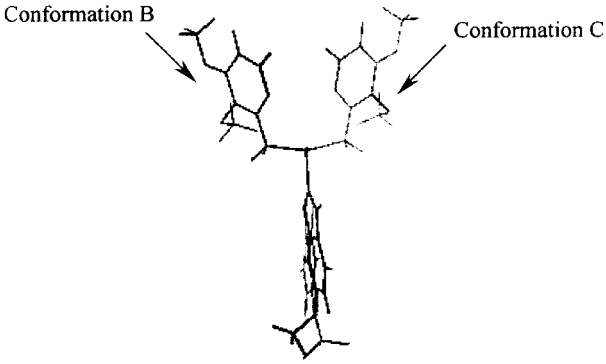


Figure 10. Pantoprazole-optimized conformations comparison: superimposed conformations B and C.

are very similar to those found for the basic structure. Moreover, for omeprazole, pantoprazole, and lansoprazole, the B and C conformations are optical isomers.

With respect to the energy values obtained for all conformations in the substituted compounds, the trend observed in the PM3 semi-

Table 11. Regions Obtained by PCA for Each Angle (Lansoprazole).

Angle	Corresponding Rotation	Initial Value	PCA Separated Region
1(a)	180–240°	65.02°	245.02–305.02°
1(b)	300–0°	65.02°	5.02–65.02°
2	60–120°	53.29°	113.29–173.29°
3(a)	330–30°	251.64°	221.641–281.64°
3(b)	120–180°	251.64°	11.64–71.64°
4	330–60°	67.35°	52.35–112.35°

Table 12. Regions Obtained by PCA for Refinement (Lansoprazole).

Angle	Corresponding Rotation	Initial Region	PCA Obtained Value
1(a)	15°	245.02–305.02°	260.02°
1(b)	45°	5.02–65.02°	50.02°
2	45°	113.29–173.29°	158.29°
3(a)	25°	221.64–281.64°	246.64°
3(b)	40°	11.64–71.64°	51.64°
4	20°	52.35–112.35°	72.35°

empirical method is the same as for the 6-31G** *ab initio* method. The basic structure has the highest energy, followed by omeprazole, lansoprazole, and finally pantoprazole. Based on these facts, it can be concluded that the substituents are responsible not only for stabilization of the basic structure, but also for the energetic difference between the studied compounds.

Conclusions

A new methodology of conformational analysis that controls the combinatorial explosion is proposed. Principal component analysis associated to quantum mechanic calculations is used to describe minimum structures energy. The methodology can be useful to handle small- and medium-size molecules. The maximum size which the method can efficiently handle is being investigated. 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. The method introduced was exemplified through the analysis of substituted benzimidazoles.

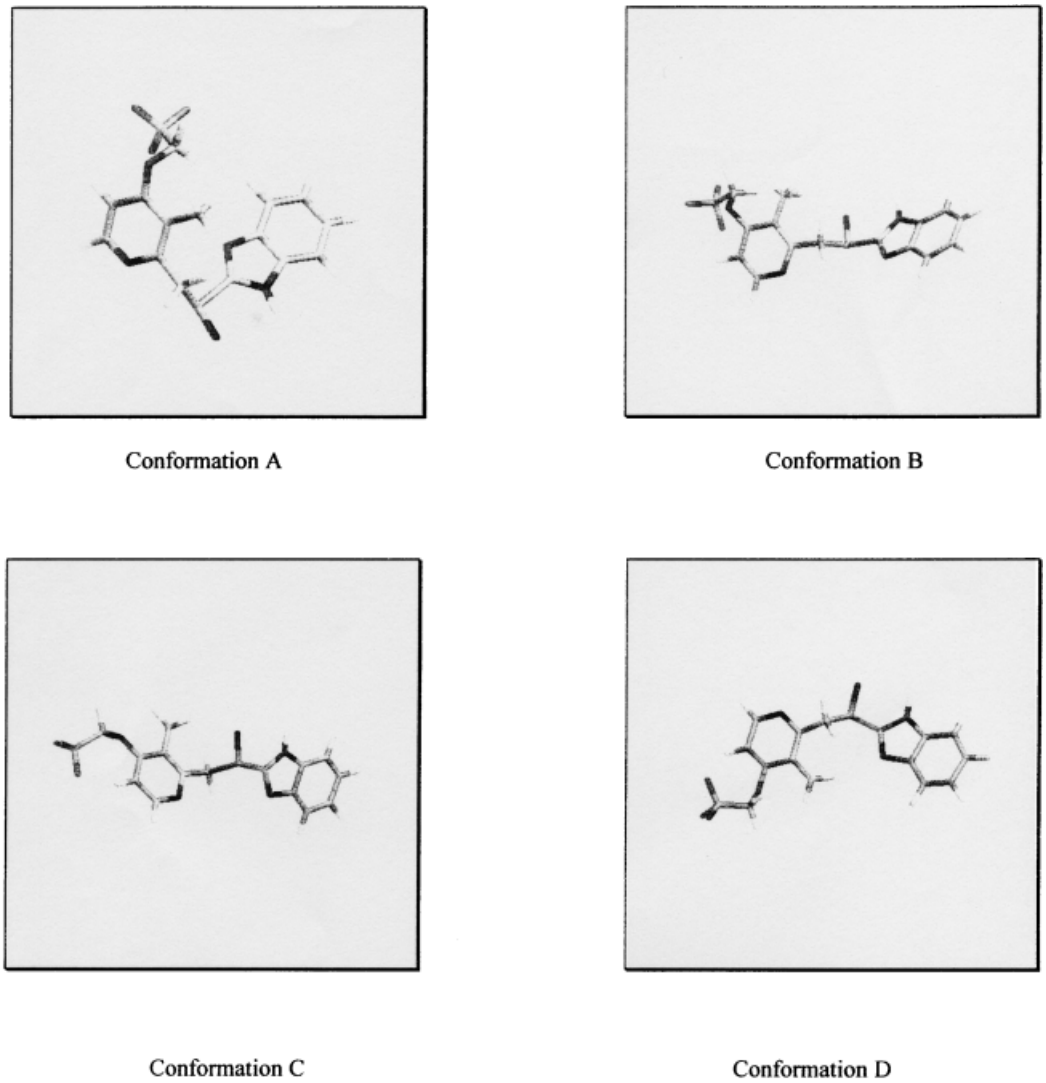


Figure 11. Optimized conformations for the lansoprazole molecule.

Table 13. Characteristics of Minimum Conformations (Lansoprazole).

Conformation	Angle	Obtained Value	$\Delta H_{f(PM3)}$ (kcal mol^{-1})	$E_{e(6\text{-}31G^{**})}$ (hartrees)
A	1	65.39°	−145.87	−1622.87
	2	80.29°		
	3	293.29°		
	4	89.38°		
B	1	81.32°	−146.10	−1622.88
	2	179.82°		
	3	61.12°		
	4	97.37°		
C	1	277.72°	−144.72	−1622.88
	2	170.00°		
	3	298.69°		
	4	96.00°		
D	1	276.50°	−144.01	−1622.88
	2	225.40°		
	3	63.55°		
	4	96.14°		

An important remark must be made when the number of minimum conformations found for basic structure is compared to those found for the substituted compounds. If the calculations were initially performed only for the basic structure and the substituents were subsequently added and followed by geometry optimization, all the compounds would only show two minimum energy conformations. A larger number of conformations could be found only when the analysis was performed in the substituted molecule from the beginning of the whole procedure. Thus, the addition of each substituent prior to the complete conformational analysis calculations is extremely important in such investigations.

The same methodology proposed here has been applied to a neolignanic derivative where four angles were taken into account and the results are in perfect agreement with those obtained from the complete systematic search. Applications of this method to a variety of larger molecules, including ring conformational search, are currently being carried out.

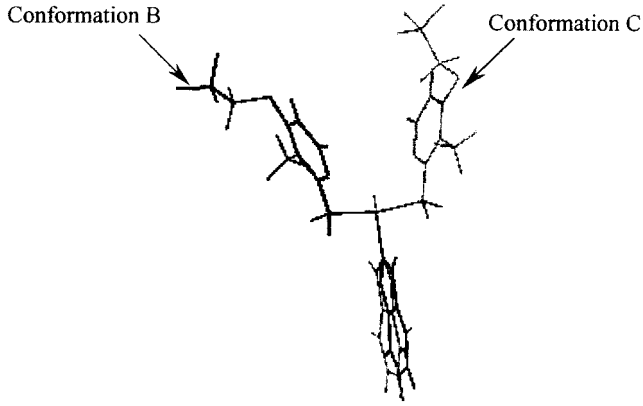


Figure 12. Lansoprazole-optimized conformation comparison: superimposed conformations B and C.

Acknowledgments

We thank CENAPAD-SP and Dr. Rogério Custódio for helpful discussions and Gaussian support.

References

1. Patrick, G. L. An Introduction to Medicinal Chemistry; Oxford University Press: Oxford, 1995.
2. Loew, G. H.; Villar, H. O.; Alkorta, I. Pharmaceut Res 1993, 10, 475.
3. Gasteiger, J.; Sadowski, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. J Chem Inf Comput Sci 1996, 36, 1030.
4. Bürgi, H. B. Acta Crystallogr A 1998, 54, 873.
5. Leach, A. R. Molecular Modelling; Addison–Wesley Longman: London, 1996.
6. Jensen, F. Introduction to Computational Chemistry; John Wiley & Sons: New York, 1999.
7. Becker, O. M. J Comput Chem 1998, 19, 1255.
8. Anet, F. A. L. J Am Chem Soc 1990, 112, 7172.
9. Kolossváry, I.; Guida, W. C. J Am Chem Soc 1993, 115, 2107.
10. Spellmeyer, D. C.; Wong, A. K.; Bower, M. J.; Blaney, J. M. J Mol Graph Model 1997, 15, 18.
11. Derks, E. P. P. A.; Beckers, M. L. M.; Melssen, W. J.; Buydens, L. M. C. Comput Chem 1996, 20, 439.
12. Jordan, S. N.; Leach, A. R.; Bradshaw, J. J Chem Inf Comput Sci 1995, 35, 640.
13. Derks, E. P. P. A.; Beckers, M. L. M.; Melssen, W. J.; Buydens, L. M. C. Comput Chem 1996, 20, 449.
14. Lucasius, C. B.; Kateman, G. Chem Intell Lab Sys 1994, 25, 99.
15. Jogensen, W. L.; Tirado-Rives, J. J Phys Chem 1996, 100, 14508.
16. Allen, M. P.; Tildesley, D. J. Computer Simulation of Liquids; Clarendon Press: Oxford, 1989.
17. Leach, A. R.; Smellie, A. S. J Chem Inf Comput Sci 1992, 32, 379.
18. Naidoo, K. J.; Brady, J. W. Chem Phys 1997, 224, 263.
19. Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. J Am Chem Soc 1990, 112, 1419.
20. Li, Z.; Scheraga, H. A. Proc Natl Acad Sci USA 1987, 84, 6611.
21. Beusen, D. D.; Sphands, E. F. B.; Karasek, S. F.; Marshall, G. R.; Dammkoehler, R. A. J Mol Struct (Theochem) 1996, 370, 157.
22. Allen, F. H.; Doyle, M. J. Acta Crystallogr B 1991, 47, 412.
23. Carugo, O. Acta Crystallogr B 1995, 51, 314.
24. Raithby, P. R.; Shields, G. P.; Allen, F. H. Acta Crystallogr B 1997, 53, 241.
25. Raithby, P. R.; Shields, G. P.; Allen, F. H. Struct Chem 1997, 8, 385.
26. Pastor, M.; Cruciani, G.; Clementi, S. J Med Chem 1991, 40, 1455.
27. Cecchetti, V.; Filipponi, E.; Fravolini, A.; Tabarrini, O.; Bonelli, D.; Clementi, M.; Cruciani, G.; Clementi, S. J Med Chem 1997, 40, 1698.
28. Nilson, J.; Wikström, H. J Med Chem 1997, 40, 833.
29. Becker, O. M. J Mol Struct (Theochem) 1997, 398, 507.
30. Becker, O. M. J Comput Chem 1998, 19, 1255.
31. Ajay, J Med Chem 1993, 36, 3565.
32. Sau So, S.; Karplus, M. J Med Chem 1996, 39, 1521.
33. Sau So, S.; Karplus, M. J Med Chem 1996, 39, 5246.
34. Lindberg, P.; Nordberg, P.; Alminger, T.; Brändström, A.; Wallmark, B. J Med Chem 1986, 29, 1329.
35. Tanaka, M.; Yamazaki, H.; Hakusui, H.; Nakamichi, N.; Sekino, H. Chirality 1997, 9, 17.
36. Landes, B. D.; Petite, J. D.; Flouvat, B. Clin Pharmacokinet 1995, 28, 1158.

37. Anderson, T. Clin Pharmacokinet 1996, 31, 9.
38. Wold, S.; Esbensen, K.; Geladi, P. Chem Intell Lab Sys 1987, 2, 37.
39. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. Chemometrics: A Practical Guide; John Wiley & Sons: New York, 1998.
40. Golub, G. H.; Van Loan, C. F. Matrix Computations; The Johns Hopkins University Press: London, 1989.
41. Matlab, Mathworks Inc., Natick, MA, 1998.
42. Hehre, W. J.; Huang, W. W. Chemistry with Computation: An Introduction to Spartan; Wavefunction Inc.: Irvine, California, 1995; Hehre, W. J.; Burke, L. D.; Shusterman, A. J. A Spartan Tutorial, Version 4.0; Wavefunction, Inc.: Irvine, California, 1995; Hehre, W. J. Practical Strategies for Electronic Structure Calculations; Wavefunction, Inc.: Irvine, California, 1995; Hehre, W. J.; Burke, L. D.; Shusterman, A. J.; Pietro, W. J. Experiments in Computational Organic Chemistry; Wavefunction, Inc.: Irvine, California, 1993.
43. Stewart, J. J. P. J Comput Chem 1989, 10, 209, 221.
44. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head–Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98, Revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.
45. Pirouette Multivariate Data Analysis for IBM PV Systems, Versio 2.0, Infometrics, Seattle, WA, 2000.
46. Clark, T. A Handbook of Computational Chemistry; John Wiley & Sons: New York, 1985.
47. Ohishi, H.; In, Y.; Ishida, T.; Inoue, M. Acta Crystallogr C 1989, 45, 1921.