ELSEVIER

Journal of Molecular Structure (Theochem) 536 (2001) 173–188

THEO CHEM

www.elsevier.nl/locate/theochem

Partially relaxed ring closure conditions for geometrical algorithm to search the conformational space for minimum energy conformations

L.N. Santagata, F.D. Suvire, R.D. Enriz*

Departamento de Quimica, Facultad de Quimica, Bioquimica y Farmacia, Universidad Nacional de San Luis, Chacabuco 917, 5700 San Luis, Argentina

Received 1 May 2000; accepted 19 June 2000

Abstract

We have reported a systematic search method, GASCOS, for open chain compounds. An extension to cyclic structures was recently reported. In this paper, we propose a more complete range of tolerance for some of the bond angles of the cyclic molecules studied. Specifically, we will accept that three consecutive bond angles associated with the ring closure condition can vary their values between certain previously fixed limits. We also introduce the concept of "family" which permit the "agglutination" of neighborhood configurations when their number is too large. Two small systems, cyclopentane and cyclohexane, were chosen to illustrate the GASCOS partially relaxed ring closure method in optimizing cyclic structures.

Keywords: Systematic conformational search; Analytical ring closure conditions; Minimum energy conformational; Cyclic compounds; Conformational space search

1. Introduction

Computer simulation has become an important tool in studying the properties and behaviors of complex biological molecules [1–3]. The use of "computer-assisted molecular design" (CAMD) [4] in displaying the three-dimensional structure of molecules is of invaluable help. It allows researches to quickly grasp the essential features of interactions involving numerous components of the complex. Molecular computation is a powerful tool for the analysis and understanding of the intimate mechanisms involved in drug-action.

One of the problems associated with CAMD is the identification of a representative set of conformations of a molecule. The principal problem is related to the uncertainty over how to account for receptor flexibility without drastically increasing the computational cost required. Obviously, one can define many receptor "snapshots" and run through each in turn. However, it must be taken into account that even small changes in the active site of a receptor in response to the ligand, e.g. rotation of an

* Corresponding author.

E-mail addresses: lnsantag@unsl.edu.ar (L.N. Santagata), fsuvire@unsl.edu.ar (F.D. Suvire), denriz@unsl.edu.ar (R.D. Enriz).

0166-1280/01/$ - see front matter
PII: S0166-1280(00)00630-8

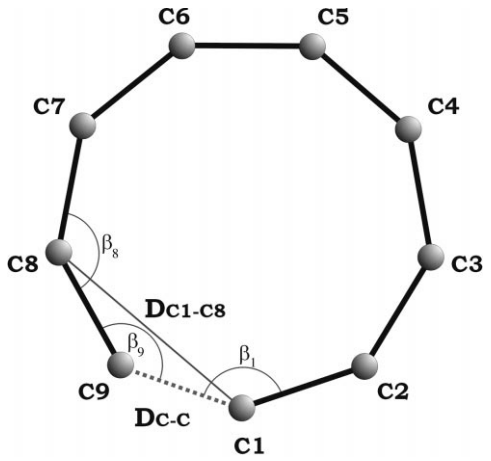

Fig. 1. Ring closure procedure for cyclononane with specific emphasis given to $D_{C\text{-}C}$ (C_1–C_9 bonding distance) and $D_{C\text{-}C}$ (C_1–C_8 non-bonding distance), and the bond angles β_1, β_8 and β_9.

hydroxymethylene group in serine residue [5–7], sulphydril methylene group in cysteine [8] or of the carboxylate methylene in an aspartate [9], will have a stabilizing effect on the ligand, but this is not taken into account in most of the fixed receptor models. Such changes are too numerous and are considered to be too small in their effect to be worthy of investigation with multiple snapshots. Consequently, they are not generally considered in design strategies.

When studying molecular conformations, there is a tendency for chemists to be preoccupied with finding the lowest energy conformation of a molecule. However, it should be kept in mind that the most flexible molecules in solution, as well as in the gas phase, exist in a variety of conformations. Also, when one molecule interacts with another, such as an inhibitor with an enzyme, it is not necessarily the lowest energy conformation of each molecule that occurs in the resulting complex. The conformation that a molecule adopts at a binding site may be several kilocalories above the global energy minimum of the isolated molecule [10]. It is often desirable to select a manageable number of conformations of a molecule to represent its accessible conformational space.

Global optimization methods aim to identify the structure, which is presumed to be the solution of a given problem. Methods such as simulated annealing and genetic algorithms are two particularly popular examples being currently applied in molecular modeling. Most global optimization methods are stochastic: they inherently rely upon a random exploration of the search space, albeit with sophisticated mechanisms for directing the search towards optimal solutions. However, random-based methods provide no guarantee to find the optimal solution required. This is particularly true if all conformations that are located only within some energy cut-off (e.g. 5 kcal/mol above the global minimum) are considered.

The alternative to the stochastic methods is the systematic search. However, these methods have an Achilles heel due to the combinatorial explosion in the number of possibilities.

Thus, the systematic search is often believed to be impractical for problems of biological interest.

Recently, we reported a systematic search method [11], which is particularly useful for flexible compounds. More recently, we introduced analytical ring closure conditions [12] for the study of cyclic molecules. However, taking into account that the search of cyclic conformations is carried out from rigid rotations of the atoms using discrete steps of $\Delta\theta$, the final number of conformations to be obtained might be too restrictive.

One of the aims of the work reported herein is to show that it is practical to systematically explore the conformational space of complex flexible cyclic compounds.

In this paper, we propose a more complete range of tolerance for some of the bond angles of the cyclic molecule studied. Specifically, we will accept that three consecutive bonds angles associated with the ring closure can vary their values between certain previously fixed limits.

The extension of this procedure does introduce a significant increase in the number of geometries for the cyclic configurations. Thus, in order to maintain a "manageable" number of configurations to be optimized, we also introduce a new concept. This is the concept of "families" which permit the "agglutination" of neighborhood configurations. From these "families", we obtain the "seeds" which are subsequently used for the geometrical optimizations using Molecular Mechanics and ab initio calculations.

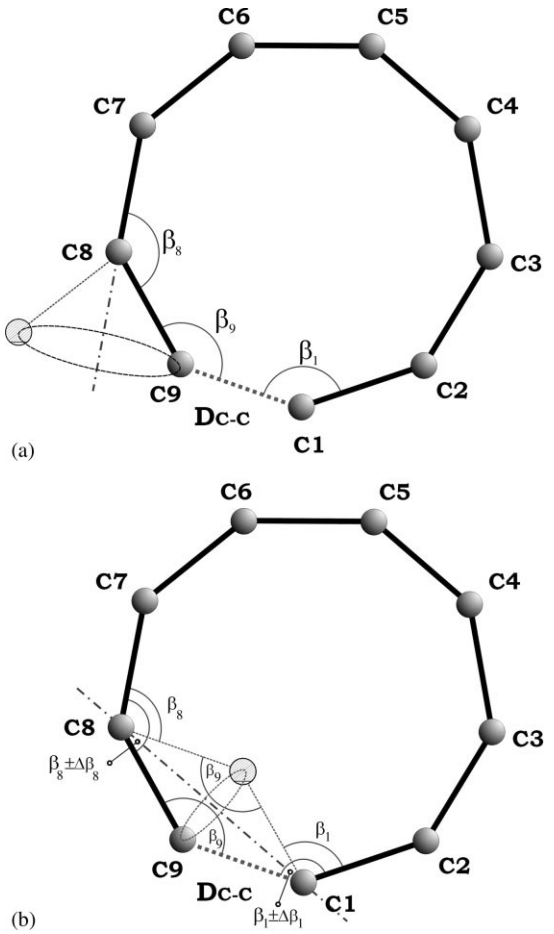

Fig. 2. Ring closure procedure for cyclononane showing the rotation of C_9: (a) rotational cone along the bond C_7–C_8 which is associated with the determination of the last torsional angle; (b) second rotation of C_9 along the C_1–C_8 direction.

2. Mathematical background[1]

2.1. Ring closure conditions for cyclic molecules

Recently, we reported the basic equations needed to establish the analytical conditions used to find the torsional angles that determine the cyclization of a molecule [12]. We have chosen cyclononane as an example. Thus, the location the last atom (C9), at C–C bond distance from the first atom (C1), result in two possible values for the torsional angles, i.e. the rotation about the C7–C8 bond.

However, this is not the only condition. We need the bond angle β_1, formed by the C9–C1–C2 atoms, to be located in a range of tolerance, $\Delta\beta_1$, with respect to the characteristic tetrahedral bond ($\beta_{\text{teth}} = 109.5°$).

The restriction of the C1–C9 distance and the β_1 range in the interval ($109.5° - \Delta\beta_1$, $109.5° + \Delta\beta_1$), predetermine all available molecular configurations (Fig. 1).

The remaining bond angles are fixed 109.5°, as the condition of the initial configuration constructed.

In the present paper, we propose to introduce a new tolerance interval on the next bond angle to be formed by the C8–C9–C1 atoms (denoted β_9 in Fig. 1).

The value of this bond angle β_9 is fixing the non-bonding distance $D_{\text{C1–C8}}$ between atoms C8···C1 because the separation between atoms C8–C9 and C9–C1 are equivalent to the bond length of C–C.

Fixing a tolerance $\Delta\beta_9$ to the bond angle β_9 as follows:

$$(109.5 - \Delta\beta_9,\ 109.5 + \Delta\beta_9), \tag{1}$$

it is obvious that the separation distance between the atoms C8···C1 can vary from a maximum value (for $\beta_9 = 109.5 + \Delta\beta_9$) to a minimum value (for $\beta_9 = 109.5 + \Delta\beta_9$).

It is easy to obtain these distances, applying the cosine theorem:

$$D_1 = \sqrt{2D_{\text{C–C}}^2[1 - \cos(\beta_9 - \Delta\beta_9)]}, \tag{2a}$$

$$D_2 = \sqrt{2D_{\text{C–C}}^2[1 - \cos(\beta_9 + \Delta\beta_9)]}. \tag{2b}$$

Thus, it is possible to locate atom C8 at a variable distance ($D_{\text{C1–C8}}$) with respect to atom C1 satisfying the following inequality:

$$D_1 \le D_{\text{C1–C8}} \le D_2. \tag{3}$$

It is important to note that the above procedure is quite different with respect to the previously used method, in which the considered β_9 is fixed. In the previous case, only one $D_{\text{C1–C8}}$ distance existed as a fixed value. In contrast, satisfying inequality (3) will permit us to obtain a variation of β_9 as a function of the tolerance fixed by $\Delta\beta_9$ (Eq. (1)). It should be emphasized that in order to obtain a cyclic structure, it is also necessary to

[1] All vectorial quantities are denoted, throughout this paper, by bold letters.

satisfy all previous conditions imposed on the C9 atom.

In addition, it is possible to impose an analogous tolerance interval condition on the bond angle β_8 defined by the atoms C7–C8–C9 (see Fig. 1). The first step is to locate the C8 atom satisfying condition (3). Then, we rotate C9 about the C7–C8 bond in order to obtain a C1–C9 distance equal to a characteristic C–C bond (Fig. 2a). The line joining the atoms C1–C8 is determining a new direction of rotation for the C9 atom (see Fig. 2b).

It can be observed that during the rotation about this new direction along the line C1···C8, the angle β_9 is maintained constant but the angles β_1 and β_8 are changing.

Satisfying the following conditions:

$$\beta_1 \in (109.5^\circ - \Delta\beta_1,\ 109.5^\circ + \Delta\beta_1), \tag{4a}$$

$$\beta_8 \in (109.5^\circ - \Delta\beta_8,\ 109.5^\circ + \Delta\beta_8) \tag{4b}$$

and fixing a priori the respective tolerances $\Delta\beta_1$ and $\Delta\beta_8$, we obtain a number of possible configurations of the cyclic molecule.

It should be noted that the two remaining conditions:

$$D_{C1\text{–}C9} = D_{C\text{–}C}, \tag{5a}$$

$$\beta_9 \in (109.5^\circ - \Delta\beta_9,\ 109.5^\circ + \Delta\beta_9), \tag{5b}$$

were fixed in the previous steps.

In conclusion, the general ring closure conditions are:

1. The distance between the first and last atom fixed at a characteristic bond length.
2. The tolerance interval conditions $\Delta\beta_1$, $\Delta\beta_8$ and $\Delta\beta_9$ for the respective bond angles are also fixed. (For convenience different tolerance values might be chosen.)
3. The remaining bond angles are chosen from the standard values (120 or 109.5°), depending on whether the atoms are sp^2 or sp^3 hybridized carbon.

2.2. Maximum and minimum distances in a rotation

During a rotation for any atom there is a maximum proximity and maximum separation with respect to any fixed atom in the molecule. These separations are defining the maximum and minimum distances in a complete revolution of the torsional mode of motion.

The evaluation of the range of distances between minimum and maximum separation is very important to avoid the superposition of atoms during the rotation but, it is also important to confirm the fulfillment of condition (3). It is possible to know, in an analytic form, the values of the minimum and maximum distances and their respective torsional angles for which they appear.

An equation for the determined distances was recently reported in Ref. [11]. Here we use it to determine the variation of the distance between a fixed atom (i.e. C1 in Fig. 1) and another moving atom (i.e. C8 in Fig. 1)

$$D_{C1\text{–}C8} = \sqrt{|\mathbf{R}_8^i - \mathbf{R}_1|^2 + 2(\mathbf{R}_8^i - \mathbf{R}_1 - \mathbf{r}_8^i)\cdot\mathbf{r}_8^i(\cos\theta_8 - 1) + 2(\mathbf{R}_8^i - \mathbf{R}_1)\cdot\mathbf{P}_8^i \sin\theta_8} \tag{6}$$

where

$\mathbf{R}_8^i$ = initial position vector for atom C8, $\mathbf{R}_1$ = position vector of atom C1 (fixed).
$\mathbf{r}_8^i$ = initial revolution radius vector of C8 around the $\hat{n}$ direction pointing along the C6–C7 bond, associated with the previous rotation by O.
$\mathbf{P}_8^i = \hat{n} \times \mathbf{r}_8^i$ = cross product vector between the direction of revolution vector $\hat{n}$ (determined by the C6–C7 atoms) and the initial revolution radius vector $\mathbf{r}_8^i$ (Fig. 3).

If we identify $\mathbf{R}_8^i$ with $\mathbf{R}_C^i$, and consider that the C1 atom is at the origin of the coordinates system (i.e. $\mathbf{R}_1 = 0$), then Eq. (6) becomes the same as the previously reported Eq. (40b) in Ref. [11].

Defining the new constants A, B and C as follows:

$$A = 2(\mathbf{R}_8^i - \mathbf{R}_1)\cdot\mathbf{P}_8^i, \tag{7a}$$

$$B = 2(\mathbf{R}_8^i - \mathbf{R}_1 - \mathbf{r}_8^i)\cdot\mathbf{r}_8^i, \tag{7b}$$

$$C = |\mathbf{R}_8^i - \mathbf{R}_1|^2, \tag{7c}$$

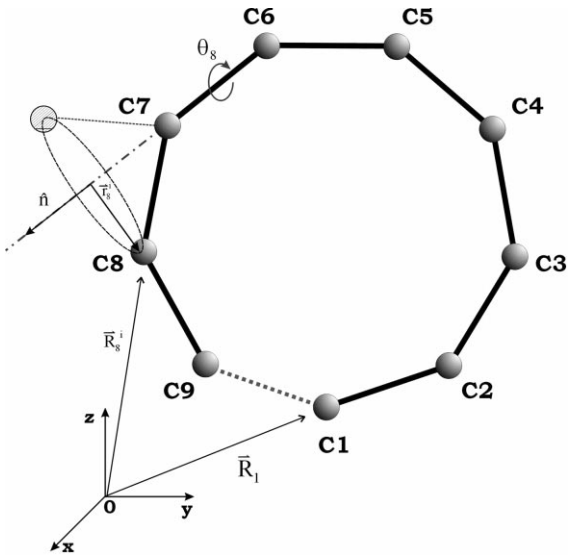

Fig. 3. Spatial view of the vectors defined in Eq. (6).

we can now rewrite Eq. (6) in the following form:

$$D_{C1\text{–}C8} = \sqrt{C + B(\cos\theta_8 - 1) + A\sin\theta_8}. \quad (8)$$

Considering the extreme condition $dD_{C1\text{–}C8}/d\theta_8 = 0$, it is possible to solve (9) for the θ_8 values:

$$\frac{dD_{C1\text{–}C8}}{d\theta_8} = \frac{A\cos\theta_8 - B\sin\theta_8}{2D_{C1\text{–}C8}} = 0. \quad (9)$$

Eq. (10a) determines the value of tg θ_8, in terms of constants A and B:

$$\text{tg}\,\theta_8 = \frac{A}{B}. \quad (10a)$$

Although Eq. (10a) is derived for the eighth torsional angle (θ_8), the relationship is valid for any torsional angle as defined for Eq. (10b):

$$\text{tg}\,\theta = \frac{A}{B} \quad (10b)$$

provided that A and B are defined correspondingly.

In a complete range (0°, 360°), there are two values for the angle having the same tangent. One of them corresponds to the value of minimum distance and the other to the maximum distance. We designate them as θ_{min} and θ_{max}.

Replacing θ_8 in Eq. (8) by the values of θ_8^{min} and θ_8^{max} obtained, it is possible to calculate the minimal and the maximal separation distances for the C8 atom with respect to the C1 atom:

$$(D_{C1\text{–}C8})_{min} = \sqrt{C + B(\cos\theta_8^{min} - 1) + A\sin\theta_8^{min}}, \quad (11a)$$

$$(D_{C1\text{–}C8})_{max} = \sqrt{C + B(\cos\theta_8^{max} - 1) + A\sin\theta_8^{max}}. \quad (11b)$$

These extreme values of distances permit us to know the range of interatomic distance in any rotation according to the following inequality:

$$(D_{C1\text{–}C8})_{min} \le D_{C1\text{–}C8} \le (D_{C1\text{–}C8})_{max}. \quad (12)$$

It should be mentioned that this inequality (12), which is given in terms of the torsional angle θ, represents a different condition for bracketing $D_{C1\text{–}C8}$ than the previous inequality (3), which was based on fixed values of bond lengths, bond angle β and bond angle tolerance $\Delta\beta$.

This point will be discussed in the next section. It will be seen that both conditions (3) and (12) must be met for successful ring-closure.

2.3. Interpreting the range of torsional angles in terms of the minimum and maximum distances

It is important to determine the extreme values of distances between two atoms, in order to know at which range of torsional angles it is possible to find ring-closed configurations.

Fig. 4 represents a complete revolution with a circumference in which the minimum (D_{min}) and maximum (D_{max}) distances are located at the angles θ_{min} and θ_{max}, respectively.

Both angles (θ_{min} and θ_{max}) have the same value of tangent, therefore they cut the circumference into two halves. As shown in Fig. 4, the distance D_{min} corresponds to θ_{min} and D_{max} corresponds to θ_{max}. The magnitude of the distances change according to the following pattern:

1. When θ changes from θ_{min} to θ_{max}, the distances increase.
2. When θ changes from θ_{max} to θ_{min}, the distances decrease.

It is implicit in this figure that in a complete

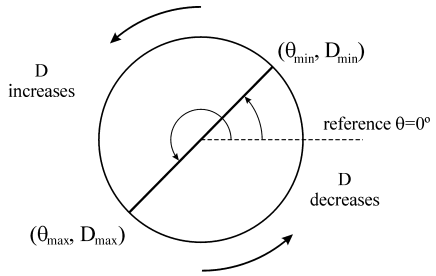

Fig. 4. Complete revolution showing the location of minimum and maximum distances. This figure also illustrates Case I.

revolution, the interatomic distances D satisfy the following relations:

$$D_{\min} \leq D \leq D_{\max}, \tag{13}$$

$$D_1 \leq D \leq D_2. \tag{14}$$

If we are interested in distances that are included in the previously determined interval of condition (3): $D_1 \leq D \leq D_2$ (for example the distances between C1 and C8 atoms in cyclononane), different situations are possible for the range of interatomic distances. These are presented in terms of (13) and (14) as five different cases:

Case I: If $D_2 < D_{\min}$, then there are no torsional angles which determine an interatomic distance in the range $D_1 \leq D \leq D_2 < D_{\min}$. This case corresponds to Fig. 4.

Case II: If the opposite case is operative, namely $D_1 \leq D_{\min} \leq D_2 \leq D_{\max}$, then there are interatomic distances located in a interval $D_{\min} \leq D \leq D_2$. The range for the rotation is shown in Fig. 5 as a zone marked by a bold - - - line. The torsional angles that correspond to interatomic distances greater than D_2 are included in the non-shaded arc of the circumference line in Fig. 5, possessing D_2 as extremes and including $D_{\max}$.

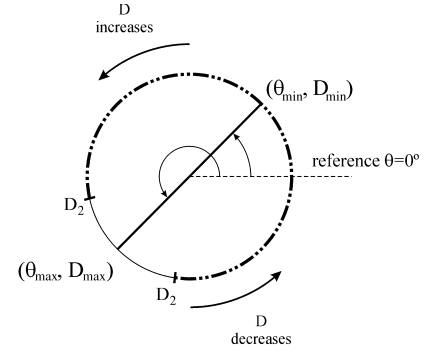

Fig. 5. Range of rotation (Case II), denoted by the bold - - - line.

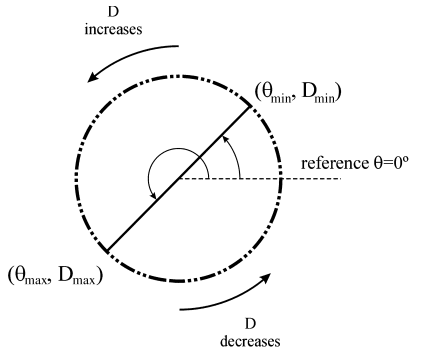

Fig. 6. Range of rotation (Case III), denoted by the bold - - - line.

Case III: If $D_1 \leq D_{\min} \leq D_{\max} \leq D_2$, then there are possible distances located in the entire range of θ (from $\theta = 0°$ to $\theta = 360°$), because Eqs. (13) and (14) are always satisfied (Fig. 6).

Case IV: If $D_{\min} \leq D_1 \leq D_2 \leq D_{\max}$, then there are two zones of angles which are determining the possible interatomic distances. These intervals are shown in Fig. 7 as zones marked by a bold - - - line. The two complete non-shaded arcs of revolution S_1 and S_2 having the extremes D_1 and D_2, respectively, are the zones of dihedrals which are not accessible for the interatomic distances required by Eqs. (13) and (14). The unmarked arcs of revolution in Fig. 7 with D_1 and D_2 extremes containing $D_{\min}$ and $D_{\max}$, respectively, represent torsional

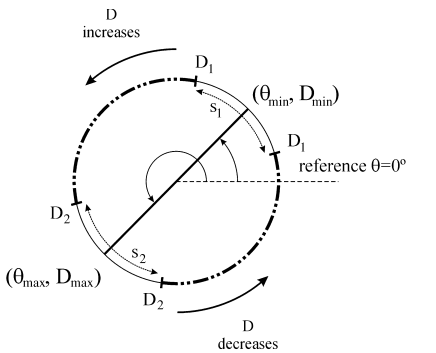

Fig. 7. Range of rotation (Case IV), denoted by the bold - - - line.

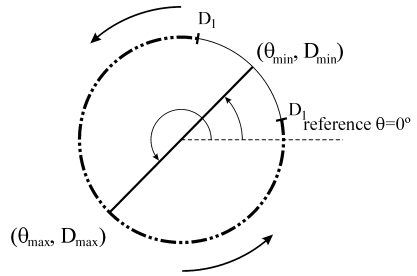

Fig. 8. Range of rotation (Case V), denoted by the bold - - - line.

angles resulting in distances out of the range of interest.

Case V: If $D_{min} \leq D_1 \leq D_{max} \leq D_2$, then there is one zone of possible angles (see shadowed zone of Fig. 8). The non-shaded arc of revolution in Fig. 8 with D_1 extremes containing D_{min}, represents torsional angles resulting in distances out of the range of interest.

It should be noted that the fact that a distance is located in a determined range does not necessarily mean that we have an acceptable molecular configuration. In order to obtain an acceptable molecular configuration, it is also necessary to satisfy the conditions imposed on the bond angles β and the ring closure condition for the last atom in order to locate it within an acceptable interatomic distance.

Using the basic equations reported previously [12]:

$$[1 + (A^2/B^2)]X^2 - (2AC/B^2)X + (C/B)^2 - 1 = 0, \quad (15)$$

where we set $X = \sin\theta$ and it is possible to determine the torsional angles which can determine the D_1 and D_2 distances of interest and then to vary the torsional angle θ in the range established for the above cases. Maintaining a constant number of steps in the rotation in those cases in which the range of the torsional angles is reduced results in a lower value for the angular increment. It is necessary to apply exactly the same criteria as above in order to avoid that a rotating atom is superimposed with another non-bonding atom. Thus, imposing a minimal distance condition, it is possible to avoid any configuration that is chemically not acceptable (for example the distance for non-bonded carbons lower than 1.53 Å).

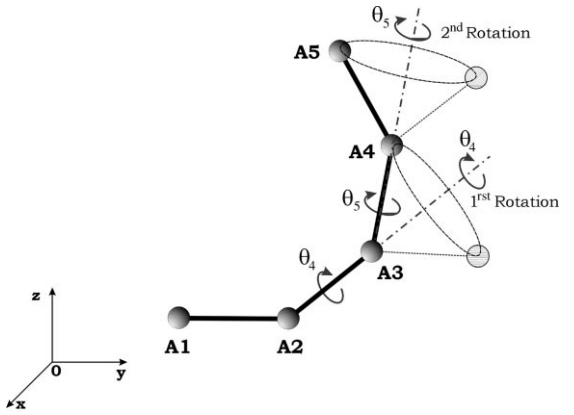

Fig. 9. Schematic spatial view of the torsional angles defining the first and subsequent rotations.

2.4. Generating the "seed bed"

2.4.1. Initial conditions

The first step in generating the atomic coordinates of the possible molecular configurations, which requires an "initial geometrical configuration" is as follows:

1. The bonded atoms are separated by standard interatomic distances.
2. The bond angles are considered characteristic bond angles for trigonal planar or tetrahedral atoms.

The initial configuration is constructed to be planar and is regarded to be the "starting geometrical configuration". We termed "initial geometrical configuration" because all the torsional angles are set to be zero degree. The last atom is not necessarily located at a distance of a typical bond length; usually there is superposition of atoms and thus it is not a chemically acceptable configuration.

This initial configuration represents only the starting point from which it is possible to begin the rotations.

2.4.2. Torsional angles

In order to determine the position of rotating atoms, we fix three atoms (A1, A2 and A3) in the initial coordinates and the first rotation is that of the next atom A4 (Fig. 9). Contrary to GASCOS searches for open chain compounds, atom A3 is not rotated with

respect to the direction A1–A2 in the cyclization search. However, when we complete all the cycles of rotations, we carry out the configuration search within the A1–A2–A3 plane, rotating all the bonds with respect to the initial plane. Thus, starting with the atom A4 we economize the calculation time and avoid repeated configurations.

Subsequently, the next atoms A5,A6,…,A($N-2$) are rotated in the second rotation, third rotation and so on. Only the two final atoms, including AN, are submitted to the analytic closure conditions. Thus, in a molecule of N atoms there are $N-5$ free rotations. The five rotations that are not free include three rotations from the first three atoms and two rotations of the two last atoms associated with the ring closure. Each rotation might be performed using different $\Delta\theta$ steps. However, imposing conditions to avoid the approach of one atom to another non-bonded atom in the environment, it is possible to find for θ, a range lower than 0–360° and therefore the value of the $\Delta\theta$ step will be lower too. Thus, it is possible to maintain the number of steps of each rotation fixed at a constant value. Since the range of restriction of torsional angles is variable, the $\Delta\theta$ steps of each rotation are changing as a consequence of the geometrical forms adopted by the molecule during the rotations. For example, in cyclononane we fix the atoms C1, C2 and C3 and rotate C4, C5, C6 and C7 (together with their respective hydrogen atoms). The analytical closure conditions are imposed to atoms C8 and C9.

If we choose for all rotations a step of $\Delta\theta = 20°$, each revolution is completed in 18 steps. Keeping constant the number of steps and satisfying the condition $\theta_{init} < \theta < \theta_{fin}$ and if:

$$\theta_{fin} - \theta_{init} < 360°, \tag{16}$$

then the variation in θ is $\Delta\theta'$:

$$\Delta\theta' = (\theta_{fin} - \theta_{init})/18 < 20°. \tag{17}$$

Thus, the maximum variation in θ is in starting step ($\Delta\theta = 20°$), and also it is determining the corresponding number of steps.

2.4.3. Defining a geometrical configuration

Once all the atoms are rotated in a sequential order and the closing conditions are satisfied, we obtain a geometrical configuration. A geometrical configuration is characterized by the values of their dihedral angles. If we have N atoms in a cyclic molecule, $N-5$ dihedrals are obtained from the free rotation, two dihedrals by rotations determined by the final closure conditions and the other three are determined by the final position of the two last atoms.

The values for the dihedral angles are defined, according to the IUPAC convention, as follows:

1. Dihedral positive if $0 <$ dihedral $< 180°$.
2. Dihedral negative if $180 <$ dihedral $< 360°$. In this case, the angle is presented as (dihedral $- 360°$).

Thus, a geometrical configuration is characterized by both the absolute values of the dihedrals and their respective signs. In order words, each geometrical configuration has only one succession of signs and values which is characteristic.

In molecules where all the atoms are equivalent (for example in all alicyclic hydrcarbons such as cyclononane, where all the carbons are equivalent), it is possible to find geometrical configurations with the same succession of signs, but with the order permuted. In that case the absolute values of the torsional angles are similar, therefore we have practically the same geometrical configuration, with the only difference that one of these configurations is spatially rotated with respect to that of the reference. Thus, the spatial orderings are irrelevant and therefore the way to analyze the extent to which a molecule is "folded" is by performing a dihedral by dihedral evaluation.

For molecules having a heteroatom the comparison must be carried out taking into account this different feature.

It should be emphasized that all geometrical configurations possessing the same succession of signs for their dihedrals (even when the absolute values are different) are located in the same subspace on the hyper-space N-dimensional of the molecule with N torsional angles. The characters of geometrical configurations and their respective positions in the subspaces are conditions used to generate the "seeds" for the future calculations of energetic minimization.

2.4.4. Separation of configurations in subspaces

The N-dimensional hyper-space of a cyclic molecule possessing N torsional angles is divided to

subspaces defined by a characteristic succession of signs for the dihedrals of geometrical configurations.

The first $N-5$ rotations are rotated in a sequential order and, therefore, the corresponding dihedrals are not repeated in the complete cycle of all the rotations.

However, the five dihedrals that remained are to be re-determined by the closure conditions. It is possible to obtain the same succession of signs and approximate values of dihedral angles, but in permuted order with respect to the original order. In this case, at least two (in general more) geometrical configurations are included in the same subspace, but they are representing "points" separated by a well-determined "distance". These "points" have "coordinates" which are values of dihedral angles in the hyper-space represented by the torsional angles.

From a mathematical point of view, if a geometrical configuration is characterized by N dihedrals θ_k^0 ($k = 1,\ldots,N$) and another geometrical configuration included in the same subspace is defined by θ_k ($k = 1,\ldots,N$), it is possible to define a "separation distance" between these configurations using an extension of the Pythagorean theorem for the case of a hyper-space, as:

$$D_\theta = \sqrt{\sum_{k=1}^{N} (\theta_k - \theta_k^0)^2}, \tag{18}$$

where previously we establish by comparison the correspondence:

$$\text{Dihedral } \theta_k \rightarrow \text{Dihedral } \theta_k^0. \tag{19}$$

Eq. (18) is defining a hyper-sphere in a subspace possessing as a center the point determined by the N dihedrals θ_k^0 ($k = 1,\ldots,N$).

For molecules possessing equivalent atoms (like for example the nine carbons of cyclononane), it is not possible to adopt the above definition in different subspaces.

This is a direct consequence of the possible permutations of the otherwise equivalent dihedrals making it impossible to fix any of the dihedral angles as initial. In contrast to the above, in molecules possessing a distinctive atom (i.e. a heteroatom) it is possible to choose this atom like the initial atom and the application of Eq. (18) will lead to a unique first dihedral angle. Thus, it is possible to establish a "separation distance" D among geometrical configurations in different subspaces.

From Eq. (18), it is evident that $D_\theta \rightarrow 0$ only when $\theta_k \rightarrow \theta_k^0$ ($k = 1,\ldots,N$); this leads to a pair of superimposed configurations and they therefore represent a unique configuration. If only one of the dihedrals is different, this is enough to obtain a "non-zero separation distance", i.e. $D_\theta \neq 0$.

In summary:

1. Geometrical configurations possessing the same succession of signs of dihedrals can be represented by "points" in a subspace on the hyper-space of configurations.
2. These "points" are more or less separated from each other in terms of their "separation distance" D_θ.
3. If $D_\theta = 0°$, then the "points" have identical locations and therefore correspond to the same configuration.

Probably the most critical result of a systematic search is that usually a huge number of geometrical configurations (with probably very few true energy minima) are obtained.

It is reasonable to assume that "points" located close to each other could be converging to the same minimum in a geometrical optimization process.

Thus, it is possible to "agglutinate" near geometrical configurations in order to lower the number of seeds for the energy calculations. This process effectively amounts to cluster analysis.

2.4.5. Considering neighborhood configurations

If $\Delta\theta$ should be infinitesimally small, we would have the highest precision in the conformational search. When, a priori, a finite value of $\Delta\theta$ is used, we are relinquishing some of that precision in the search of geometrical configurations. The magnitude of dihedral rotation is a discrete value in terms of $\Delta\theta$ and, therefore, we have a difference with respect to the ideal situation which would be the continuous variation (i.e. $\lim \Delta\theta \rightarrow \delta\theta$). On the basis of the above considerations, we might be missing geometrical configurations if some dihedral angles fall into the

following range:

$$(j-1) \qquad \Delta\theta \leq \text{Dihedral} \leq j\Delta\theta \tag{20}$$

with $j = 1, \ldots, m$, where $m = 360°/\Delta\theta$.

Some of these missed geometrical configurations might be located within the same subspace.

Thus, the interdependence among consecutive dihedrals is strong enough in order to prevent the other geometrical configurations to be located in a different subspace. Note that small variations in the values of the first dihedrals (first and second rotation) involves significant variations in the values of final dihedrals which determine that some configurations might appear in different subspaces.

However, during the comparison of configurations using dihedral permutations, it is possible to obtain values for the dihedrals that belong to the interval determined by Eq. (20).

If $\Delta\theta$ is at its maximum value, i.e. 20°, that 20° will correspond to the maximum inaccuracy, we define a hyper-sphere inside each subspace, which is determined by a succession of signs, possessing the following "radius":

$$D_{\Delta\theta} = \sqrt{\sum_{k=1}^{N} (\theta_k - \theta_k^0)^2} = \sqrt{N}\Delta\theta. \tag{21}$$

This equation allowed us to define a "cluster" of geometries or "geometrical configuration neighborhood". We define a "neighborhood of geometrical configurations" as a single geometrical configuration included in the same subspace and located at a "distance" D_θ with respect to the configuration of reference which satisfy (Fig. 10):

$$D_\theta \leq D_{\Delta\theta}. \tag{22}$$

In other words if we centralize a hyper-sphere of "radius" $D_{\Delta\theta}$ at the point of the subspace defining the reference configuration, all the representative "points" of the configuration located inside of the hyper-sphere will represent a "neighborhood of geometrical configurations".

This "agglutination" of neighborhood geometrical configurations" permit us to define a "seed" which is representative of the reference configuration. It should be emphasized that the creation of a "neighborhood of geometrical configurations" is the result of lack of precision caused by a non-infinitesimal $\Delta\theta$.

2.4.6. Defining a "family" of geometrical configuration

Once all the cycles of the atomic rotations are completed, we obtain a set of n_{initial} number of geometrical configurations. However, there are some that are not acceptable due to atomic superposition, especially of the non-bonding type H–H and C–H distances. At this time, we perform a check of the n_{initial} configurations and we discard those that satisfy:

1. non-bonded H–H distances lower than 1.01 Å;
2. non-bonded C–H distances lower than 1.01 Å.

This selection gives a number n_{final} ($<n_{\text{initial}}$) of geometrical configurations which are available to be used like "seeds" for the energetic calculations, because they do not have atomic superposition which can abort the calculation.

Nevertheless, the number of configurations n_{final} is still large enough to realize the minimization. We need to reduce the n_{final}, without the loss representativity of the geometrical configurations initially obtained.

A "family" is formed by a number of "neighborhoods of geometrical configurations" and only one will represent the "family".

In order to form the "families", we consider the following steps:

1. Choose in a sequential order, one geometrical configuration from the n_{final} structures.
2. Locate this configuration in a subspace from the succession of signs of dihedral angles.
3. Compare with the rest of configurations and calculate D_θ (Eq. (18)). If it satisfies:

$$D_\theta \leq D_{\Delta\theta} = \sqrt{N}\Delta\theta, \tag{23}$$

   then the configuration belongs to the "family".
4. All the geometrical configurations belonging to a "family" are not considered as new generators of new "families".
5. A geometrical configuration belonging to a "family" might also belong to another one. We cannot exclude a configuration belonging to a "family" like a member of another "family", because we do not know "a priori" which of the

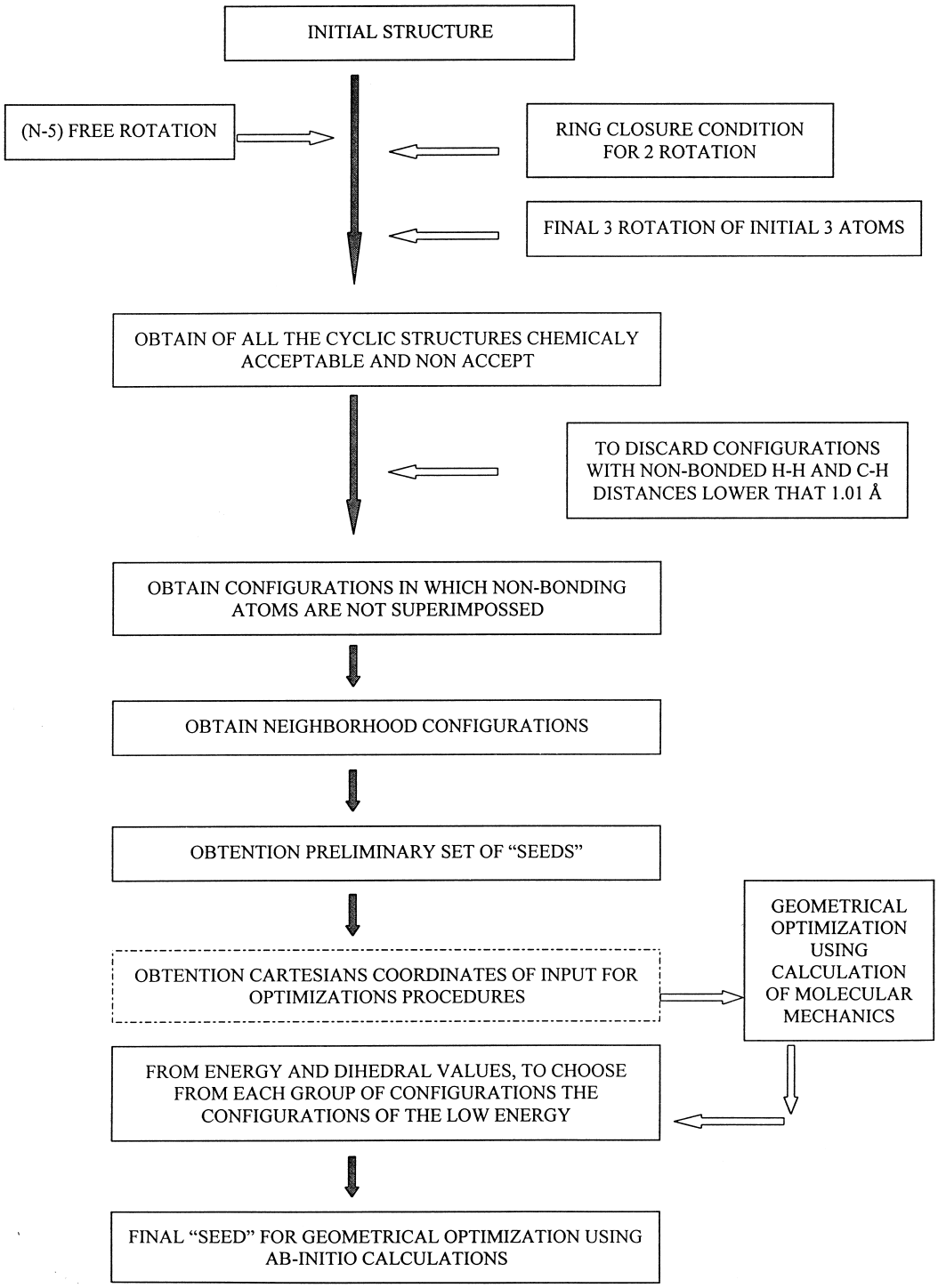

Fig. 10. Flow chart of GASCOS for cyclic compounds.

n_{final} configurations are the best generators of "families".

6. Because each geometrical configuration belonging to one "family" is not a generator of a new "family", the above process permits us to obtain a number of "families" which are lower than the initial number n_{initial}.
7. Only one geometrical configuration is representative of a "family". We call this representative configuration the "seed".

2.4.7. How to select the representative geometrical configuration for a "family"

Since all the members of a "family" belong to the same subspace and since they are also neighboring configurations, it is possible to establish a one-to-one relationship for the dihedrals of the members and the dihedrals of the configuration generator of the "family".

We can take the average values of all the dihedrals over all the members of the "family" and adopt this value as the representative one of "seed".

If a "family" has n_{family} members, the average value of each dihedral θ_k $(k = 1, \ldots, N)$ is:

$$\bar{\theta}_i = \frac{\sum_{j=1}^{n_{\text{family}}} \theta_{ij}}{n_{\text{family}}}, \qquad i = 1, \ldots, N, \;\; j = 1, \ldots, n_{\text{family}}. \tag{24}$$

For a molecule with N atoms, the set of average dihedrals:

$$(\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_N) \tag{25}$$

represent an average geometrical configuration comprising all the geometrical configurations located close to the generator configuration. These are the "seeds" used later, for geometrical optimizations via energy minimization.

2.4.8. Generating the cartesian coordinates which represents "seeds"

Once the set of average dihedrals representing the "families" are obtained, the next step is to determine the cartesian coordinates governing the spatial ordering of atoms in the average geometrical configuration. Thus, the atoms of the molecular skeleton must be located in order to reproduce the sequence:

$$\text{Family } j \qquad j = 1, \ldots, n_{\text{family}} \qquad (\bar{\theta}_{1j}, \bar{\theta}_{2j}, \ldots, \bar{\theta}_{Nj}). \tag{26}$$

Due to fact that there are conditions for the bond angles β (imposed by the closure conditions) and also that the comparison among dihedrals of geometrical configurations is performed by permuting their order during the formation of "families", we are also taking an average over the bond angles β.

Analogous to Eq. (24), if we have n_{family} members in the "family", the average value of angles β are defined by:

$$\bar{\beta}_i = \frac{\sum_{j=1}^{n_{\text{family}}} \beta_{ij}}{n_{\text{family}}}, \qquad i = 1, \ldots, N; \;\; j = 1, \ldots, n_{\text{family}}. \tag{27}$$

With this set of average angles β_k $(k = 1, \ldots, N)$:

$$(\bar{\beta}_1, \bar{\beta}_2, \ldots, \bar{\beta}_N) \tag{28}$$

we construct a planar initial geometrical form by using typical bonds lengths.

This initial geometrical form determines the initial coordinates for the atoms of the molecule, with the atoms (for example H) joined to each atom of the skeleton.

This initial geometrical form is submitted to the torsions fixed by the average dihedrals obtained above in the process of the "family" construction.

3. Algorithm and computational details

In Fig. 10, we show a flow chart of GASCOS used for flexible cyclic compounds. The rectangle marked with a dashed line in Fig. 10 indicates that cartesian coordinates of the configurations which represent the "seeds" must be obtained. The details of this process are given in Fig. 11.

The method adopted to generate the Cartesian coordinates is as follows:

1. Define a "starting planar geometrical configuration" using the set of average bond angles $(\bar{\beta}_1, \bar{\beta}_2, \ldots, \bar{\beta}_N)$ and typical bond lengths.

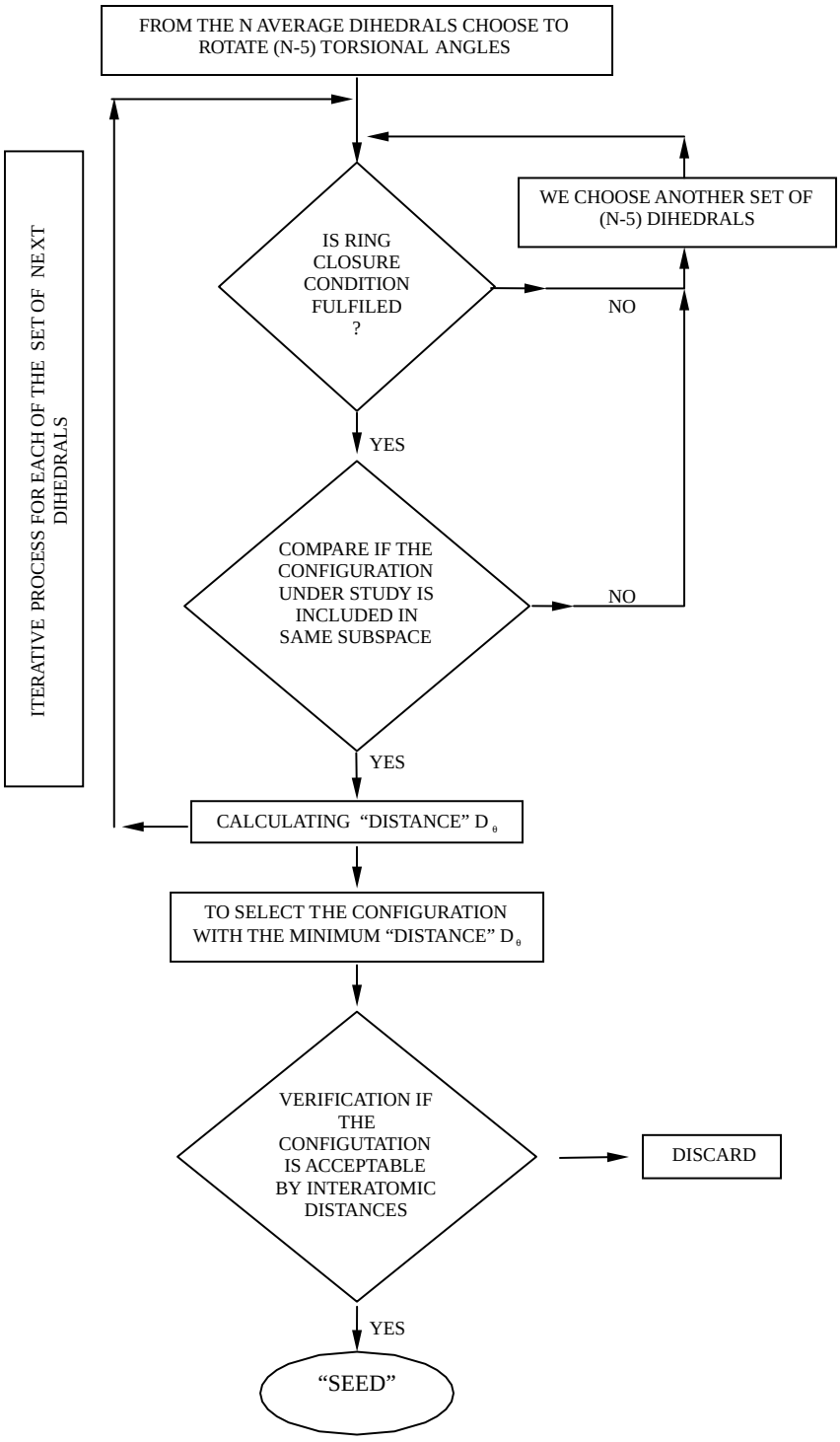

Fig. 11. Flow chart showing the generation of cartesian coordinates of the configurations which represent the "seeds".

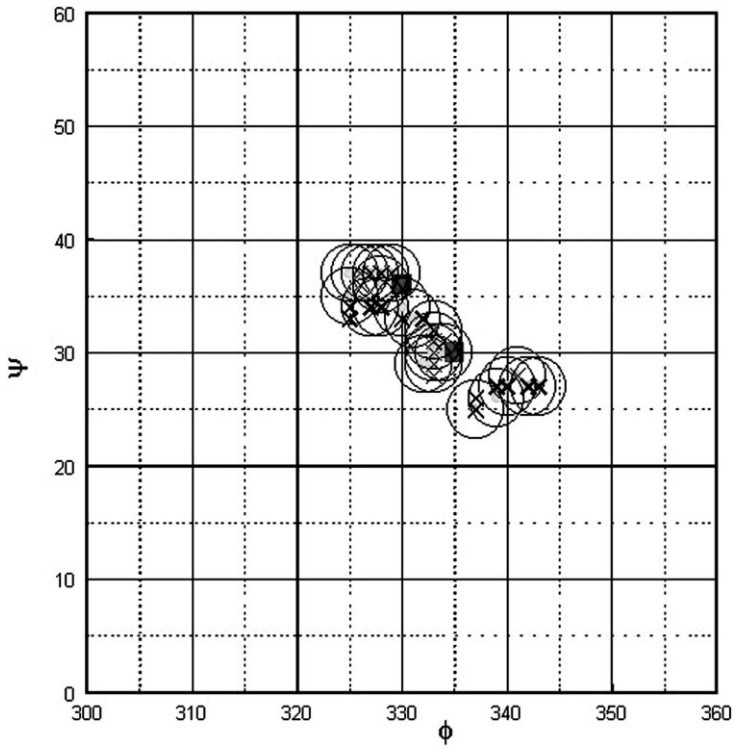

Fig. 12. Schematic 2D-representation for the GASCOS results of cyclopentane. The analytical solutions are marked by ■ and "seeds" by ×. The families are shown by spheres with radii = $(5)^{1/2}(1)$ (from Eq. (21)) and their centers are denoted by shaded circles.

2. Only $(N-5)$ atoms are submitted to the free rotations of dihedrals. Thus, from N dihedrals, we choose $(N-5)$.
3. The two last atoms are submitted to the closing conditions. Thus, we can close the ring.
4. The remaining three dihedral angles are determined by (1) and (2).

This process is submitted to the following proofs, which is a subroutine in the calculation programs.

4. Illustration of the method

Two small systems, cyclopentane and cyclohexane, were chosen to illustrate the GASCOS partially relaxed ring closure method in optimizing cyclic structures.

Cyclopentane is the simplest example because this molecule has three originally fixed atoms and two atoms are involved in the closing condition. Although the conformation of cyclopentane is determined by the values of the five dihedrals angles, it is possible to represent the GASCOS process in a schematic way in a 2D-space (φ, ψ), maintaining the rest of the dihedral angles in approximately constant values (Fig. 12).

The analysis of cyclopentane has been carried out in two steps:

1. First we use only the analytical ring closure conditions. We obtain two different conformations denoted by solid squares in Fig. 12.
2. In a second step we use the free rotation for the atoms varying the torsional angles each by 1°. We obtain 21 seeds denoted by crosses in Fig. 12.

The families are shown as circles. This figure is constructed in the (φ, ψ) space. The radii of these circles are the values of $D_{\Delta\theta}$ and the center is the average dihedrals $(\bar{\varphi}, \bar{\psi})$.

The geometry of the seeds is denoted by a cross.

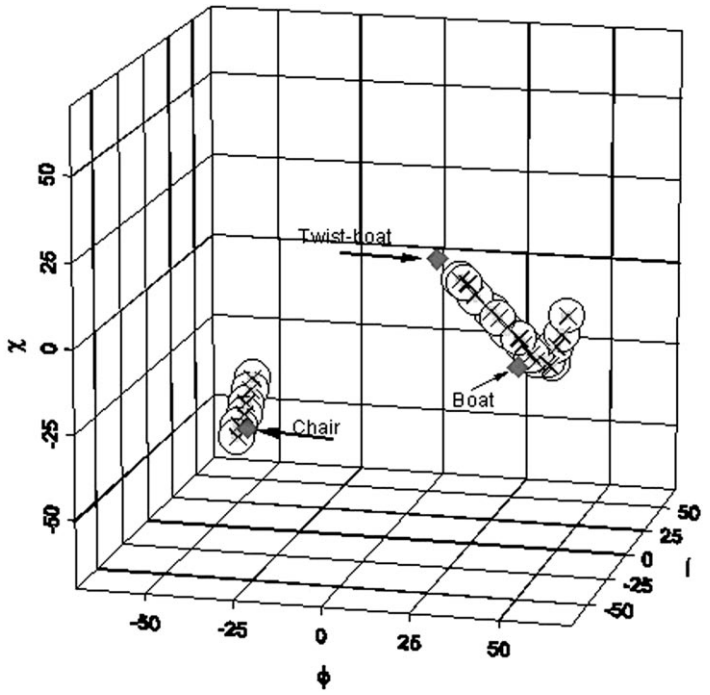

Fig. 13. Schematic 3D-representation for the GASCOS results of cyclohexane. MM2 conformations reported in Ref. [13] are shown by ■. The "seeds" are denoted by crosses (×) and the families are shown by spheres with radii = $(6)^{1/2}(2.5)$. The centers of the families are denoted by shaded spheres.

The conformations obtained by the analytic ring closure conditions are shown as solid squares.

Cyclohexane is a bit more involved because in addition to the three originally fixed atoms and the two atoms involved in the ring closure conditions, there is one atom which has a variable position in the GASCOS search. Thus, the process may be represented in a 3D-space (φ, ψ, χ) maintaining the rest of dihedral angles in approximately constant values.

The analysis of cyclohexane was carried out in an analogous fashion using the GASCOS procedure with the analytical ring closure condition. In this case, the angle of the free rotation was varied by 2.5 Å. Thus we obtain 22 seeds (Fig. 13).

The families are shown as spheres. This figure is constructed in the (φ, ψ, χ) space. The radii of these spheres are the values of $D_{\Delta\theta}$ and the center is the average dihedrals $(\bar{\varphi}, \bar{\psi}, \bar{\chi})$. The geometry of the seeds is marked by a cross. Previously published MM2 geometry [13] are shown as solid squares. The ab initio (HF/321G) optimized structure is located not very far from the MM2 geometries.

All of these clearly give validity to the GASCOS as an effective procedure for starting geometry to be used for energy minimization.

5. Conclusions

Algorithms, which find optimal solutions, are very important in computational chemistry and in particular in molecular modeling. Thus, the development and application of new methods using novel strategies continues to be an area of significant activity and paramount importance.

The results reported here clearly reveal that GASCOS is an effective procedure for the starting geometry to be used for energy minimization.

In general, systematic search methods are traditional, subject to the effects of combinatorial explosion. However, it is clear that the development of new search techniques like GASCOS may provide a method to directly and efficiently identify all solutions of interest.

Acknowledgements

The authors are grateful to Professors I.G. Csizmadia and S.J. Salpietro for helpful discussion. The research was supported by grants from Universidad Nacional de San Luis (UNSL). Dr R.D. Enriz is a career researcher of CONICET (Consejo Nacional de Ciencia y Tecnologia), Argentina.

References

[1] P.M. Dean, G. Jolles, C.G. Newton (Eds.), New Perspectives in Drug Design Academic Press, New York, 1995.
[2] M.A. Murcko, in: B. Lipkowitz, B. Boyd (Eds.), Recent Advances in Ligand Design Methods, Reviews in Computational Chemistry, vol. 11, Wiley, New York, 1997.
[3] D.E. Clarck, C.W. Murray, J. Li, in: K.B. Lipkowitz, B. Boyd (Eds.), Current Issues in De Novo Molecular Design, Reviews in Computational Chemistry, vol. 11, Wiley, New York, 1997, pp. 67–117.
[4] J.P. Doucet, J. Weber (Eds.), Computer-Aided Molecular Design: Theory and Applications Academic Press, New York, 1996.
[5] O. Farcas, A. Perczel, J.F. Marcoccia, M. Hollosi, I.G. Csizmadia, J. Mol. Struct. (Theochem) 331 (1996) 27.
[6] A. Perczel, O. Farkas, I.G. Csizmadia, J. Comput. Chem. 17 (1996) 821.
[7] A. Perczel, O. Farkas, I.G. Csizmadia, J. Am. Chem. Soc. 118 (1996) 7809.
[8] M.A. Zamora, H.A. Baldoni, A.M. Rodriguez, R.D. Enriz, C.P. Sosa, J.C. Vank, A. Perczel, A. Kucsman, O. Farkas, E. Deretey, I.G. Csizmadia, submitted for publication.
[9] S.J. Salpietro, A. Perczel, O. Farcas, R.D. Enriz, I.G. Csizmadia, J. Mol. Struct. (Theochem) 497 (2000) 39.
[10] C. Hausch, A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley-Interscience, New York, 1979.
[11] L.N. Santagata, F.D. Suvire, R.D. Enriz, J.L. Torday, I.G. Csizmadia, J. Mol. Struct. (Theochem) 465 (1999) 33.
[12] L.N. Santagata, F.D. Suvire, R.D. Enriz, Analytical ring closure condition for GASCOS (geometrical algorithm to search the conformational space), J. Mol. Struct. (Theochem) 507 (2000) 89.
[13] D.M. Ferguson, D.J. Raber, An approach to probing conformational space with molecular mechanics: random incremental pulse search, J. Am. Chem. Soc. 111 (1989) 4371.