ELSEVIER

Journal of Molecular Structure (Theochem) 507 (2000) 89–95

THEO
CHEM

www.elsevier.nl/locate/theochem

An analytic ring closure condition for geometrical algorithm to search the conformational space

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

Departamento de Química, Facultad de Química, Bioquímica y Framacia, Universidad Nacional de San Luis, Chacabuco 917, 5700 San Luis, Argentina

Received 15 August 1999; accepted 16 September 1999

Abstract

The recently reported geometrical algorithm to search the conformational space (GASCOS) scans conformational space exhaustively using an internal coordinate tree search. Using only geometrical operations and a set of criteria for eliminating chemically unreasonable atomic arrangements, the algorithm generates starting geometries for optimizations by molecular mechanics or by molecular orbital procedures. Up until now GASCOS has been used for linear structures, but an extension to cyclic structures is reported here. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Systematic conformational search; Analytic ring closure condition; Cyclic compounds; Conformational space search

1. Introduction

Three-dimensional geometry is an important property for understanding or predicting the behaviour of any molecular system. It is the necessary starting point for the derivation of structural features (surface, volume, shape, etc.), the estimation of steric requirements or the calculation of electronic properties. Geometry also plays a key role in the study of stereo-electronic complementarity which is essential for the ligand–receptor complex formation. Rigid molecules usually have one structure. Nearly rigid molecules, such as aspirin, have an enumerable number of conformers. Furthermore, it is relatively easy to construct input files for molecular computations for all such conformers. However, nobody can make, even an educated guess, of how many conformers might be associated with a large flexible cyclic molecules such as cyclosporin (cf. Scheme 1). Furthermore, the construction of input files for molecular computations for each and every plausible structure of cyclosporin is totally out of question. It is clear therefore that the choice of good starting geometries for conformationally flexible molecules is one of the greatest challenges, in applying quantitative molecular mechanics and molecular orbital calculations.

For systems with numerous degrees of freedom, the hypersurface of potential energy (in the space of the parameters defining the system) may have a substantial number of local minima, and to determine the global minimum is not always an easy task. Furthermore, finding that global minimum energy conformation alone may be not sufficient. In flexible molecules, several conformations may be significantly populated in given conditions and the observed physicochemical properties correspond to an average over this conformational mixture.

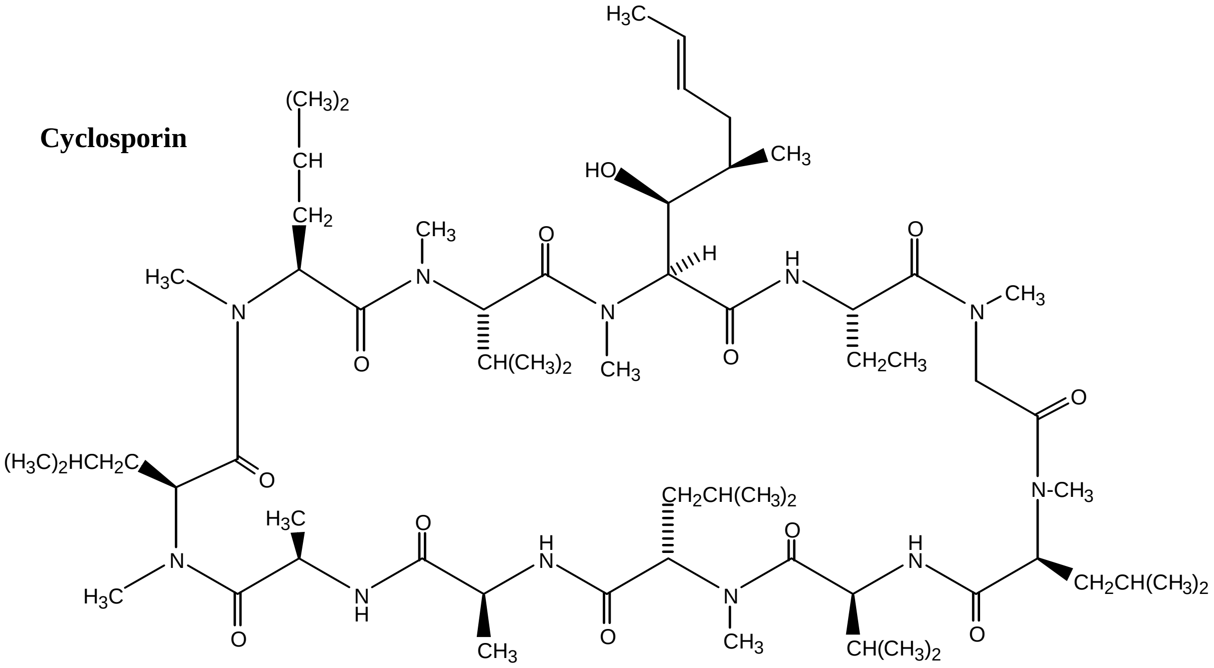
---

* Corresponding author.

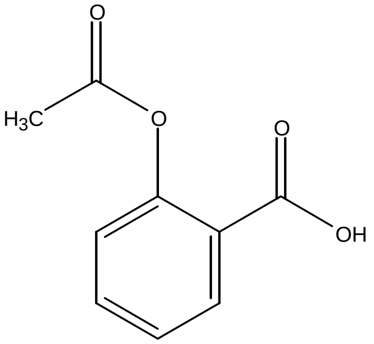
E-mail addresses: lnsantag@unsl.edu.ar (L.N. Santagata), fsuvire@unsl.edu.ar (F.D. Suvire), denriz@unsl.edu.ar (R.D. Enriz).

PII: S0166-1280(99)00353-X

Cyclosporin



Aspirin



Scheme 1.

Assignment of preferred conformations is not simple either because the number of possible conformers can be very large. This is illustrated for the case of conformation generation via a simple systematic search technique, and limited to single bonds incorporated into a straight chain. Given X single bonds and a step of $360°/Z$, the number of conformations, which need to be considered in such a search, is Z^X. Hence, we are dealing with a problem of considerable computational magnitude, even for relatively simple flexible acyclic molecules. The conformational problem becomes even more complex when flexible rings are involved.

There are several reasons for extending the search for new algorithm and to improve the available methods. The principal problem is that the various energy minimization processes do not go through potential energy barriers. They only move down-hill from the trial starting structure towards the nearest minimum, which may of course be only a local minimum. Even worse, after a search has been completed, there is no immediate indication of whether important conformers

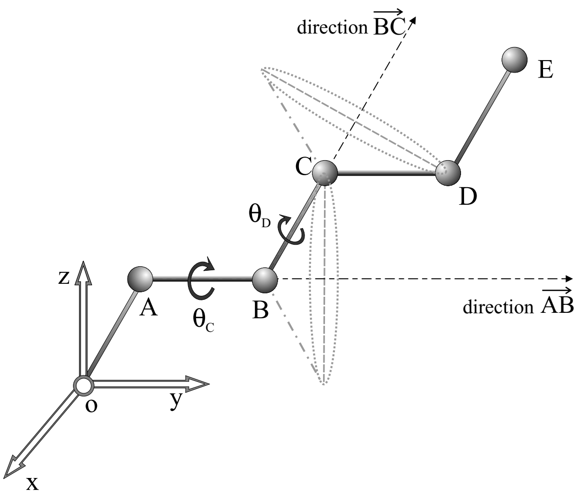


Fig. 1. Vector model for the initial phase of GASCOS search.

have been missed. It might be tempting to assume that the conformers missed, with some particular search method, are likely to be either high in energy or kinetically very unstable and thus less significant. However, it is more prudent to locate all the conformers one can possibly find on an energy surface first and then to examine their stability and their ease of conversion to other conformers to determine their relative importance. The most straightforward way to search the entire conformational space of a flexible molecule is by rotation around single bonds to construct geometries approximating all the anticipated low energy conformations (or transition states) of the system, followed by local minimization of these starting geometries.

Recently, we have reported a systematic search method, GASCOS [1], for open chain compounds. In the present report, we extend the capability of this algorithm to study conformationally flexible cyclic molecules. Thus, in this article we describe an internal coordinate conformer generator for cyclic compounds which has a number of advantages over methods described previously using tree search [2], random incremental pulse search (RIPS) [3] and stochastic techniques [4].

First, it is quite fast. No energy gradient needs to be evaluated and no distance matrix equations need to be solved with the result that the speed increases for conformer generation over the above methods. Second, operation of the method, in an internal coordinate framework, allows the sampling over all reasonable regions (including those of difficult accessibility) of the conformational space. Third, GASCOS has introduced a novel analytical ring closure constraint, which substantially improved the performance of these search techniques. Finally, the method is readily applied as a tree search, which makes feasible analyses of cyclic molecules containing many internal degrees of freedom (e.g. several dihedral angles).

2. Method of calculations

The approach we employ to generate molecular conformers computationally is analogous to the way a chemist might go about the same task using hand-held molecular models such as Dreiding or CPK molecular representations. However, most, if not all, chemists had bad experiences using these models because the generation of the actual conformers is limited by the knowledge, imagination, patience and manual dexterity of the researcher. Cyclic compounds produce an additional complication in the sense that the ring closure cannot always be maintained during the modelling process.

In contrast, GASCOS can provide a rapid method for generating molecular geometries approximating the various conformers of small-, medium- and large-sized acyclic and cyclic molecules.

2.1. Conformational space scan

GASCOS approach begins with a standard CH_2–CH_2–CH_2 geometry where the three carbon atoms are denoted by O–A–B in Fig. 1. Rotation about the A–B bond is possible only after carbon atom C is included. The torsional angle about the A–B bond is labelled as θ_C because this rotation moves atom C about a circle. The next rotational angle (θ_D) moves atom D about a full circle as illustrated in Fig. 1. This process involves therefore stepwise torsional rotations about all rotating bonds, retaining only those conformations which were passing geometrical tests designed to reject chemically unreasonable structures. Needless to say that for the accepted carbon skeleton conformations, the H-atoms are attached with the appropriate orientation.

GASCOS is a systematic search method. The

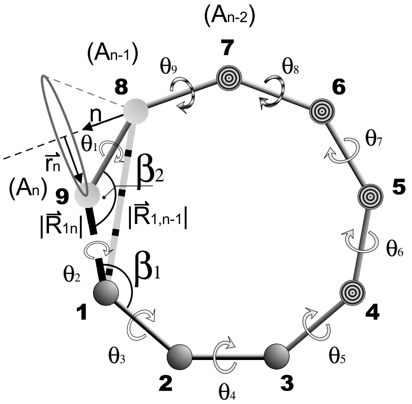


Fig. 2. Ring closure procedure for cyclononane. Note that β_1 and β_2 are part of the ring-closure condition. The rotational cone is associated with the determination of the last torsional angle: θ_9.

disadvantage of the existing search methods is that without some way to select only those conformers that are unique and structurally reasonable local minima on the conformational hypersurface, the generation and energy evaluation of conformations may become intractable. GASCOS have filtering algorithms, which eliminates the unreasonable portions of the conformational space from consideration. These filtering criteria include: ring-closure (described in Section 2.3), steric interactions and rejection of degenerate structures.

2.2. Non-bonded interatomic distance constraint

By applying constraint tests (e.g. tests for non-bonded interatomic contacts) to the structures generated by the systematic scan, we can eliminate in a single operation all structures having some undesirable component.

The most general constraint checks are minimum allowable separations among all non-bonded atoms. It is a simple distance constraint test which eliminates conformations which would have high steric proximity due to severe non-bonded contacts. We apply a stringent 1.53 Å cut off between carbon–carbon separations and 1.05 Å cut off between hydrogen–hydrogen and carbon–hydrogen separations. The non-bonded distance test provides one of the most efficient ways to eliminate chemically unreasonable structures from those generated by the systematic scan because the tests can be applied early in the structure generation process, and thus allow highly effective structure tree pruning. A second distance test is employed with cyclic molecules; the ring-closure condition.

2.3. Ring-closure constraints

Whereas in other tree search methods [2] rings are temporarily opened to form a pseudo-acyclic molecule which is then processed as the acyclic case but with additional constraints enforcing ring closure. GASCOS introduces a novel analytical ring-closure condition which permits the complete conformer generation.

Let us consider a cyclic molecule containing n atoms $A_k (k = 1, \ldots, n)$. Fig. 2 depicts the situation for the $n = 9$ case. Since we wish to retain the cyclic nature of the final structure, several geometrical constraints must be introduced. The most important of these are:

1. The closure distance constraint $|\vec{R}_{1n}|$ which is simply an acceptable distance between the two atoms (A_1 and A_n) forming the ring-closure bond (in the case of a C–C bond formation $R_{1n} = 1.53$ Å).
2. Two bond angle constraints:
    (a) bond angle β_1 which is determined by A_n, A_1 and A_2
    (b) bond angle β_2 determined by A_{n-1}, A_n and A_1.

Starting from Eq. [1] which was reported earlier [1], it is possible to obtain conditions to close the ring

$$|\vec{R}_{1n}|^2 = |\vec{R}_{1n}^{(\text{init.})}|^2 + 2\vec{r}_n^{(\text{init.})}\cdot(\vec{r}_n^{(\text{init.})} - \vec{R}_{1n}^{(\text{init.})})(1 - \cos\theta_n) + 2\vec{R}_{1n}^{(\text{init.})}\cdot(\hat{n} \times \vec{r}_n^{(\text{init.})}) \sin\theta_n \qquad (1)$$

We define:

(i) $|\vec{R}_{1n}|$ = distance between atoms A_1 and A_n after rotating by θ_n (note that in the case of a nine member ring $\theta_n = \theta_9$.

(ii) $|\vec{R}_{1n}^{(\text{init.})}|$ = initial distance between atoms A_1 and A_n before rotating by θ_n.

(iii) $\vec{r}_n^{(\text{init.})}$ = rotation vector associated with atom A_n giving the radius of the rotational cone.

(iv) n = the unit vector is determining the rotational axis which is associated with the motion of atom A_n along the bond direction of A_{n-1} and A_{n-2}.

We consider θ_k $(k = 1, \dots, n)$ as the torsion angles associated with the motion of the A_k atoms in the directions of their respective bonds (Fig. 2). All the conformations are obtained when we rotate in an arbitrary way about all the bonds $\theta_4, \theta_5 \dots \theta_{n-1}$. However, we are interested in only those in which the interatomic distance between atoms A_1 and A_n has the adequate value in order to complete the cyclization process.

Rotating all the torsional angles (from θ_4 to θ_{n-1}), it is possible to calculate the above values (i)–(iv). Once the $|\vec{R}_{1n}|$ value is fixed in Eq. [1], the unknown remaining is the torsional angle θ_n.

In order to solve Eq. [1] for $\theta_n$we define:

$$A = 2\vec{R}_{1n}^{(\text{init.})} \cdot (\hat{n} \times \vec{r}_n^{(\text{init.})}) \tag{2a}$$

$$B = 2\vec{r}_n^{(\text{init.})} \cdot (\vec{r}_n^{(\text{init.})} - \vec{R}_{1n}^{(\text{init.})}) \tag{2b}$$

$$C = |\vec{R}_{1n}|^2 - |\vec{R}_{1n}^{(\text{init.})}|^2 - 2\vec{r}_n^{(\text{init.})} \cdot (\vec{r}_n^{(\text{init.})} - \vec{R}_{1n}^{(\text{init.})}) \tag{2c}$$

then we can write Eq. (1) in the following form:

$$C = -B \cos \theta_n + A \sin \theta_n \tag{3}$$

Expressing $-\cos \theta_n$ we obtain:

$$-\cos \theta_n = (C/B) - (A/B) \sin \theta_n \tag{4}$$

Squaring Eq. (4) and using the $\cos^2 \theta_n = 1 - \sin^2 \theta_n$ identity, we obtain

$$1 - \sin^2 \theta_n = (C/B)^2 + (A/B)^2 \sin^2 \theta_n - 2(AC/B^2) \sin \theta_n \tag{5}$$

Eq. (5) is second order in the unknown $\sin \theta_n$. It is possible to express Eq. (5) in the following form by setting $X = \sin \theta_n$

$$[1 + (A^2/B^2)]X^2 - (2AC/B^2)X + (C/B)^2 - 1 = 0 \tag{6}$$

Solving Eq. (6) we obtain two roots (X_1 and X_2) which determine the two possible values for the torsional angle θ_n.

Two conditions are necessary for the roots in order to obtain ring-closure solutions:

$$X_1 \text{ and } X_2 \text{ must be real numbers} \tag{7a}$$

$$|X_1| \le 1 \text{ and } |X_2| \le 1 \tag{7b}$$

If both of the above conditions are satisfied then definitely there are two solutions for closure at the fixed $|\vec{R}_{1n}|$ distance resulting in two structures. Of the above two conditions, the first one (i.e. Eq. (1)) is stronger than the second one (i.e. Eq. (2)). If only condition (1) is satisfied, and condition (2) is not, then we have three possibilities. Either we have two roots and therefore two structures or one root and therefore one structure or no root at all which means no structure at all. In addition to the above, namely to satisfy condition (1), discussed earlier, we must also fulfil the ring closure conditions set by bond angles β_1 and β_2 as condition (1).

With respect to β_1 it is possible to fix a priori a tolerance range for the final value of this angle. In the present study, we may choose $\beta_1 = 109.5 \pm 5°$ In the case of β_2 this angle could be fixed at 109.5° Applying the cosine theorem, it is possible to determine the distance between A_1 and A_{n-1} (i.e. $|\vec{R}_{1,n-1}|$):

$$|\vec{R}_{1,n-1}|^2 = |\vec{R}_{1,n}|^2 + |\vec{R}_{n-1,n}|^2 - 2|\vec{R}_{1,n}||\vec{R}_{n-1,n}| \cos \beta_2 \tag{8}$$

In the above equation $|\vec{R}_{n-1,n}|$, the interatomic distance between A_{n-1} and A_n, is taken to be 1.53 Å (Fig. 1).

Taking into account that $|\vec{R}_{1,n-1}|$ is fixed, because β_2 is fixed, it is possible to apply again Eq. (5) and its solutions in order to solve the torsional angle θ_{n-1} associated with the motion of atom A_{n-1}.

In this case, $X = \sin \theta_{n-1}$ and:

$$A = 2\vec{R}_{1n-1}^{(\text{init.})} \cdot (\vec{n}' \times \vec{r}_{n-1}^{(\text{init.})}) \tag{9a}$$

$$B = 2\vec{r}_{n-1}^{(\text{init.})} \cdot (\vec{r}_{n-1}^{(\text{init.})} - \vec{R}_{1n-1}^{(\text{init.})}) \tag{9b}$$

$$C = |\vec{R}_{1n-1}|^2 - |\vec{R}_{1n-1}^{(\text{init.})}|^2 - 2\vec{r}_{n-1}^{(\text{init.})} \cdot (\vec{r}_{n-1}^{(\text{init.})} - \vec{R}_{1n-1}^{(\text{init.})}) \tag{9c}$$

where we define:

(i) $|\vec{R}_{1n-1}|$ = distance between atoms A_1 and A_{n-1} after rotating byθ_{n-1} (note: in the present example $\theta_{n-1} = \theta_8$).

(ii) $|\vec{R}_{1n-1}^{(\text{init.})}|$ = initial distance between atoms A_1 and A_{n-1} before rotating by θ_{n-1}.

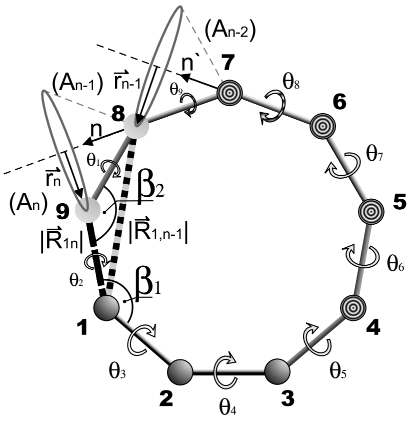


Fig. 3. Ring-closure procedure for cyclononane. The two rotational cones are associated with the determination of the last two torsional angles: θ_8 and θ_9.

(iii) $\vec{r}^{(\text{init.})}_{n-1}$ = rotation vector associated with atom A_{n-1} giving the radius of the rotational cone.

(iv) n' = the unit vector is determining the rotational axis which is associated with the motion of atom A_{n-1} along the bond direction of A_{n-2} and A_{n-3}. Analogously to Eq. (1) we now have Eq. (10)

$$|\vec{R}_{1n-1}|^2 = |\vec{R}^{(\text{init.})}_{1n-1}|^2 + 2\vec{r}^{(\text{init.})}_{n-1}\cdot(\vec{r}^{(\text{init.})}_{n-1} - \vec{R}^{(\text{init.})}_{1n-1}) \times (1 - \cos\theta_{n-1}) + 2\vec{R}^{(\text{init.})}_{1n-1}\cdot(n' \times \vec{r}^{(\text{init.})}_{n-1}) \sin\theta_{n-1} \tag{10}$$

Thus, it is possible to apply the analytical ring-closure condition reported here to the last two torsional angles in the molecule (θ_{n-1} and θ_n) as illustrated in Fig. 3.

2.4. Calculations

We wish to give only a preliminary outline of the calculations to be performed, cyclononane is used in the graphical presentation as it is a more general structure than cycloheptane or even cyclopentane, if we wish to concentrate on rings with odd number of carbon atoms.

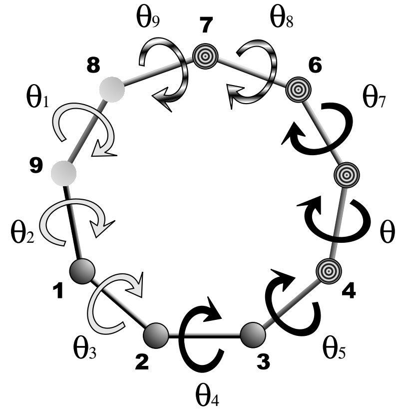


Fig. 4. The set of the ring closure. Torsional angle θ_k is associated with the rotating motion of atom k. First three atoms (1, 2 and 3) come as an initial standard geometry. With the addition of an atom a new torsional angle is created. A total of $(n-5)$ such angles can be determined by the GASCOS search (in the present case θ_4, θ_5, θ_6 and θ_7). Two torsional angles (θ_8 and θ_9) are determined by the ring-closure condition by varying β_1. The final three torsional angles (θ_1, θ_2 and θ_3) are the consequences of the ring closure.

The first three torsional angles ($\theta_1, \theta_2, \theta_3$) may be chosen arbitrarily. These three torsional angles are to be fixed. GASCOS then can search torsional angles from θ_4 to θ_7 (Fig. 4) as they may be rotated each by a preselected increment (e.g. 15, 20, 25 or 30°). Torsional angles θ_8 and θ_9 will then be analytically obtained as result of the ring-closure condition. The new values of torsional angles θ_1, θ_2 and θ_3 will become the consequence of the ring closure. The carbon–carbon and carbon–hydrogen interatomic distances can be considered to be 1.53 and 1.05 Å, respectively.

In order to assess the stability of the compounds which passed all criteria molecular mechanics (MM2) or quantum mechanics calculations, at semi-empirical (e.g. AM1 [5,6]), or ab initio (e.g. RHF/3-21G), levels of theory can be performed using SPARTAN [7] or the Gaussian [8] program system.

3. Conclusions

It appears that surprisingly little effort has been made toward developing optimal algorithms to explore the conformational hyperspace using systematic searches. The recently developed GASCOS method which was previously designed to treat open chain system has been extended to treat cycloalkanes with the help of the analytic ring-closure constraint.

Acknowledgements

The research was supported by grants from the Universidad Nacional de San Luis (UNSL). Dr R.D. Enriz is a carrier researcher of CONICET (Consejo Nacional de Ciencia y Tecnología) Argentina.

References

[1] L.N. Santagata, F.D. Suvire, R.D. Enriz, L.L. Torday, I.G. Csizmadia, J. Mol. Struct. (Theochem) 465 (1999) 33–67.

[2] M. Lipton, W.C. Still, J. Comput. Chem. 9 (1988) 343–355.

[3] D.M. Ferguson, W.A. Glauser, D.J. Raber, J. Comput. Chem. 10 (1989) 903–910.

[4] M. Saunders, J. Comput. Chem. 12 (1991) 645–663.

[5] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.

[6] J.J.P. Stewart, QCPE 51 (1985) 523.

[7] SPARTAN version 4.0. Wavefunction, Inc. Irvine, CA, USA, 1995.

[8] M.J. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.A. Robb, J.R. Cheeseman, T. Keith, G.A. Petersson, J.A. Montgomery, K. Raghavachari, M.A. Al-Laham, V.G. Zakrzewski, J.V. Ortiz, J.B. Foresman, J. Cioslowski, B.B. Stefanov, A. Nanayakkara, M. Challacombe, C.Y. Peng, P.Y. Ayala, W. Chen, M.W. Wong, J.L. Andres, E.S. Replogle, R. Gomperts, R.L. Martin, D.J. Fox, J. S. Binkley, D.J. Defrees, J. Baker, J.P. Estewart, M. Head-Gordon, C. Gonzalez, J.A. Pople, GAUSSIAN 94, Revision E.2, Gaussian, Inc., Pittsburgh, PA, 1995.