Chain closure: A problem in molecular CAD *

M. D. Di Benedetto† P. Lucibello† A. L. Sangiovanni-Vincentelli‡ K. Yamaguchi‡

†Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Roma, Italy
‡Department of EECS, University of California, Berkeley, CA 94720

Abstract

Conformational analysis is the problem of finding all minimal energy three-dimensional configurations of molecules. Cyclic structures are of particular interest. An efficient algorithm based on a purely geometric approach that generates feasible configurations very efficiently is presented thus making full conformational analysis possible even for fairly large cyclic structures.

1 Introduction

The design of a new drug is a complex and extremely expensive task. The design process has been based on a trial-and-error procedure: given certain properties of the drug to be developed, several thousands of compounds are generated and tested for activity and for side effects. This results in a development process that may take as long as ten years.

Recently, new ways of approaching the problem (rational drug design) have been proposed that make use of a fairly rigorous top-down methodology somewhat similar to the methodology used for VLSI design. As in the case of the early days of VLSI design, most of the design steps are now carried out mostly in a heuristic way and with little help from the computer. In our opinion, an extensive use of computer aids will soon take place. There are many interesting computational problems to be solved. The algorithms and methodology used in other design fields such as electronic system design could be of great help. In this paper we address an interesting geometrical problem arising from the desire to determine a molecule such that its three-dimensional configuration matches the structure of a given receptor molecule. The solution to this problem is, for example, of key importance in synthesizing anti-rejection drugs. Note that the three-dimensional configuration of a molecule depends on the force field (molecular energy model) in which the molecule is immersed. In fact, three-dimensional configurations that minimize energy are the actual configurations of interest.

Three-dimensional configurations for molecules have to satisfy a set of constraints, for example the distance between two atoms (*bond length*) and the angle between lines connecting two consecutive atoms (*bond angle*) cannot vary by more than a given small amount. The configurations that satisfy all the constraints on the positions of the atoms of the molecule are called *conformations*. *Conformers* are minimal energy conformations. If the molecules under study have cyclic structure, then the positions of the atoms are further constrained, and while the additional constraints allow the exploration of larger systems than in the acyclic case, the imposition of the constraints is complicated. Even finding configurations that satisfy all constraints, let alone the minimal energy ones, is difficult. There is a definite interest in an efficient solution to the problem of determining conformers for cyclic molecules since many biologically important molecules, e.g. peptides, have ring structures.

All conformers are of interest in studying the physical and chemical properties of a given molecule. Hence, standard optimization algorithms that find *a* local minimum are not satisfactory. To obtain all conformers, the conformation space has to be searched thoroughly. Most of the algorithms proposed in the literature follow this scheme: first a set of molecular conformations is generated that populate uniformly the search space. Then these structures are used as starting points for an energy minimization routine [10]. The various algorithms differ according to the procedure used to generate the initial conformations and the energy minimization process. Three approaches have been followed thus far with some success: one is based on stochastic optimization algorithms such as simulated annealing and its predecessor Monte Carlo analysis (e.g., [8, 12, 13, 14]); the second is based on descent algorithms of several types (e.g., [1, 5, 7, 11]); the third is based on molecular dynamics, i.e., the initial conformations are used as initial conditions for differential equations that describe the motions of the molecules under a given force field [10]. Conformers are obtained as equilibrium points. Interesting variations are presented in [9] and in [4] where some of the conformers are generated from other conformers directly without performing a space search. While elegant and effective, these approaches do not guarantee that all conformers of interest are found. In all cases the computational complexity increases greatly with the number of atoms in the molecule, and the quality of the search, i.e., the percentage of conformers that are identified and the running time, depends on the initial conformations

*This work is supported in part by a National Science Foundation Graduate Research Fellowship.

chosen. A sufficient number of initial conditions has to be generated to guarantee that the search space is visited thoroughly. According to [10], "surprisingly little effort has been given toward determining optimal dihedral angle increments" in the systematic space searches.

For the case of cyclic structures, any time a conformation is needed either as initial conditions or in the inner loops of the optimization algorithms, most of the proposed algorithms first randomly generate three-dimensional configurations and then discard the ones that do not satisfy the chain closure condition. This approach is clearly wasteful, and a procedure that can generate conformations directly is of interest.

This paper deals with efficient generation of conformations of cyclic molecules. Being able to generate conformations efficiently allows a brute-force approach to energy minimization—at least in the case of molecules with fairly small number of degrees of freedom—by substituting the use of optimization algorithms with exhaustive enumeration. Even when an exhaustive search with fine grain resolution is not computationally feasible, efficient generation of conformations allows the population of the search space with a larger number of initial conditions, thereby increasing the probability of finding all conformers and speeding up the calculation of feasible solutions in the inner loop of optimization algorithms.

The paper by Gō and Scheraga [7] approaches the problem of finding conformations of cyclic structures by writing a set of six equations that determines the position of the atoms in a chain given the position of the first atom and imposes the additional constraint that the final atom of the chain must coincide with the first. In this paper a pure geometrical approach is proposed that reduces the problem of finding conformations to one of finding the solution to a quadratic equation in *one* variable thus yielding an algorithm that is largely insensitive to the number of atoms in the molecule.

The paper is organized as follows: in Section 2 the problem is defined and the basic terminology introduced. In Section 3 the geometric procedure is derived for chains with six atoms. It is shown that the problem can be reduced to the problem of solving one quadratic equation in one unknown. The properties of the procedure, its correctness and its complexity are also investigated. Then the procedure is extended to the case of chains with greater than six atoms by decomposing the problem into simpler subproblems that reduce the calculations to the solution of a quadratic equation as in the six atom case. In Section 4 the implementation of the procedure is described, and numerical results are presented that demonstrate the capability of the procedure to find all conformations of some fairly large molecular chains. In Section 5, concluding remarks are offered.

2 Problem Setting

In this section the molecular chain is described in a geometric setting. Then, on the basis of this mathematical description, the chain closure problem is formulated.

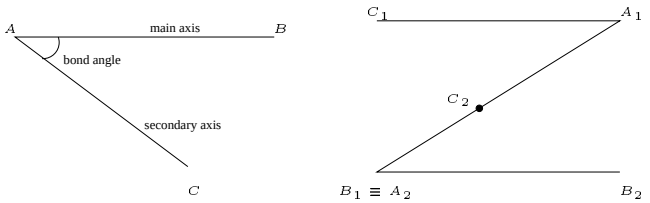

Figure 1: A single "bond" and two connected bonds

Let a *molecular chain* be a set of ordered bonds. A *bond* is a rigid body where two oriented axes are defined and referred to as the *main axis* and the *secondary axis*. Let A be the point of incidence of the two axes, and let B and C be two given points on the main and secondary axes respectively, in the positive directions. The distance between A and B is the *bond length* and the angle $\widehat{BAC}$, i.e. the angle spanned by the rotation of AC on AB according to the right-hand rule, is the *bond angle*. Fig. 1 visualizes a bond.

Two consecutive bonds, denoted by (A_1, B_1, C_1) and (A_2, B_2, C_2), are said to be *connected* when (i) the point B_1 of the antecedent bond coincides with the point A_2 of the subsequent bond and (ii) the main axis of the antecedent bond coincides with the secondary axis of the subsequent bond with opposite directions (see Fig. 1).

Two connected bonds can freely rotate around the common axis. This rotation, called *dihedral rotation*, can be arbitrary, and no distinction is made between rotations which differ modulo a full round angle. The set of all rotations, denoted by S, is diffeomorphic to a circle.

Given a chain with n bonds, numbered from 1 to n, the chain can be *closed*, or *chain closure* can be achieved, if there exist dihedral angles such that consecutive bonds are connected and bond n is connected with bond 1. Let $S^n = S \times S \times \ldots \times S$ be n-times the circle S. The subset C of S^n of all dihedral angles for which chain closure can be achieved is defined as the *conformation space* of the closed chain. In general, C is not a connected set [9].

The dihedral angles that close a chain must satisfy a set of six algebraic equations, defined by smooth functions [7]. Hence, there are $n-6$ degrees of freedom in a closed molecular chain. More precisely, let r be the rank of the Jacobian of the six algebraic equations at a given point P of the conformation space C. Two cases are possible:

1. $n > 6$. There exists a neighborhood O of P such that $C \cap O$ is a smooth manifold of dimension $n - r \geq n - 6$.
2. $n \leq 6$. If the Jacobian is full rank, P is an isolated point. Otherwise, there exists a neighborhood O of P such that $C \cap O$ is a smooth manifold of dimension $n - r$.

In the existing literature the current approach to the chain closure problem consists of searching for solutions to the previously mentioned six algebraic equations. The main drawback of this approach is that, because of the complexity of

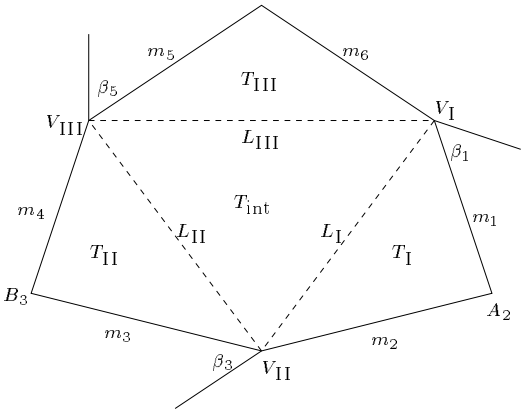

Figure 2: A chain with six bonds

these equations, numerical procedures are considerably time-consuming. Furthermore, since in general C is not a connected set [9], there is no guarantee that all the solutions are found. For rings of size $n = 8$ and larger, the number of configurations is so large that any analysis is usually incomplete [2]. In the following section the chain closure problem is tackled in a different way. First, a pure geometric algorithm is given which guarantees that all possible solutions are found. Then, the algebraic implementation of the algorithm is illustrated. This reduces to the search for the zeros of one algebraic equation in one unknown and is much simpler than existing algorithms.

3 Chain Closure

In this section a geometric procedure is proposed for solving the chain closure problem. First, in Section 3.1, the case of a chain with six or fewer bonds is solved. Then, in Section 3.2, the procedure for the six-or-fewer-bonds case is extended to a generic chain.

3.1 Chains with Six or Fewer Bonds

Consider a chain with six bonds, numbered from 1 to 6, whose main axes are denoted as $m_1, \ldots, m_6$, secondary axes as $s_1, \ldots, s_6$ and bond angles as $\beta_1, \ldots, \beta_6$.

Procedure 1 (Chain Closure)

1. (a) Connect pairs of consecutive bonds: $(1,2)$, $(3,4)$, and $(5,6)$. With reference to Fig. 2, let $L_{\rm I}$, $L_{\rm II}$ and $L_{\rm III}$ be the segments which connect point B of the first bond of each pair with point A of the second bond of the same pair. The lengths of $L_{\rm I}$, $L_{\rm II}$ and $L_{\rm III}$ are uniquely determined since they are not affected by dihedral rotations.

   (b) Construct, if possible, the triangle $T_{\rm int}$ whose edges are the segments $L_{\rm I}$, $L_{\rm II}$ and $L_{\rm III}$ and let $V_{\rm I}$, $V_{\rm II}$ and $V_{\rm III}$ denote its vertices.

2. Let $T_{\rm I}$, $T_{\rm II}$ and $T_{\rm III}$ be the triangles formed, respectively, by $(L_{\rm I}, m_1, m_2)$, $(L_{\rm II}, m_3, m_4)$, $(L_{\rm III}, m_5, m_6)$. Satisfaction of bond angle constraints at the vertices $V_{\rm I}$, $V_{\rm II}$ and $V_{\rm III}$ is searched for by means of rotation of the triangles $T_{\rm I}$, $T_{\rm II}$ and $T_{\rm III}$ around $L_{\rm I}$, $L_{\rm II}$ and $L_{\rm III}$ respectively, and by means of rotation of the bonds 1, 3 and 5 around the axes m_1, m_3 and m_5.

   (a) Let $\theta_{\rm I} \in S$ be a generic rotation of $T_{\rm I}$ around $L_{\rm I}$.

   (b) Find all rotations $\theta_{\rm II} \in S$ of $T_{\rm II}$ around $L_{\rm II}$ so that the angle $A_2\widehat{V_{\rm II}}B_3$ between m_2 and m_3 is equal to the bond angle β_3.

   (c) For each rotation $\theta_{\rm II}$, if any, determined at Step 2b, find all rotations $\theta_{\rm III} \in S$ of $T_{\rm III}$ around $L_{\rm III}$ so that the angle $A_4\widehat{V_{\rm III}}B_5$ between m_4 and m_5 is equal to the bond angle β_5.

   (d) For each pair $(\theta_{\rm II}, \theta_{\rm III})$, if any, if the angle $A_6\widehat{V_{\rm I}}B_1$ between m_6 and m_1 is not equal to the bond angle β_1, chain closure cannot be achieved for $(\theta_{\rm I}, \theta_{\rm II}, \theta_{\rm III})$. Otherwise, bonds 1, 3 and 5 are rotated around m_1, m_3 and m_5 to make the secondary axes s_1, s_3 and s_5 coincide with opposite directions with m_6, m_2 and m_4 respectively, thus achieving chain closure for $(\theta_{\rm I}, \theta_{\rm II}, \theta_{\rm III})$.

Theorem 1 *A point of S^n belongs to the conformation space if and only if it is generated by Procedure 1 (Chain Closure).*

Proof (By contradiction) Assume that a point of the conformation space is not generated by Procedure 1. This means that one of the steps has failed. In Step 1, Procedure 1 fails if the segments $L_{\rm I}$, $L_{\rm II}$ and $L_{\rm III}$ do not define a triangle. However, the existence of such a triangle is necessary for satisfying condition (i) of the bond connection between bonds $(2,3), (4,5)$ and $(6,1)$ (the bond angle constraints at the vertices $V_{\rm I}$, $V_{\rm II}$ and $V_{\rm III}$ are not necessarily satisfied). Hence, if this cannot be done, chain closure cannot be achieved, yielding a contradiction. In Step 2, Procedure 1 fails if one of the steps 2b, 2c or 2d fails. In this case there is no rotation that makes condition (ii) of bond connection satisfied at $V_{\rm I}$, $V_{\rm II}$ or $V_{\rm III}$, and this again would yield a contradiction. Finally, Procedure 1 does not generate spurious solutions since, given $(\theta_{\rm I}, \theta_{\rm II}, \theta_{\rm III})$ that satisfy the bond connection constraints, a closed chain can be constructed. ∎

An important question is under which conditions the chain can be closed, i.e. under which conditions Procedure 1 does not fail. A trivial necessary and sufficient condition for the completion of Step 1 is that $L_{\rm I}$, $L_{\rm II}$ and $L_{\rm III}$ satisfy the geometric conditions that allow building the triangle $T_{\rm int}$. A deeper necessary and sufficient condition under which Step 2 can be completed is given in the following theorem whose proof can be found in [3].

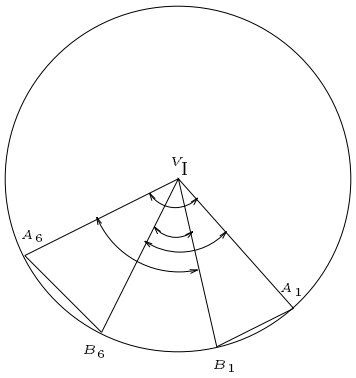

Figure 3: Cones formed by rotation of triangles

Theorem 2 *Consider the cone C_1 described by m_1 when the triangle T_{I} is rotated around L_{I}, and the cone C_6 described by m_6 when the triangle T_{III} is rotated around L_{III} (see Fig. 3).*

If the cones C_1 and C_6 do not intersect, a solution to Step 2b exists if and only if the angle β belongs to the interval defined by the smallest and largest angles of the set $\{A_1\widehat{V_{\mathrm{I}}}A_6, B_1\widehat{V_{\mathrm{I}}}B_6, A_1\widehat{V_{\mathrm{I}}}B_6, B_1\widehat{V_{\mathrm{I}}}A_6\}$. If the cones C_1 and C_6 intersect, a solution to Step 2b exists if and only if the angle β belongs to the interval defined by the null angle and $A_1\widehat{V_{\mathrm{I}}}A_6$.

Another important question answered in the following theorem (proof in [3]) is how many solutions exist.

Theorem 3 *The number of solutions n_{2b} of Step 2b is either 0, 1, 2 or infinitely many, i.e. the whole S (degenerate case).*

Remark If $(\theta_{\mathrm{I}}, \theta_{\mathrm{II}}) \in R^2$ is a pair of rotations which satisfies the bond constraint at V_{I}, then the pair $(-\theta_{\mathrm{I}}, -\theta_{\mathrm{II}}) \in R^2$ also satisfies the bond constraint.

The case of a chain with less than six bonds can be seen as a particular case of the one with six bonds. Indeed, it is sufficient to collapse one or more triangles T in segments.

3.2 Chains with More Than Six Bonds

A chain with more than six bonds can be handled by cutting it in arbitrary groups, numbered I, II and III, of two or more connected bonds. With reference to Fig. 4, the internal edge L_{I} connects point B of the first bond of the first group with point A of the last bond of the same group. L_{II} and L_{III} are defined in a similar way. In this case, as opposed to the cases analyzed in Section 3.1, the internal edge, the secondary axis of the first bond and the main axis of the last bond of a group are in general not co-planar. Each group of bonds is an open chain. The configuration of a group is defined by the dihedral rotation of the bonds belonging to the group except the first and last ones. Once a configuration has been chosen, the distance between point A of the first bond of the group and point B of the last bond of the group is not modified by the first or last dihedral rotations, as it is not in the case of

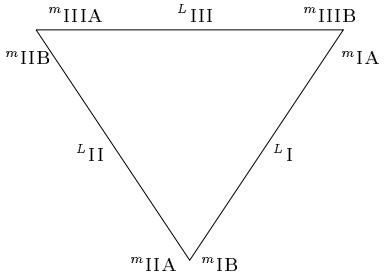

Figure 4: A generic reduced chain

chains containing six or fewer bonds. On the basis of these observations, Procedure 1 is modified as follows.

Procedure 2

1. (a) Cut the chain into groups I, II and III of two or more connected bonds.
   (b) For each group, select a configuration.
2. (a) Define L_{I} connecting point B of the first bond of group I with point A of the last bond of group I. Define L_{II} and L_{III} in a similar way.
   (b) Construct, if possible, the triangle of edges L_{I}, L_{II} and L_{III}.
3. Consider the triangle in Fig. 4, where m_{IA} denotes the main axis of the first bond and m_{IB} the main axis of the last bond of the group of bonds backing L_{I}, and so on.
   (a) Let $\theta_{\mathrm{I}} \in S$ be a generic rotation of group I around L_{I}.
   (b) Find all rotations $\theta_{\mathrm{II}} \in S$ of group II around L_{II} so that the angle between m_{IB} and m_{IIA} is equal to their bond angle.
   (c) For each rotation θ_{II}, if any, determined at Step 3b, find all rotations $\theta_{\mathrm{III}} \in S$ of group III around L_{III} so that the angle between m_{IIB} and m_{IIIA} is equal to their bond angle.
   (d) For each pair $(\theta_{\mathrm{II}}, \theta_{\mathrm{III}})$, if any, if the angle between m_{IIIB} and m_{IA} is not equal to their bond angle, chain closure cannot be achieved for $(\theta_{\mathrm{I}}, \theta_{\mathrm{II}}, \theta_{\mathrm{III}})$. Then, change the angle of rotation around L_{I}. Otherwise, find rotations around m_{IA}, m_{IIA} and m_{IIIA}, so that chain closure is achieved for $(\theta_{\mathrm{I}}, \theta_{\mathrm{II}}, \theta_{\mathrm{III}})$.

The results proven in the previous subsection apply *mutatis mutandis* to this new procedure. In particular, it is possible to prove that the above procedure generates all points of the conformation space.

4 Implementation and Results

In this section we present some implementation details and numerical results obtained from some molecular structures in order to underline the properties of the method.

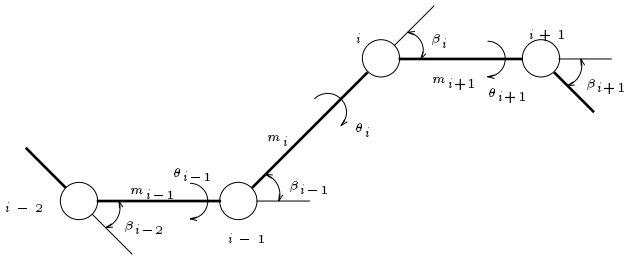

Figure 5: Definition of bond lengths m_i, bond angles β_i and dihedral angles θ_i

4.1 Coordinate System

The algorithm uses internal coordinates, i.e. bond lengths, bond angles and dihedral angles, for representing the geometric structure [10]. Using internal coordinates permits all regions of the conformationally accessible space to be sampled [11]. However, there are steps of the procedure that need external (Cartesian) coordinates for the atom positions, for instance when the length of the segment joining two atoms is calculated. The same transformations as in [7] are used; see Fig. 5 for notation. For atoms at position $\mathbf{r}_i$ and $\mathbf{r}_{i-1}$ in the ith and $(i-1)$th coordinate system, respectively, the relation is

$$\mathbf{r}_{i-1} = \mathbf{T}_{i-1}\mathbf{R}_i\mathbf{r}_i + \mathbf{p}_{i-1} \quad (1)$$

$$\mathbf{T}_{i-1} = \begin{pmatrix} \cos\beta_{i-1} & -\sin\beta_{i-1} & 0 \\ \sin\beta_{i-1} & \cos\beta_{i-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

$$\mathbf{R}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_i & -\sin\theta_i \\ 0 & \sin\theta_i & \cos\theta_i \end{pmatrix} \quad (3)$$

$$\mathbf{p}_{i-1} = \begin{pmatrix} m_{i-1} & 0 & 0 \end{pmatrix}^T. \quad (4)$$

4.2 Six-membered Ring

The crucial part of the implementation of the chain closure procedure is Step 2b through Step 2d of Procedure 1 (p. 3). See Fig. 6. Given $\theta_{\mathrm{III}}, \theta_1, \theta_n, \alpha, \beta$, the set of solutions for θ_{I} must be found. Using the coordinate systems introduced above, the solution angles θ_{I} are obtained from the solution of a quadratic equation of the form $ax^2 + 2bx + c = 0$, where $x = \cos\theta_{\mathrm{I}}$. This equation may have 0, 1, 2 or infinitely many solutions in the degenerate case, in agreement with Theorem 3. The derivation of this result follows.

Let the x-axis be defined by the segment L_{III} and the y-axis be in the plane of T_{int} such that the positive direction is towards T_{int}. Then the z-axis is $x \times y$. Let p_n be point A of bond m_n, and let p_1 be point B of bond m_1 (as defined in Fig. 1). A rotation $\theta_{\mathrm{I}} = 0$ puts triangle T_{I} coplanar and external to T_{int}. Positive rotation is defined by using the right-hand rule when traversing the closed chain. The Cartesian coordinates are then

$$p_n = \begin{pmatrix} \cos\theta_n - \sin\theta_n\cos\theta_{\mathrm{III}} - \sin\theta_n\sin\theta_{\mathrm{III}} \end{pmatrix}^T \quad (5)$$

$$p_1 = \begin{pmatrix} \cos\theta_1\cos\alpha - \sin\theta_1\cos\theta_{\mathrm{I}}\sin\alpha \\ -\sin\theta_1\cos\theta_{\mathrm{I}}\cos\alpha + \cos\theta_1\sin\alpha \\ -\sin\theta_1\sin\theta_{\mathrm{I}} \end{pmatrix}. \quad (6)$$

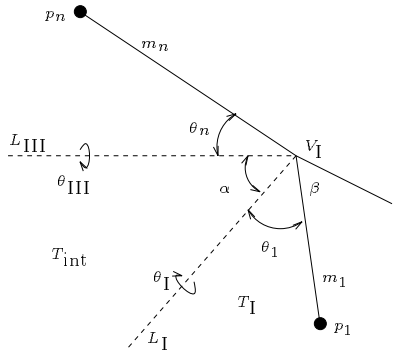

Figure 6: Angle definitions for determining rotations around segments

The derivation of θ_{I} follows:

$$\cos\beta = 1 - |p_n - p_1|^2/2 \quad (7)$$

$$\begin{aligned} |p_n - p_1|^2 = {} & 2\cos\theta_n\sin\theta_1\sin\alpha\cos\theta_{\mathrm{I}} \\ & + 2\sin\theta_n\sin\theta_1\cos\alpha\cos\theta_{\mathrm{III}}\cos\theta_{\mathrm{I}} \\ & - 2\sin\theta_n\sin\theta_1\sin\theta_{\mathrm{III}}\sin\theta_{\mathrm{I}} \\ & + 2\sin\theta_n\cos\theta_1\sin\alpha\cos\theta_{\mathrm{III}} \\ & - 2\cos\theta_n\cos\theta_1\cos\alpha + 2 \end{aligned} \quad (8)$$

$$\sin\theta_{\mathrm{I}} = \sqrt{1 - \cos\theta_{\mathrm{I}}{}^2} \quad (9)$$

$$\text{Let } A = \sin\theta_n\sin\theta_1\sin\theta_{\mathrm{III}} \quad (10)$$

$$\begin{aligned} \text{Let } B = {} & \cos\theta_n\sin\theta_1\sin\alpha \\ & + \sin\theta_n\sin\theta_1\cos\alpha\cos\theta_{\mathrm{III}} \end{aligned} \quad (11)$$

$$\begin{aligned} \text{Let } C = {} & \cos\beta + \sin\theta_n\cos\theta_1\sin\alpha\cos\theta_{\mathrm{III}} \\ & - \cos\theta_n\cos\theta_1\cos\alpha \end{aligned} \quad (12)$$

$$\text{Let } x = \cos\theta_{\mathrm{I}}. \quad (13)$$

Substituting and rearranging yields

$$A\sqrt{1 - x^2} = Bx + C \quad (14)$$

$$(A^2 + B^2)x^2 + 2BCx + C^2 - A^2 = 0. \quad (15)$$

Defining $a = A^2 + B^2$, $b = BC$ and $c = C^2 - A^2$ yields finally $ax^2 + 2bx + c = 0$.

The algorithm applies the above equations at each vertex of the internal triangle and hence *divides* the problem into three more easily solved subproblems.

There is a number of improvements to the basic algorithms that can make the computation faster and more robust. A few have been implemented, for example

- in Step 2a, θ_{I} is selected arbitrarily. A bisection root-finding loop robustly refines θ_{I} when a change of sign is detected indicating the presence of a root of the equation;
- θ_{I} is stepped through only 180 degrees to exploit the symmetry arising from $\cos\theta_{\mathrm{I}} = \cos(-\theta_{\mathrm{I}})$.

The operation of the algorithm has been verified on a number of molecules whose conformations are known. First the program implemented in C++ and running on a DEC 3000 Model 400 AXP has been applied to cyclohexane (bond lengths

of 1.536 angstroms and bond angles of 111.4°), a six-atom cyclic molecule, yielding a neighborhood of solutions as predicted by the theory for cases with non-full-rank Jacobian. The program also locates the chair conformation of cyclohexane, where the computed values for the angles are equal to the accepted experimental values of 54.6°. The calculation of all conformations of cyclohexane requires 13 ms. For this case, [7] reports a running time of 0.1 s on an IBM 360/65, and [1] lists 120 ms on a VAX 11/780.

The program has been applied to the problem of finding the conformations of larger molecules: to eight-membered rings; to cyclodecane, a ten-atom ring; and to a pentapeptide, a fifteen-membered ring. In all these cases, the solutions are infinitely many because of the additional degrees of freedom. Hence, they define a surface in an n-dimensional space. The procedure samples the surface with a resolution which depends on the increment given to the independent variables. In most cases we are interested in locating conformations of minimal energy; hence the resolution of the sampling procedure must be such that these conformations are not missed. Increments of the order of 1 to 5 degrees seem appropriate and computationally feasible. We have used increments of 5 degrees for all molecules. Given the efficiency of our procedure, the eight-atom chains require 13.7 s and cyclodecane about 23 hours. The algorithm extends easily to handle chains such as the biologically important peptide chains, which have additional constraints. The amide groups (N—C=O) in these chains constrain many of the dihedral angles to be planar. For the pentapeptide (Gly-Gly-Gly-Pro-Pro) analyzed in [1, 6], there are 15 intracyclic atoms (and so 15 dihedral angles) but only two dihedral degrees of freedom because of planar and ring constraints. The calculation of all chain-closing angles for the pentapeptide requires 21 s of CPU time.

For even larger molecules the computational task may be prohibitive. As pointed out in the introduction, in this case the resolution has to be made coarser, thus reducing substantially the probability of locating the conformers. In this case an energy optimizer can be invoked using the conformations generated by the procedure as starting points to obtain all conformers. While there is no theoretical analysis to determine what is the optimal sampling that allows an optimization algorithm to locate all conformers in minimum time, [11] found a step size of 60° was adequate for some molecules and even proposed that 120° may suffice. Being able to generate the configurations as efficiently as we can allows the population of the space with a much finer resolution and permits the optimization algorithm to run faster. The interactions between energy minimization and conformation generation need to be explored more carefully to yield an efficient conformer search and is the subject of our future work.

5 Conclusions

In this paper a procedure that finds conformations of a cyclic molecular structure is presented. The approach is purely geometric and is particularly effective for fairly large structures. An implementation of the algorithm has verified the theory, and initial results indicate that the new approach achieves an efficiency that allows complete conformational analyses of ring systems which were previously not achievable.

Future work includes the development of acceleration techniques such as the exploitation of symmetry properties of the configuration space to eliminate the search of large subsets of the space. In addition, if particular conformations are sought such as minimal energy ones (conformers), then the properties of these conformers can be used to guide the search, thus eliminating useless explorations. Another research direction is to study how best to use the conformation generation procedure to find conformers in conjunction with energy optimizers.

References

[1] R. Bruccoleri and M. Karplus. Chain closure with bond angle variations. *Macromolecules*, 18(12):2767–2773, 1985.

[2] U. Burkert and N. L. Allinger. *Molecular Mechanics.* ACS Monographs. American Chemical Society, Washington, D.C., 1982.

[3] M. D. Di Benedetto, P. Lucibello, A. L. Sangiovanni-Vincentelli, and K. Yamaguchi. A geometric approach to conformational analysis of cyclic structures. internal report, 1994.

[4] P. R. Gerber, K. Gubernator, and K. Muller. Generic shapes for the conformation analysis of microcyclic structures. *Helv. Chim. Acta*, 71:1429–1441, 1988.

[5] N. Gō. Conformational entropy of ring polymers. *Macromolecules*, 19(7):2054–2058, 1986.

[6] N. Gō and H. A. Scheraga. Calculation of the conformation of the pentapeptide cyclo-(glycylglycylglycylprolylprolyl). I. A complete energy map. *Macromolecules*, 3(2):188–194, March–April 1970.

[7] N. Gō and H. A. Scheraga. Ring closure and local conformational deformations of chain molecules. *Macromolecules*, 3(2):178–187, March–April 1970.

[8] N. Gō and H. A. Scheraga. Calculation of the conformation of cyclo-hexaglycyl. *Macromolecules*, 11(3):552–559, 1978.

[9] Hitoshi Gotō and Eiji Ōsawa. Corner flapping: A simple and fast algorithm for exhaustive generation of ring conformations. *J. Am. Chem. Soc.*, 111:8950–8951, 1989.

[10] A. E. Howard and P. A. Kollman. An analysis of current methodologies for conformational searching of complex molecules. *J. Med. Chem.*, 31(9):1669–1675, September 1988.

[11] M. Lipton and W. C. Still. The multiple minimum problem in molecular modeling. Tree searching internal coordinate conformational space. *J. Comp. Chem.*, 9(4):343–355, 1988.

[12] M. Saunders, K. N. Houk, Yun-Dong Wu, W. C. Still, M. Lipton, G. Chang, and W. C. Guida. Conformations of cycloheptadecane. A comparison of methods for conformational searching. *J. Am. Chem. Soc.*, 112(4):1419–1427, 1990.

[13] M. Saunders and H. A. Jiménez-Vázquez. Stochastic searches for lactone and cycloalkene conformers. *J. Comp. Chem.*, 14(3):330–348, 1993.

[14] S. R. Wilson, W. Cui, J. Moskowitz, and K. Schmidt. *Tetrahedron Lett.*, 29:43–73, 1988.