A deterministic global optimization approach for molecular structure determination

Costas D. Maranas and Christodoulos A. Floudas
Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544-5263

A deterministic global optimization algorithm is introduced for locating global minimum potential energy molecular conformations. The proposed branch and bound type algorithm attains finite ϵ-convergence to the global minimum through the successive refinement of converging lower and upper bounds on the solution. These bounds are obtained through a novel convex lowering bounding of the total potential function and the subsequent solution of a series of nonlinear convex optimization problems. The minimization of the total potential energy function is performed on an independent set of internal coordinates involving only dihedral angles. A number of example problems illustrate the proposed approach.

I. INTRODUCTION

The search for the global minimum potential energy conformation of a molecule faces the existence of a plethora of minima in the multivariable potential energy hyperspace making it as fascinating a subject as it is utterly complex. Wille[1] has shown, that the complexity of determining the global minimum energy of a system of atoms interacting via simple two-body forces belongs to the class NP, which implies that there is no known algorithm that can solve this simplified problem in nonexponential time.[2]

Although the basic building blocks of molecules, the atoms, remain virtually unchanged in different compounds, the versatility of the ways that they can be combined and reconfigure the resulting atom chains results in numerous different conformations for a given molecule. One of these conformations, the most stable one, is of particular importance because it dictates most of the properties of the molecule. Experimental evidence[3,4] shows that in the great majority of cases the most stable conformation corresponds to the one involving the global minimum potential energy. This enables molecular conformation identification based solely on the energetics of the interactions between the atoms composing the molecule.

Molecular mechanics calculations employ an empirically derived set of potential energy contributions for approximating these atomic interactions. This set of potential energy contributions, called the *force field*, contains adjustable parameters that are selected in a such a way as to provide the best possible agreement with experimental data. The main assumption introduced in molecular mechanics is that every parameter is associated with *a specific interaction* rather than a specific molecule. These parameters can be bond lengths; covalent bond angles; bond stretching, bending, or rotating constants; nonbonded atom interaction constants, etc. Thus whenever a specific interaction is present, the same value for the parameter can be used even if this interaction occurs in different molecules.[5] Note that experimental results provide sufficient evidence that it is a reasonable assumption in most cases. Many different models have been proposed for approximating the *force field*, and some of the most popular ones are ECEPP,[6–8] MM2,[9] ECEPP/2,[10] CHARMM,[11] AMBER,[12] GROMOS87,[13] MM3,[14] and ECEPP/3.[15]

A very large number of methods have been proposed for finding the most stable conformation of a molecule through the identification of the global minimum point of the potential energy surface. Most methods attempt to locate this point by tracing paths on the potential energy surface conjecturing that some of them will converge to the global minimum point. Molecular dynamics (MD) methods trace the time evolution or trajectory of a molecule, described by the principles of Newtonian mechanics, by integrating the equations of motion. The necessity of taking small step sizes limits the ability of molecular mechanics to simulate long trajectories. Despite the above caveat, people have utilized this method for obtaining conformational possibilities for small and moderately sized, as well as macromolecular systems.[16–27] In Monte Carlo (MC) simulations the dynamic behavior of a molecule is examined by performing random perturbations in the positions of the atoms. A trial configuration is accepted if it results in reduction of the total potential energy. Otherwise, it is accepted with a probability set by typically the Metropolis algorithm.[28] Although the Monte Carlo method has been primarily used for the simulation of liquids, variations of this method in conjunction with other methods (e.g., molecular dynamics, energy minimization, electrostatics, adaptive importance sampling) have been used to study conformations of small molecules and constrained motions in proteins.[29–39]

Simulated annealing (SA), introduced by Kirkpatrick *et al.*,[40] is a stochastic nonequilibrium procedure designed to cope with large energy barriers, based on an analogy to the process by which a many-body system is brought to a low energy state by annealing. A number of researchers have recently used simulated annealing to locate good molecular conformations.[41–43] A similar approach with simulated annealing is the one based on stochastic conformational searches on a number of candidate conformations. These initial conformations are either selected independently, without sharing of information,[45–50] or alternatively the history of the optimization process is utilized in the selection of the next candidate conformation.[51]

Relaxation of dimensionality (RD) methods are mainly motivated by the fact that the number of local minima of the potential energy function are greatly reduced by relaxing the requirement of maintaining the molecule in the three-dimensional space and allowing it to span a higher than a three dimensional space. By doing so individual potential energy terms can be *independently* driven to their respective minima since the additional dimensions absorb some or even all overlapping constraints. The heart of the problem, however, is how to project the molecule back to the three-dimensional space in such a way that the increase in potential energy is minimum. This is accomplished either by the use of the Caley–Menger determinants,[52,53] or by energy embedding.[54–58]

Distance geometry methods attempt to satisfy external constraints (e.g., maintain covalently bonded atoms at their equilibrium bond length separation; keep nonbonded atoms from getting closer than the sum of their van der Waals radii; satisfy supplied upper and lower bounds on interatomic distances), rather than directly minimizing the potential energy of the resulting configuration. They have been applied to small as well as larger molecular systems.[59–63] In spite of the simplicity and elegance of distance geometry methods, the indirectly performed minimization of the total potential energy is not always adequate. A technique which is conceptually related to the distance geometry methods but avoids working on the atomic coordinate space is the ellipsoid algorithm.[64,65] The advantage of this method is that the dimensionality of the problem is defined by the number of independent dihedral angles whereas in distance geometry the dimensionality is determined by the Cartesian coordinates of all atoms.

Gradient type methods, in connection with other approaches, have been extensively used for locating minimum energy conformations of molecules.[66,67] It must be noted, however, that since the total potential energy function is in general nonconvex gradient type methods can guarantee convergence to a local minimum at best. By recognizing that bond lengths and bond angles do not deviate considerably from their equilibrium values in molecules, internal coordinate systematic search methods attempt to generate molecular conformations by systematically varying each of the dihedral angles in a molecule by some small increment, while keeping the bond lengths and bond angles fixed. Grid search methods utilizing sufficiently small increments,[68,69] in conjunction with tree searching algorithms,[70,71] and filtering algorithms[72,73] have been applied to a number of conformational problems.

Finally, a host of diverse approaches including "build up" methods,[74–77] the random incremental pulse search (RIPS) method,[78–80] optimization of electrostatics (SCEF),[81] neural networks,[82,83] genetic algorithms,[84] dynamic programming,[85] pattern recognition importance sampling minimization (PRISM),[86–88] the diffusion equation method (DEM),[89–91] sequence homology methods,[92,93] the probabilistic quasiquantal (QQ) method,[94,95] and the self-consistent multitorsional field (SCMTF) method[96] have been proposed to predict the structure of macromolecules.

Despite the plethora of the currently available methods and the intriguing ways that they propose to circumvent the existence of myriads of local minima in the potential energy hypersurface, little progress has been made towards proving convergence to the global minimum. The key limitation shared by all the aforementioned methods is that unless there is a single potential well, the obtained minimum energy conformation depends heavily on the supplied initial conformation. This is why in practice many trial geometries serve as initial points for the employed optimization method. These geometries are usually chosen from Dreiding models, or other similar considerations and thus there is no guarantee that important conformations are not overlooked. Therefore, the obtained minimum energy conformations are limited by which initial configurations seemed appropriate to the researcher.[97] The need for a method that can guarantee convergence to the global minimum potential energy conformation motivated our initial effort to introduce such a method for microclusters,[98,99] and small molecules[100] interacting with relatively simple force fields and apply it to a number of molecular conformational problems in this work.

II. PROBLEM DEFINITION AND ANALYSIS

The main focus of this work is to introduce a systematic procedure for locating the most stable conformation of a given molecule in the three-dimensional space based solely on the energetics of the interactions between the atoms that compose the molecule. The simplifications employed herein start with the well established conjecture in molecular mechanics that the most stable conformation of the molecule is the one involving the global minimum potential energy. Tractable expressions for this potential energy function are estimated based on the Born–Oppenheimer approximation. Because the most stable conformations of nonpolar molecules are sought in this work, only pairwise, additive, two-body interaction terms are considered. Furthermore, since covalent bond lengths and angles do not deform significantly from their equilibrium values without substantial increase in the potential energy of the molecule, they are assumed to remain fixed at their equilibrium values, which is a fairly good approximation in most cases.

Under the aforementioned simplifications the expression for the total potential energy V of a molecule involves only the sum of a number of pairwise potential interaction terms. Each pairwise potential term is a function of only the Euclidean distances r_{ij} between the interacting atoms i and j which are directly related to the Cartesian coordinates of the atoms forming the molecule. Therefore, V can be fully represented in the coordinate space x_i, y_i, z_i of the atomic coordinates.

Note, however, that a number of equality constraints must be included in the formulation to reflect the fact that all covalent bond lengths and covalent bond angles are assumed to be fixed at their equilibrium values. After defining $\mathscr{B}$ to be the set of bonded atoms and $\mathscr{NB}$ the set of nonbonded atoms, V can be expressed in the Cartesian atomic coordinate space as follows:

$$V= \sum_{\substack{i<j \\ ij\in \mathcal{NB}}} v(r_{ij}),$$

where

$$r_{ij}= \sqrt{(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2}, \quad \forall ij\in \mathcal{NB},$$

subject to

$$r_{ij}^0= \sqrt{(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2}, \quad \forall ij\in \mathcal{B},$$

$$\theta_{ijk}^0= -\frac{(x_i-x_j)(x_j-x_k)+(y_i-y_j)(y_j-y_k)+(z_i-z_j)(z_j-z_k)}{r_{ij}^0 r_{jk}^0},$$

$\forall ijk$ such that $ij,jk\in \mathcal{B}$,

and

$$x_1=y_1=z_1=y_2=z_2=z_3=0,$$

where V is the total potential energy of the molecule; $v(r_{ij})$ is the pairwise potential expression; r_{ij} is the Euclidean distance between nonbonded atom i with atom j; r_{ij}^0, θ_{ijk}^0 are the fixed covalent bond lengths and angles, respectively; and x_i, y_i, z_i are the atomic Cartesian coordinates. Note that in the summation $i<j$ so that we avoid double-counting pair interactions and the interaction of an atom with itself. Furthermore, by imposing $x_1=y_1=z_1=y_2=z_2 =z_3=0$ we eliminate the translational and rotational degrees of freedom of the molecule.

The use of Cartesian coordinates in local optimization techniques greatly facilitates the calculation of the total potential energy V. However, by employing Cartesian coordinates we introduce a number of complex, highly nonlinear equality constraints, one for each r_{ij}^0, θ_{ijk}^0, which are very difficult to handle in the context of a deterministic global optimization algorithm. On the other hand, the set of independent internal coordinates reduces to a set of independent torsion angles because the covalent bond lengths and angles are assumed to be constant. Therefore, it suffices to select one torsion angle per rotational degree

of freedom. In polycyclic molecules, however, it may become more involved to define a set of independent internal coordinates because the bond and torsion angles are correlated. In principle, the independent set of torsion angles is smaller than the set of atomic coordinates. This is a considerable advantage for global optimization algorithms where the number of variables is typically the bottleneck of the computational effort. Therefore, in this work independent internal coordinates rather than Cartesian coordinates are employed. The total potential energy can then be written in the set of independent internal coordinates as follows:

$$V= \sum_{\substack{i<j \\ ij\in \mathcal{NB}}} v[r_{ij}(t_k)],$$

where t_k is the set of independent internal coordinates, $k=1,...,K$. Note that the above formulation does not involve any equality constraints; however, the functionality between r_{ij} and the set of independent internal variables (independent torsion angles) needs to be established.

Although it is quite straightforward to express all r_{ij}'s in the Cartesian coordinate set, unfortunately this is not the case when r_{ij} must be expressed as a function of an independent set of internal coordinates. In doing so it is first necessary to establish a connecting path, formed by a sequence of covalent bonds, between every atom i and atom j. For all pairs of atoms (i,j) in a molecule there is always an acyclic chain of atoms, connected by covalent bonds, which links atom i with atom j, because the connectivity of atoms in a molecule always defines a *undirected connected graph*.[101] This introduces a natural way of *partitioning* the set of nonbonded r_{ij}'s according to the minimum number of atoms participating in an acyclic chain connecting atom i with atom j. The aforementioned partitioning of r_{ij}'s can be realized with the definition of the following sets:

$$\mathcal{B}=\{(i,j): i,j \text{ bonded atoms}\}$$
$$\mathcal{NB}=\{(i,j): i,j \text{ nonbonded atoms}\}$$
$$\mathcal{B}^2=\{(i,j): i<j, \text{ and } (i,j)\in \mathcal{B}\}$$
$$\vdots \quad = \quad \vdots$$
$$\mathcal{B}^m=\{(i_1,i_2,...,i_m): i_1<i_m, \text{ and}$$
$$(i_1,i_2)\in \mathcal{B},...,(i_{m-1},i_m)\in \mathcal{B},$$

where $(i_1, i_2,...,i_m)$ acyclic chain}.
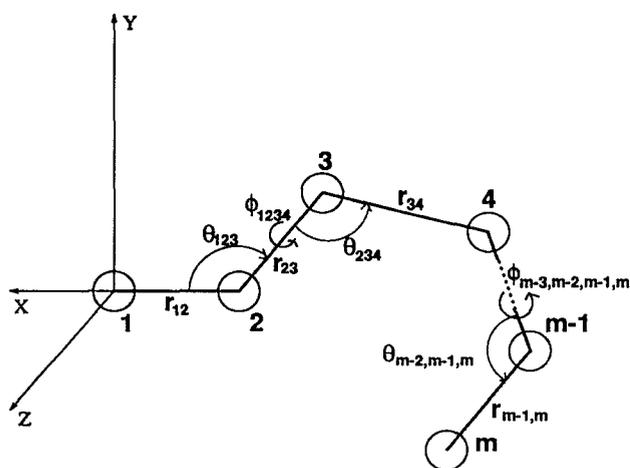
FIG. 1. Coordinate set of atomic chain.

Note that $\mathscr{B}$ is the set of pairs of atoms which are connected with covalent bonds and $\mathscr{NB}$ is the set of pairs of atoms which are not. Also, $\mathscr{B}^2,...,\mathscr{B}^m$, correspond to *directed* sets of chains of covalently bonded atoms involving two, three, and m atoms, respectively.

If only a single atom (2) is between atoms (1) and (3) such that $(1,2,3)\in\mathscr{B}^3$, then it is well known that,

$$r_{13}^2=r_{12}^2+r_{23}^2-2r_{12}r_{23}\cos(\theta_{123}),$$

where θ_{123} is the angle formed by covalent bonds (1–2) and (2–3). However, when more than one atom is required to connect atoms i and j the derivation of r_{ij} is much more involved (see Fig. 1). The coordinates of the mth atom in an m-atom acyclic chain, composed by atoms connected sequentially with covalent bonds, in the coordinate system that the first three atoms define are given by[102]

$$\begin{bmatrix}x_m\\y_m\\z_m\\1\end{bmatrix}=B_2 B_3\cdots B_m\begin{bmatrix}0\\0\\0\\1\end{bmatrix},$$

where B_2, B_3,..., and B_m are 4×4 transformation matrices

$$B_2=\begin{bmatrix}-1 & 0 & 0 & -r_{12}\\0 & 1 & 0 & 0\\0 & 0 & -1 & 0\\0 & 0 & 0 & 1\end{bmatrix},$$

$$B_3=\begin{bmatrix}-\cos(\theta_{123}) & -\sin(\theta_{123}) & 0 & -r_{23}\cos(\theta_{123})\\\sin(\theta_{123}) & -\cos(\theta_{123}) & 0 & r_{23}\sin(\theta_{123})\\0 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix},$$

and B_m is

$$B_m=\begin{bmatrix}-\cos(\theta_m) & -\sin(\theta_m) & 0 & -r_m\cos(\theta_m)\\\sin(\theta_m)\cos(\phi_m) & -\cos(\theta_m)\cos(\phi_m) & -\sin(\phi_m) & r_m\sin(\theta_m)\cos(\phi_m)\\\sin(\theta_m)\sin(\phi_m) & -\cos(\theta_m)\sin(\phi_m) & \cos(\phi_m) & r_m\sin(\theta_m)\sin(\phi_m)\\0 & 0 & 0 & 1\end{bmatrix}$$

following the simplified notation,

$$r_{m-1,m}\leftarrow r_m,\quad \theta_{m-2,m-1,m}\leftarrow\theta_m,$$

$$\phi_{m-3,m-2,m-1,m}\leftarrow\phi_m.$$

Based on the above analysis, explicit expressions for the Cartesian coordinates x_m, y_m, z_m of the mth atom in a m-atom chain can be obtained as functions of the bond lengths, covalent bond angles, and torsion angles. Since $x_1=y_1=z_1=0$, the squared Euclidean distance r_{1m}^2 is,

$$r_{1m}^2=x_m^2+y_m^2+z_m^2.$$

After expanding the squared terms, the expression for r_{1m}^2 becomes the sum of a very large number of terms involving the product of bond lengths with sines and/or cosines of covalent bond lengths and/or dihedral (torsion) angles. Simplified explicit expressions for r_{1m}^2, $m=2,...,6$ as functions of ϕ_{ijkl}

$$r_{1m}^2=r_{1m}^2(\phi_{ijkl},ijkl\in\mathscr{B}^4)$$

are obtained via symbolic computations and are given in Appendix A of Ref. 100. Note that, although in this paper explicit expressions of the Euclidean distances as functions of dihedral angles are utilized this is not a requirement of the proposed global optimization algorithm. In fact, work currently under way have shown to us that it is possible to couple the local optimization algorithm with a procedure designed to efficiently evaluate all nonbonded atoms Euclidean distances for a given set of values for the dihedral angles. This negates the need of having explicit expressions

for the Euclidean distances as functions of dihedral angles which can potentially become the bottleneck of our approach for larger molecules.

Note, however, that not all dihedral angles are independent. In fact, for acyclic molecules, a set of independent dihedral angles can be defined involving one independent dihedral angle per rotational degree of freedom. For example, propanelike molecules require only three dihedral angles in order to fully specify all other dihedral angles and therefore the molecular conformation. In general, for each dihedral angle ϕ_{ijkl} of an acyclic molecule an independent dihedral angle t_k can be found such as,

$$\phi_{ijkl}=t_k+\omega_{ijkl},$$

where ω_{ijkl} counts how many degrees ϕ_{ijkl} is lagging or is ahead of t_k. This means that all nonbonded distances r_{ij}, and consequently the expression for the total potential energy V can be expressed as a function of only these independent dihedral angles, defined one for each rotational degree of freedom

$$V=\sum_{j=4}^{m}\sum_{(i_1,...,i_j)\in\mathscr{B}^j}v(r_{i_1,i_j})$$

where

$$r_{i_1,i_j}=r_{i_1,i_j}[\phi_{ijkl},(i,j,k,l)\in\mathscr{B}^4],$$

$$\forall(i_1,...,i_j)\in\mathscr{B}^j,\quad j=4,...,m$$

and $\forall (i,j,k,l) \in \mathcal{B}^4$, $\exists k$, ω_{ijkl} such that $\phi_{ijkl} = t_k + \omega_{ijkl}$. Where m is the number of atoms in the longest chain connecting two interacting atoms; r_{i_1,i_j} is the expression derived earlier for the Euclidean distance between the two end atoms i_1 and i_j; and $v(r_{i_1,i_j})$ is the pairwise potential interaction function. Note that j starts from ($j = 4$) because there must be at least two atoms between the interacting atoms before considering their contribution to the total potential energy of the molecule.[97]

This implies that the problem of minimizing the total potential energy of an acyclic molecule can be formulated as an unconstrained optimization problem in the space of independent dihedral angles. The expression, however, for the total potential energy function is nonconvex therefore existing optimization algorithms cannot guarantee convergence to the global minimum. To overcome this problem, a deterministic branch and bound type global optimization algorithm is proposed which locates the global minimum solution by constructing converging lower and upper bounds. These bounds are successively refined by iteratively partitioning the initial feasible region into many subregions. Upper bounds to the global minimum can be obtained by simply calculating the value of V at some point. Lower bounds, however, are more difficult to derive. One way is to underestimate the original nonconvex total potential energy function with a convex function whose single global minimum can be routinely found. This global minimum can then serve as a lower bound on the global minimum solution of the total potential energy function. In the next section, such a convex lower bounding function is proposed and its properties are analyzed.

III. CONVEX LOWER BOUNDING OF V

A convex lower bounding function $\mathcal{L}$ of V can be defined by augmenting V using the same principles that have been proposed in Refs. 98 and 99

$$\mathcal{L} = V + \alpha \sum_{k=1}^{K} (t_k^L - t_k)(t_k^U - t_k)$$

where

$$\alpha \geq \max\{ \max_{\substack{k \\ t_k^L < t_k < t_k^U}} 0, (-\tfrac{1}{2}\lambda_k^V)\}.$$

Note that t_k^L, t_k^U correspond to the lower and upper bounds on t_k which are typically set to $t_k^L = 0$ and $t_k^U = 2\pi$. Also, α is a nonnegative parameter which must be greater or equal to the negative one half of the minimum eigenvalue of V over $t_k^L < t_k < t_k^U$. This parameter α can be estimated through the solution of an optimization problem, or with the use of the concept of the measure of a matrix.[100] The effect of adding the term,

$$\alpha \sum_{k=1}^{K} (t_k^L - t_k)(t_k^U - t_k)$$

to V is to make $\mathcal{L}$ convex by overpowering the nonconvexity characteristics of V with the addition of the term 2α to all of its eigenvalues,

$$\lambda_k^{\mathcal{L}} = \lambda_k^V + 2\alpha.$$

Here $\lambda_k^{\mathcal{L}}$, λ_k^V are the kth eigenvalues of $\mathcal{L}$, V, respectively. This function $\mathcal{L}$ involves a number of very important properties which enable us to construct a global optimization algorithm for finding the global minimum of V in the space defined by the independent dihedral angles t_k. These properties, whose proof is given in Ref. 100, are illustrated in the following.

Property 1: $\mathcal{L}$ is always a valid *underestimator* of V inside the box constraints.

$$\forall t_k \in [t_k^L, t_k^U], \quad \mathcal{L}(t_k) \leq V(t_k).$$

Property 2: $\mathcal{L}$ matches V at all corner points.

$$\forall t_k \text{ such that } t_k = t_k^L \text{ or } t_k = t_k^U, \quad \mathcal{L}(t_k) = V(t_k).$$

Property 3: $\mathcal{L}$ is *convex* in $[t_k^L, t_k^U]$.

Property 4: The maximum separation between $\mathcal{L}$ and V is *bounded* and proportional to α and to the square of the diagonal of the current box constraints

$$\max_{t_k^L < t_k < t_k^U} (V - \mathcal{L}) = \frac{1}{4}\alpha \sum_{k=1}^{K} (t_k^U - t_k^L)^2.$$

Property 5: The underestimators constructed over supersets of the current set are always *less tight* than the underestimator constructed over the current box constraints for every point within the current box constraints. Therefore, the values of different definitions of $\mathcal{L}$ at any point, if $\mathcal{L}$ is constructed over a tighter box of constraints each time, define a *nondecreasing* sequence. $\forall t_k \in [t_k^{L1}, t_k^{U1}]$, $\forall [t_k^{L2}, t_k^{U2}]$ such that $[t_k^{L1}, t_k^{U1}] \subseteq [t_k^{L2}, t_k^{U2}]$, $\mathcal{L}(t_k, \alpha^1, t_k^{L1}, t_k^{U1}) \geq \mathcal{L}(t_k, \alpha^2, t_k^{L2}, t_k^{U2})$, after imposing $\alpha^2 \geq \alpha^1$ in the selection process for α^1, α^2, where

$$\alpha^1 \geq \max \left\{ 0, \quad \max_{\substack{k \\ t_k^{L1} < t_k < t_k^{U1}}} [-\tfrac{1}{2}\lambda_k^V(t_k)] \right\}$$

and

$$\alpha^2 \geq \max \left\{ 0, \quad \max_{\substack{k \\ t_k^{L2} < t_k < t_k^{U2}}} [-\tfrac{1}{2}\lambda_k^V(t_k)] \right\}.$$

Furthermore, in Appendix C of Ref. 100 it is shown that function $\mathcal{L}$ corresponds to a relaxed dual bound of the original function V.[103-107] In the next section, based on the aforementioned properties, a branch and bound type global optimization algorithm is discussed.

IV. GLOBAL OPTIMIZATION ALGORITHM

Based on the properties introduced in the previous section a deterministic branch and bound type global optimization algorithm is proposed for locating the global minimum potential energy V^* by constructing converging lower and upper bounds on V^*. A lower bound on V^*, denoted as V^L, within some box constraints is derived by invoking *Properties (1) and (3)*. Based on these properties $\mathcal{L}$ is a convex lower bounding function of V. Therefore, its single global minimum within some box constraints is a

valid lower bound V^L on the global minimum solution V^* and can be guaranteed to be found with available local optimization algorithms. An upper bound on V^*, denoted as V^U, is then simply the value of V at the global minimum point of $\mathscr{L}$.

From *Property (4)* we know that the gap between these upper and lower bounds $V^U - V^L$ is at most,

$$(V^U - V^L) < \frac{1}{4} \alpha \sum_{k=1}^{K} (t_k^U - t_k^L)^2.$$

The next step, after establishing an upper and a lower bound on the global minimum, is to refine them by using *Property (5)*. This property implies that the value of $\mathscr{L}$ at every point, and therefore at its global minimum, is increased by restricting the box constraints within which it has been defined. Tighter box constraints can be realized by partitioning the rectangle that the initial box constraints define into a number of smaller rectangles. One way of partitioning is to successively divide the current rectangle in two subrectangles by halving on the middle point of the longest side of the initial rectangle (bisection). At each iteration the lower bound of V^* is simply the minimum over all the minima of $\mathscr{L}$ in every subrectangle composing the initial rectangle. Therefore, a straightforward (bound improving) way of tightening the lower bound V^L is to halve at each iteration, only the subrectangle responsible for the infenum of the minima of $\mathscr{L}$ over all subrectangles, according to the rules discussed earlier. This procedure generates a *nondecreasing* sequence for the lower bound V^L of V^*. Furthermore, we construct a *nonincreasing* sequence for the upper bound V^U by selecting it to be the infenum over all the previously recorded upper bounds. Clearly, if the global minimum of $\mathscr{L}$ in any subrectangle is greater than the current upper bound V^U we can ignore this subrectangle because the global minimum of V cannot be situated inside it (fathoming step).

Property (4) answers the question of how small these subrectangles must become before the upper and lower bounds of V inside these subrectangles are within ϵ. If δ is the diagonal of the subrectangle,

$$\delta = \sqrt{\sum_{k=1}^{K} (t_k^U - t_k^L)^2}$$

and ϵ is the convergence tolerance, from *Property (4)* we have the following condition for convergence:

$$\epsilon > \frac{1}{4} \alpha \sum_{k=1}^{K} (t_k^U - t_k^L)^2 = \frac{1}{4} \alpha \delta^2 > V^U - V^L$$

which means that if the diagonal δ of a subrectangle is,

$$\delta < \sqrt{\frac{4\epsilon}{\alpha}}$$

then ϵ convergence to the global minimum of V has been achieved. It is interesting to note that the required for convergence value of δ is proportional to the square root of ϵ. Therefore, if for example ϵ is set to be 0.0001, δ suffices to be proportional to 0.01. Note also, that δ is inversely proportional to the squared root of the parameter α reflect-

ing the fact that the smaller the value of α is, the faster the convergence rate becomes. Proof of ϵ convergence to the global minimum of the proposed global optimization algorithm is given in Ref. 100. The basic steps of the proposed global optimization algorithm are summarized in the following section.

V. STEPS OF THE GLOBAL OPTIMIZATION ALGORITHM

Step 1: Initialization. A convergence tolerance ϵ is selected and the iteration counter Iter is set to one. Appropriate global bounds t_k^{LBD}, t_k^{UBD} on t_k are chosen and local bounds $t_k^{L,\text{Iter}}$, $t_k^{U,\text{Iter}}$ for the first iteration are set to be equal to the global ones. Lower and upper bounds on the global minimum V^{LBD}, V^{UBD} are initialized and an initial current point $t_k^{c,\text{Iter}}$ is selected.

Step 2: Update of upper bound V^{UBD}. V is calculated at the current point $V^{c,\text{Iter}}$ and the upper bound V^{UBD} is updated as follows:

$$V^{\text{UBD}} = \min[V^{\text{UBD}}, V(t_k^{c,\text{Iter}})].$$

Step 3: Partitioning of current rectangle. The current rectangle $[t_k^{L,\text{Iter}}, t_k^{U,\text{Iter}}]$, $k=1,...,K$ is partitioned into the following two rectangles ($r=1,2$):

$$\begin{bmatrix} t_1^{L,\text{Iter}} & t_1^{U,\text{Iter}} \\ \vdots & \vdots \\ t_{l^{\text{Iter}}}^{L,\text{Iter}} & \dfrac{(t_{l^{\text{Iter}}}^{L,\text{Iter}} + t_{l^{\text{Iter}}}^{U,\text{Iter}})}{2} \\ \vdots & \vdots \\ t_K^{L,\text{Iter}} & t_K^{U,\text{Iter}} \end{bmatrix}, \begin{bmatrix} t_1^{L,\text{Iter}} & t_1^{U,\text{Iter}} \\ \vdots & \vdots \\ \dfrac{(t_{l^{\text{Iter}}}^{L,\text{Iter}} + t_{l^{\text{Iter}}}^{U,\text{Iter}})}{2} & t_{l^{\text{Iter}}}^{U,\text{Iter}} \\ \vdots & \vdots \\ t_K^{L,\text{Iter}} & t_K^{U,\text{Iter}} \end{bmatrix},$$

where l^{Iter} corresponds to the variable with the scaled longest side in the initial rectangle,

$$l^{\text{Iter}} = \arg \max_k \frac{(t_k^{U,\text{Iter}} - t_k^{L,\text{Iter}})}{(t_k^{\text{UBD}} - t_k^{\text{LBD}})} .$$

Step 4: Solution of convex problems in two subrectangles. Update the parameter α and solve the following convex nonlinear optimization problem in both subrectangles ($r=1, 2$) by using a nonlinear solver:

$$\min_{t_k} \mathscr{L} = \sum_{j=4}^{m} \sum_{(i_1,...,i_j) \in \mathscr{B}^j} v[r_{i_1,i_j}(t_k)] + \alpha \sum_{k=1}^{K} (t_k^{L,\text{Iter}} - t_k)$$
$$\times (t_k^{U,\text{Iter}} - t_k).$$

If a solution $L_{\text{sol}}^{r,\text{Iter}}$ is less than the current upper bound,

$$L_{\text{sol}}^{r,\text{Iter}} \leq V^{\text{UBD}}$$

then it is stored along with the value of the variable t_k at the solution point $t_{k,\text{sol}}^{r,\text{Iter}}$.

Step 5: Update iteration counter Iter and lower bound V^{LBD}. The iteration counter is increased by one,

$$\text{Iter} \leftarrow \text{Iter} + 1$$

and the lower bound V^{LBD} is updated to the minimum solution over the stored ones from previous iterations. Furthermore, the selected solution is erased from the stored set

$$V^{LBD} = L_{sol}^{r',Iter'}$$

where

$$L_{sol}^{r',Iter'} = \min_{r,I} L_{sol}^{r,I}, \quad r=1,2, \quad I=1,...,Iter-1.$$

Step 6: Update current point $t_k^{c,Iter}$ and Bounds $t_k^{L,Iter}$, $t_k^{U,Iter}$ on t_k. The current point is selected to be the solution point of the previously found minimum solution in Step 5,

$$t_k^{c,Iter} = t_{k,sol}^{r',Iter'}, \quad k=1,...,K$$

and the current rectangle becomes the subrectangle containing the previously found solution,

$$[t_k^{L,Iter}, t_k^{U,Iter}] = \begin{bmatrix} t_1^{L,Iter'} & t_1^{U,Iter'} \\ \vdots & \vdots \\ t_{fIter'}^{L,Iter'} & \dfrac{(t_{fIter'}^{L,Iter'} + t_{fIter'}^{U,Iter'})}{2} \\ \vdots & \vdots \\ t_K^{L,Iter'} & t_K^{U,Iter'} \end{bmatrix}, \quad \text{if } r'=1,$$

$$[t_k^{L,Iter}, t_k^{U,Iter}] = \begin{bmatrix} t_1^{L,Iter'} & t_1^{U,Iter'} \\ \vdots & \vdots \\ \dfrac{(t_{fIter'}^{L,Iter'} + t_{fIter'}^{U,Iter'})}{2} & t_{fIter'}^{U,Iter'} \\ \vdots & \vdots \\ t_K^{L,Iter'} & t_K^{U,Iter'} \end{bmatrix}, \quad \text{if } r'=2.$$

Step 7: Check for convergence. IF $(V^{UBD} - V^{LBD}) > \epsilon$, then return to Step 2. Otherwise, ϵ convergence has been reached and the global minimum solution, and solution point are

$$V^* \leftarrow V^{U,Iter''},$$

$$t_k^* \leftarrow t_k^{c,Iter''}, \quad k=1,...,K,$$

where

$$Iter'' = \arg_I \{V^{U,I} = V^{UBD}\}, \quad I=1,...,Iter.$$

Proof of ϵ convergence to the global minimum can be found in Ref. 100. In the next section, the proposed approach is applied to a number of molecular conformation problems.

VI. EXAMPLES

Based on the analysis and the assumptions presented in the previous sections, the problem of minimizing the total potential energy of a molecule can now be explicitly formulated in the coordinate space of independent dihedral (torsion) angles t_k
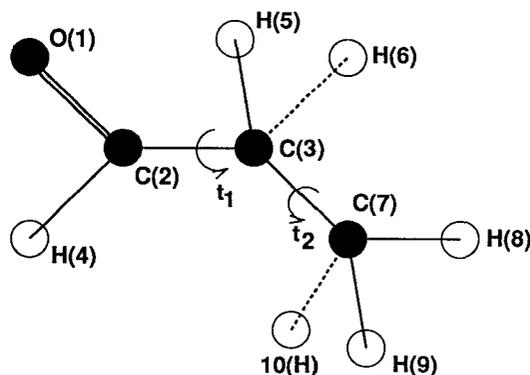
FIG. 2. Propanal molecule.

$$\min_{t_k} V = \sum_{j=4}^{m} \sum_{(i_1,...,i_j) \in \mathscr{B}^j} v(r_{i_1,i_j}),$$

where

$$r_{i_1,i_j} = r_{i_1,i_j}[\phi_{ijkl}, (i,j,k,l) \in \mathscr{B}^4],$$

$$\forall (i_1,...,i_j) \in \mathscr{B}^j, \quad j=4,...,m,$$

and $\forall (i,j,k,l) \in \mathscr{B}^4$, $\exists k$, ω_{ijkl} such that $\phi_{ijkl} = t_k + \omega_{ijkl}$

$$t_k^L \leqslant t_k \leqslant t_k^U, \quad k=1,...,K.$$

Note that V is the total potential energy composed by the sum of all pairwise potential interactions $v(r_{i_1,i_j})$ which in turn are functions of only the corresponding interatomic distances r_{i_1,i_j}. These distances are defined over all unique acyclic j-atom chains $(i_1,...,i_j)$ belonging to $\mathscr{B}^j$. Note also that j spans from $j=4$, avoiding constant distance atom chains for $j=2$, 3, to $j=m$ where m is the longest acyclic chain in the molecule. Explicit expressions for the interatomic Euclidean distances r_{i_1,i_j} for all acyclic atom chains $[(i_1,...,i_j), j=4,...,m]$ in terms of dihedral angles ϕ_{ijkl}, $(i,j,k,l) \in \mathscr{B}^4$ are provided in Appendix A of Ref. 100. Based on the topology of the molecule an independent set of dihedral angles t_k, $k=1,...,K$ can be selected so as each other dihedral angle ϕ_{ijkl} differs by a constant factor ω_{ijkl} from some independent dihedral angle t_k. Finally, lower and upper bounds t_k^L, t_k^U typically 0 and 2π are provided for the independent dihedral angles t_k.

In the following subsections, five example problems will be considered. For each individual example problem the expression for the selected pairwise potential energy function is provided along the necessary parameters. Furthermore, from the set of dihedral angles a set of indepen-

TABLE I. Covalent bond lengths r_0 in the propanal example.

Atom–atom	r_0 (Å)
C=O	1.208
C–C	1.509
C–H(ald)	1.113
C–H	1.113

TABLE II. Covalent bond angles θ_0 in the propanal example.

Atom–atom–atom	θ_0
O=C–H(ald)	120.00°
O=C–C	120.00°
H(ald)–C–C	120.00°
C–C–C	109.50°
C–C–H	109.50°
H–C–H	109.50°

FIG. 3. 1,2,3-trichloro-1-fluoro-propane molecule.

dent dihedral angles is selected and explicit relations between dependent and independent dihedral angles are established. Finally, information on the computational performance of the proposed global optimization algorithm as well as of a local optimization algorithm, MINOS 5.3[108] are summarized for each individual example.

A. Propanal

The first example is the problem of finding the global minimum potential energy conformation of the propanal molecule (CH_3CH_2CHO) (see Fig. 2). This molecule involves fifteen chains of four covalently bonded atoms and six chains of five covalently bonded atoms. The values of the covalent bond lengths and covalent bond angles are assumed to be fixed at their equilibrium values and they are given in Table I and Table II, respectively.[109] Note also that the Buckingham potential function was selected to model the nonbonded atom interactions

$$v(r_{ij}) = A_{ij} \exp(B_{ij}r_{ij}) + \frac{C_{ij}}{r_{ij}^6}.$$

The values of the interaction parameters A_{ij}, B_{ij}, and C_{ij} are given in Table III.[109]

Clearly, this problem involves two independent internal coordinates. One possible selection can be the dihedral angle t_1 between the planes defined by O(1)=C(2)–C(3) and C(2)–C(3)–H(5), and t_2 the dihedral angle between the planes H(5)–C(3)–C(7) and C(3)–C(7)–H(8). The bounds on t_1, t_2 are as follows:

$$0 \leqslant t_1 \leqslant 2\pi, \quad 0 \leqslant t_2 \leqslant \frac{2\pi}{3}.$$

All other dihedral angles are,

$$\phi_{O(1)=C(2)-C(3)-H(5)} = t_1,$$

$$\phi_{O(1)=C(2)-C(3)-H(6)} = t_1 + \frac{2\pi}{3},$$

$$\phi_{O(1)=C(2)-C(3)-C(7)} = t_1 - \frac{2\pi}{3},$$

$$\phi_{C(4)-C(2)-C(3)-H(5)} = t_1 + \pi,$$

$$\phi_{C(4)-C(2)-C(3)-H(6)} = t_1 - \frac{\pi}{3},$$

$$\phi_{C(4)-C(2)-C(3)-C(7)} = t_1 + \frac{\pi}{3},$$

$$\phi_{H(5)-C(3)-C(7)-H(8)} = t_2,$$

$$\phi_{H(5)-C(3)-C(7)-H(9)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{H(5)-C(3)-C(7)-H(10)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{H(6)-C(3)-C(7)-H(8)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{H(6)-C(3)-C(7)-H(9)} = t_2,$$

$$\phi_{H(6)-C(3)-C(7)-H(10)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{C(2)-C(3)-C(7)-H(8)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{C(2)-C(3)-C(7)-H(9)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{C(2)-C(3)-C(7)-H(10)} = t_2.$$

TABLE III. Data for parameters A_{ij}, B_{ij}, and C_{ij} in the propanal example.

Atom–atom	A (kcal/mol)	B (Å^{-1})	C (kcal Å^6/mol)
H–O	16 152.0	−3.858 00	−144.970
H(ald)–C	15 628.0	−3.434 10	−282.030
H–C	13 188.0	−3.633 70	−169.550
H(ald)–H	13 630.0	−4.166 77	−77.0920
H–H	13 630.0	−4.166 77	−77.0920

TABLE IV. Covalent bond lengths r_0 in the 1,2,3-trichloro-1-fluoro-propane example.

Atom–atom	r_0 (Å)
C–C	1.54
C–H	1.06
C–Cl	1.77
C–F	1.39

For the given set of data the following solutions have been found by repeatedly solving the problem of minimizing the total potential energy with the local optimization solver MINOS 5.3[108] using randomly selected starting points.

$$t_1 = 41.17°, \quad t_2 = 59.84° \to V_{min} = 0.769\ 101 \text{ kcal/mol},$$

$$t_1 = 198.83°, \quad t_2 = 60.16° \to V_{min} = 0.769\ 101 \text{ kcal/mol},$$

$$t_1 = 300.00°, \quad t_2 = 60.00° \to V_{min} = 0.830\ 866 \text{ kcal/mol},$$

$$t_1 = 360.00°, \quad t_2 = 60.06° \to V_{min} = 0.844\ 303 \text{ kcal/mol}.$$

The proposed global optimization algorithm with $\alpha = 5$ and tolerance $\epsilon = 10^{-4}$ converged in all 100 runs to either of the two multiple global minimum solutions,

$$(41.17°, 59.84°) \quad \text{or} \quad (198.83°, 60.16°) \to V^*$$

$$= 0.769\ 101 \text{ kcal/mol}$$

in about 400 iterations with an average CPU time of 80 s on a HP-730 workstation. For $\alpha = 1$ convergence to the global minimum is attained in all 100 runs in about 110 iterations requiring only 20 s of CPU time.

B. 1,2,3-trichloro-1-fluoro-propane

In the second example the most stable conformation of 1,2,3-trichloro-1-fluoro-propane (CHClF–CHCl–CH$_2$Cl) (see Fig. 3) is sought. This molecule features eighteen different chains of four covalently bonded atoms and nine different chains of five covalently bonded atoms. Furthermore, all angles are assumed to be tetrahedral (109.5°), the bond lengths are given in Table IV,[5] and the parameters A_{ij} and B_{ij} for the Lennard-Jones potential,

$$v(r_{ij}) = \frac{B_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}}$$

which models the nonbonded interactions are given in Table 5.[5]

The selected pair of independent variables is,

$$t_1 \leftarrow \phi_{F(1)-C(4)-C(5)-Cl(6)},$$

$$t_2 \leftarrow \phi_{Cl(6)-C(5)-C(7)-Cl(9)},$$

and the bounds are $[0, 2\pi]$. The expressions for the rest of the dihedral angles are

$$\phi_{F(1)-C(4)-C(5)-Cl(6)} = t_1,$$

$$\phi_{F(1)-C(4)-C(5)-C(7)} = t_1 + \frac{2\pi}{3},$$

$$\phi_{F(1)-C(4)-C(5)-H(8)} = t_1 - \frac{2\pi}{3},$$

$$\phi_{Cl(2)-C(4)-C(5)-Cl(6)} = t_1 - \frac{2\pi}{3},$$

$$\phi_{Cl(2)-C(4)-C(5)-C(7)} = t_1,$$
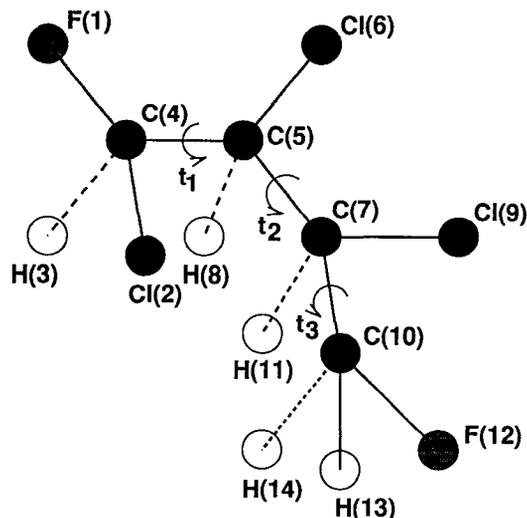
$$\phi_{Cl(2)-C(4)-C(5)-H(8)} = t_1 + \frac{2\pi}{3},$$

FIG. 4. 1,2,3-trichloro-1,4-bifluoro-butane molecule.

$$\phi_{H(3)-C(4)-C(5)-Cl(6)} = t_1 + \frac{2\pi}{3},$$

$$\phi_{H(3)-C(4)-C(5)-C(7)} = t_1 - \frac{2\pi}{3},$$

$$\phi_{H(3)-C(4)-C(5)-H(8)} = t_1,$$

$$\phi_{Cl(6)-C(5)-C(7)-Cl(9)} = t_2,$$

$$\phi_{Cl(6)-C(5)-C(7)-C(10)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{Cl(6)-C(5)-C(7)-H(11)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{C(4)-C(5)-C(7)-Cl(9)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{C(4)-C(5)-C(7)-C(10)} = t_2,$$

$$\phi_{C(4)-C(5)-C(7)-H(11)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(7)-Cl(9)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(7)-C(10)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(7)-H(11)} = t_2.$$

For the given set of parameter values and nonbonded interaction expressions the following solutions for the minimum potential energy problem have been found with the local nonlinear solver MINOS 5.3[108] by employing multiple initial points

TABLE V. Data for parameters A_{ij}, B_{ij} in the 1,2,3-trichloro-1-fluoro-propane example.

Atom-atom	A (kcal Å^6/mol)	B (10^{-4} kcal Å^{12}/mol)
F–Cl	457.6	29.6700
F–H	79.8	0.9361
F–C	223.2	10.0700
H–H	76.0	0.7220
H–C	127.4	3.7430
H–Cl	272.8	8.0740
Cl–Cl	1562.0	200.5000
Cl–C	759.3	64.6400

$t_1 = 291.50°$, $\quad t_2 = 188.75° \rightarrow V_{min} = -3.526\ 175$ kcal/mol

$t_1 = 55.97°$, $\quad t_2 = 67.22° \rightarrow V_{min} = -3.325\ 395$ kcal/mol

$t_1 = 193.52°$, $\quad t_2 = 187.88° \rightarrow V_{min} = -3.322\ 530$ kcal/mol

$t_1 = 291.30°$, $\quad t_2 = 282.73° \rightarrow V_{min} = -3.242\ 106$ kcal/mol

$t_1 = 193.65°$, $\quad t_2 = 282.65° \rightarrow V_{min} = -3.004\ 820$ kcal/mol

$t_1 = 288.39°$, $\quad t_2 = 66.48° \rightarrow V_{min} = -2.836\ 006$ kcal/mol

$t_1 = 56.60°$, $\quad t_2 = 191.77° \rightarrow V_{min} = -2.816\ 008$ kcal/mol

$t_1 = 193.50°$, $\quad t_2 = 66.91° \rightarrow V_{min} = -2.702\ 393$ kcal/mol

$t_1 = 55.91°$, $\quad t_2 = 282.40° \rightarrow V_{min} = -2.629\ 182$ kcal/mol

$t_1 = 55.97°$, $\quad t_2 = 67.22° \rightarrow V_{min} = -2.325\ 395$ kcal/mol.

After applying the presented global optimization algorithm with $\alpha = 10$ and tolerance $\epsilon = 10^{-4}$ the global minimum solution,

$t_1^* = 291.50°$, $\quad t_2^* = 188.75° \rightarrow V^* = -3.526\ 175$ kcal/mol

is obtained in all 100 runs in about 200 iterations with an average CPU time of 70 s on a HP-730 workstation. After halving the value of parameter α to $\alpha = 5.0$ the global minimum solution for all 100 runs is still found requiring about 110 iterations and 40 s of CPU time.

C. 1,2,3-trichloro-1,4-bifluoro-butane

This third example addresses the problem of finding the global minimum total potential energy conformation of 1,2,3-trichloro-1,4-bifluoro-butane (CHClF–CHCl–CHCl–CH$_2$F), (see Fig. 4). This molecule involves 27 different chains composed of four covalently bonded atoms; 18 chains of five covalently bonded atoms; and 9 chains of six covalently bonded atoms. In this example all covalent bond angles are assumed to be tetrahedral (109.5°). The equilibrium bond lengths as well as the parameters A_{ij} and B_{ij} for the Lennard-Jones potential are listed in Table IV and Table V of the previous example.

Clearly, only three dihedral angles are required to fully describe the rotational conformation of the molecule, and one possible selection is the following:

$0 \leqslant t_1 \leftarrow \phi_{F(1)-C(4)-C(5)-Cl(6)} \leqslant 2\pi$,

$0 \leqslant t_2 \leftarrow \phi_{Cl(6)-C(5)-C(7)-Cl(9)} \leqslant 2\pi$,

$0 \leqslant t_3 \leftarrow \phi_{Cl(9)-C(7)-C(10)-F(12)} \leqslant 2\pi$.

The expressions for all 27 dihedral angles are then,

$\phi_{F(1)-C(4)-C(5)-Cl(6)} = t_1$,

$\phi_{F(1)-C(4)-C(5)-C(7)} = t_1 + \dfrac{2\pi}{3}$,

$\phi_{F(1)-C(4)-C(5)-H(8)} = t_1 - \dfrac{2\pi}{3}$,

$\phi_{Cl(2)-C(4)-C(5)-Cl(6)} = t_1 - \dfrac{2\pi}{3}$,

$\phi_{Cl(2)-C(4)-C(5)-C(7)} = t_1$,

$\phi_{Cl(2)-C(4)-C(5)-H(8)} = t_1 + \dfrac{2\pi}{3}$,

$\phi_{H(3)-C(4)-C(5)-Cl(6)} = t_1 + \dfrac{2\pi}{3}$,

$\phi_{H(3)-C(4)-C(5)-C(7)} = t_1 - \dfrac{2\pi}{3}$,

$\phi_{H(3)-C(4)-C(5)-H(8)} = t_1$,

$\phi_{Cl(6)-C(5)-C(7)-Cl(9)} = t_2$,

$\phi_{Cl(6)-C(5)-C(7)-C(10)} = t_2 + \dfrac{2\pi}{3}$,

$\phi_{Cl(6)-C(5)-C(7)-H(11)} = t_2 - \dfrac{2\pi}{3}$,

$\phi_{C(4)-C(5)-C(7)-Cl(9)} = t_2 - \dfrac{2\pi}{3}$,

$\phi_{C(4)-C(5)-C(7)-C(10)} = t_2$,

TABLE VI. Solutions for the 1,2,3-trichloro-1,4-bifluoro-butane example.

t_1^0	t_2^0	t_3^0	V_{min}^{local} kcal/mol
291.95	180.42	182.86	−5.271 002
291.95	180.49	290.97	−5.159 241
291.94	180.06	64.59	−5.127 455
194.32	179.85	182.85	−4.975 758
291.08	292.43	184.17	−4.944 256
194.33	179.91	290.95	−4.865 099
291.07	292.13	65.12	−4.861 028
289.02	67.06	183.21	−4.841 873
194.30	179.49	64.62	−4.830 295
289.03	67.09	290.93	−4.697 346
289.04	67.06	65.54	−4.691 369
194.30	67.43	183.17	−4.603 630
194.61	292.47	184.16	−4.565 333
194.60	292.17	65.20	−4.485 494
194.60	292.21	291.12	−4.473 805
194.29	67.46	290.93	−4.461 059
194.29	67.44	65.53	−4.452 745
56.17	181.51	182.86	−4.451 375
56.16	181.58	290.87	−4.334 131
55.61	68.03	183.30	−4.296 107
56.18	181.14	64.58	−4.290 602
55.14	292.50	184.16	−4.192 450
55.60	68.02	290.99	−4.160 552
55.61	67.98	65.47	−4.153 306

$$\phi_{C(4)-C(5)-C(7)-H(11)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(7)-Cl(9)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(7)-C(10)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(7)-H(11)} = t_2,$$

$$\phi_{Cl(9)-C(7)-C(10)-F(12)} = t_3,$$

$$\phi_{Cl(9)-C(7)-C(10)-H(13)} = t_3 + \frac{2\pi}{3},$$

$$\phi_{Cl(9)-C(7)-C(10)-H(14)} = t_3 - \frac{2\pi}{3},$$

$$\phi_{C(5)-C(7)-C(10)-F(12)} = t_3 - \frac{2\pi}{3},$$

$$\phi_{C(5)-C(7)-C(10)-H(13)} = t_3,$$

$$\phi_{C(5)-C(7)-C(10)-H(14)} = t_3 + \frac{2\pi}{3},$$

$$\phi_{H(11)-C(7)-C(10)-F(12)} = t_3 + \frac{2\pi}{3},$$

$$\phi_{H(11)-C(7)-C(10)-H(13)} = t_3 - \frac{2\pi}{3},$$

$$\phi_{H(11)-C(7)-C(10)-H(14)} = t_3.$$

For the current set of data and interaction type the local optimization algorithm MINOS 5.3[108] generates a number of solutions which are summarized in Table VI. It is interesting to note that only 8 out of 100 runs resulted in generating the global minimum solution.

However, the proposed global optimization algorithm for $\alpha = 5.0$ and $\epsilon = 10^{-4}$ in all 100 runs converged to the global minimum,

$$t_1^* = 291.95°, \quad t_2^* = 180.42°,$$

$$t_3^* = 182.86° \rightarrow V^* = -5.271\,002 \text{ kcal/mol}$$

in about 1000 iterations with an average CPU time of 1400 s on a HP-730 workstation.

D. 1,3-bifluoro-1-chloro-2-chloromethyl-propane

The fourth example involves finding the most stable molecular conformation of 1,3-bifluoro-1-chloro-2-chloromethyl-propane [CHFCl–CH(CHCl)–CH$_2$F], (see Fig. 5). This molecule involves 27 different chains composed of four covalently bonded atoms; and also 27 chains of five covalently bonded atoms. Note again that all covalent bond angles are assumed to be tetrahedral (109.5°), and that the equilibrium bond lengths as well as the parameters A_{ij} and B_{ij} for the Lennard-Jones potential are listed in Table IV and Table V of the second example.
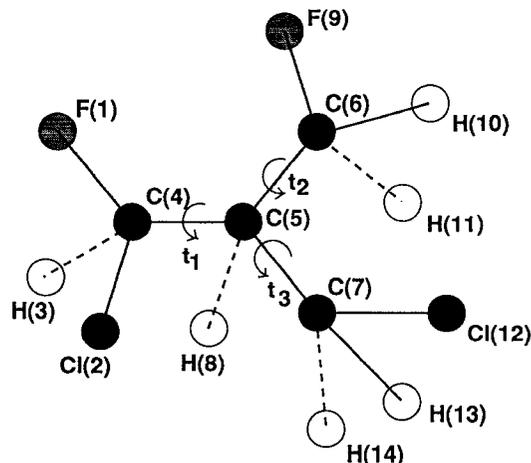
FIG. 5. 1,3-bifluoro-1-chloro-2-chloromethyl-propane.

Three dihedral angles are required in this example to satisfy all the rotational degrees of freedom of the molecule, and one possible selection is the following:

$$0 \leqslant t_1 \leftarrow \phi_{F(1)-C(4)-C(5)-C(6)} \leqslant 2\pi,$$

$$0 \leqslant t_2 \leftarrow \phi_{C(4)-C(5)-C(6)-F(9)} \leqslant 2\pi,$$

$$0 \leqslant t_3 \leftarrow \phi_{C(4)-C(5)-C(7)-Cl(12)} \leqslant 2\pi.$$

Explicit expressions for all 27 dihedral angles as functions of the above four are

$$\phi_{F(1)-C(4)-C(5)-C(6)} = t_1,$$

$$\phi_{F(1)-C(4)-C(5)-C(7)} = t_1 + \frac{2\pi}{3},$$

TABLE VII. Solutions for the 1,3-bifluoro-1-chloro-2-chloromethyl-propane.

t_1^0	t_2^0	t_3^0	V_{min}^{local} kcal/mol
293.41	296.58	189.20	−5.308 668
185.82	296.87	291.11	−5.294 174
293.40	183.59	189.20	−5.291 917
185.80	183.61	291.12	−5.291 700
293.41	59.65	189.20	−5.289 884
186.45	60.50	291.11	−5.273 069
291.30	296.59	289.05	−5.271 459
185.94	296.87	189.07	−5.255 464
291.28	183.59	289.05	−5.254 775
185.92	183.61	189.07	−5.252 985
291.29	59.65	289.05	−5.252 699
186.55	60.50	189.07	−5.234 493
60.86	59.79	291.93	−5.223 234
60.81	183.47	291.93	−5.220 839
60.42	295.69	291.93	−5.211 549
60.78	59.79	188.81	−5.198 598
60.74	183.47	188.81	−5.196 214
60.34	295.69	188.81	−5.187 023
185.74	296.87	59.41	−5.173 325
293.77	296.57	59.92	−5.172 821
185.72	183.61	59.41	−5.170 854
293.77	183.59	59.92	−5.156 062
293.78	59.65	59.92	−5.154 038
186.35	60.50	59.41	−5.152 115
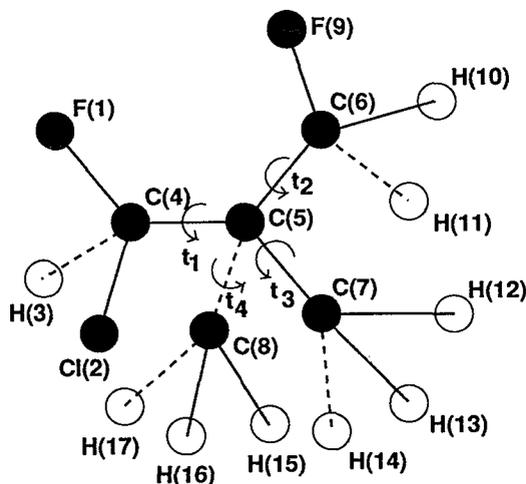
FIG. 6. 1,3-bifluoro-1-chloro-2,2-bimethyl-propane.

$$\phi_{F(1)-C(4)-C(5)-H(8)} = t_1 - \frac{2\pi}{3},$$

$$\phi_{Cl(2)-C(4)-C(5)-C(6)} = t_1 - \frac{2\pi}{3},$$

$$\phi_{Cl(2)-C(4)-C(5)-C(7)} = t_1,$$

$$\phi_{Cl(2)-C(4)-C(5)-H(8)} = t_1 + \frac{2\pi}{3},$$

$$\phi_{H(3)-C(4)-C(5)-C(6)} = t_1 + \frac{2\pi}{3},$$

$$\phi_{H(3)-C(4)-C(5)-C(7)} = t_1 - \frac{2\pi}{3},$$

$$\phi_{H(3)-C(4)-C(5)-H(8)} = t_1,$$

$$\phi_{C(4)-C(5)-C(6)-F(9)} = t_2,$$

$$\phi_{C(4)-C(5)-C(6)-H(10)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{C(4)-C(5)-C(6)-H(11)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{C(7)-C(5)-C(6)-F(9)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{C(7)-C(5)-C(6)-H(10)} = t_2,$$

$$\phi_{C(7)-C(5)-C(6)-H(11)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(6)-F(9)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(6)-H(10)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(6)-H(11)} = t_2,$$

$$\phi_{C(4)-C(5)-C(7)-Cl(12)} = t_3,$$

$$\phi_{C(4)-C(5)-C(7)-H(13)} = t_3 + \frac{2\pi}{3},$$

$$\phi_{C(4)-C(5)-C(7)-H(14)} = t_3 - \frac{2\pi}{3},$$

$$\phi_{C(6)-C(5)-C(7)-Cl(12)} = t_3 - \frac{2\pi}{3},$$

$$\phi_{C(6)-C(5)-C(7)-H(13)} = t_3,$$

$$\phi_{C(6)-C(5)-C(7)-H(14)} = t_3 + \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(7)-Cl(12)} = t_3 + \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(7)-H(13)} = t_3 - \frac{2\pi}{3},$$

$$\phi_{H(8)-C(5)-C(7)-H(14)} = t_3.$$

For this example, the local optimization algorithm MINOS 5.3[108] generates a plethora of very closely spaced solutions the best of which are tabulated in Table VII. It is interesting to note that only 39 out of 1000 runs resulted in generating the global minimum solution. This is less than 4% success rate.

However, the proposed global optimization algorithm for $\alpha = 1$ and $\epsilon = 10^{-3}$ in all runs converged to the global minimum,

$$t_1^* = 293.41°, \quad t_2^* = 296.58°,$$

$$t_3^* = 189.20° \rightarrow V^* = -5.308\ 668 \ \text{kcal/mol}$$

in about 100 iterations with an average CPU time of 300 s on a HP-730 workstation. For $\alpha = 5$ again convergence to the global minimum is achieved in about 400 iterations 1200 s of CPU time.

E. 1,3-bifluoro-1-chloro-2,2-bimethyl-propane

The fourth example involves finding the most stable molecular conformation of 1,3-bifluoro-1-chloro-2,2-bichloromethyl-propane [CHFCl–C(CH$_2$Cl)$_2$–CH$_2$F], (see Fig. 6). This molecule involves 36 different chains composed of four covalently bonded atoms; and 54 chains of five covalently bonded atoms. Note again that all covalent bond angles are assumed to be tetrahedral (109.5°), and that the equilibrium bond lengths as well as the parameters A_{ij} and B_{ij} for the Lennard-Jones potential are listed in Table IV and Table V of the second example.

Four dihedral angles are required in this example to satisfy all the rotational degrees of freedom of the molecule, and one possible selection is the following:

$$0 \leqslant t_1 \leftarrow \phi_{F(1)-C(4)-C(5)-C(6)} \leqslant 2\pi,$$

$$0 \leqslant t_2 \leftarrow \phi_{C(4)-C(5)-C(6)-F(9)} \leqslant 2\pi,$$

$$0 \leqslant t_3 \leftarrow \phi_{C(4)-C(5)-C(7)-H(12)} \leqslant \frac{2\pi}{3},$$

TABLE VIII. Solutions for the 1,3-bifluoro-1-chloro-2,2-bimethyl-propane.

β_1	β_2	β_3	β_4	V_{min}^{local} kcal/mol
299.92	59.59	59.87	60.18	−7.121,635
299.77	300.59	59.88	60.18	−7.102 329
299.68	179.37	59.88	60.18	−7.037 667
58.75	59.94	60.19	59.95	−6.933 911
181.40	300.45	59.95	59.86	−6.925 400
181.70	60.53	59.95	59.86	−6.922 366
181.34	179.89	59.95	59.86	−6.921 186
58.85	179.41	60.18	59.95	−6.893 233
58.56	300.23	60.19	59.95	−6.835 546
58.75	59.94	60.19	120.00	−6.104 734
58.74	59.94	0.00	59.95	−6.093 669
181.35	179.89	59.95	120.00	−6.089 325
181.34	179.89	120.00	59.86	−6.082 174
299.86	360.00	59.87	60.18	−5.880 910
181.68	0.00	59.95	59.86	−5.674 856
58.51	360.00	60.19 `	59.95	−5.645 061

$$0 < t_4 \leftarrow \phi_{C(4)-C(5)-C(8)-H(15)} \leq \frac{2\pi}{3}.$$

Note that due to symmetry the last two dihedral angles suffice to vary only between 0 and $2\pi/3$. Explicit expressions for all 36 dihedral angles as functions of the above four are

$$\phi_{F(1)-C(4)-C(5)-C(6)} = t_1,$$

$$\phi_{F(1)-C(4)-C(5)-C(7)} = t_1 + \frac{2\pi}{3},$$

$$\phi_{F(1)-C(4)-C(5)-C(8)} = t_1 - \frac{2\pi}{3},$$

$$\phi_{Cl(2)-C(4)-C(5)-C(6)} = t_1 - \frac{2\pi}{3},$$

$$\phi_{Cl(2)-C(4)-C(5)-C(7)} = t_1,$$

$$\phi_{Cl(2)-C(4)-C(5)-C(8)} = t_1 + \frac{2\pi}{3},$$

$$\phi_{H(3)-C(4)-C(5)-C(6)} = t_1 + \frac{2\pi}{3},$$

$$\phi_{H(3)-C(4)-C(5)-C(7)} = t_1 - \frac{2\pi}{3},$$

$$\phi_{H(3)-C(4)-C(5)-C(8)} = t_1,$$

$$\phi_{C(4)-C(5)-C(6)-F(9)} = t_2,$$

$$\phi_{C(4)-C(5)-C(6)-H(10)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{C(4)-C(5)-C(6)-H(11)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{C(7)-C(5)-C(6)-F(9)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{C(7)-C(5)-C(6)-H(10)} = t_2,$$

$$\phi_{C(7)-C(5)-C(6)-H(11)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{C(8)-C(5)-C(6)-F(9)} = t_2 + \frac{2\pi}{3},$$

$$\phi_{C(8)-C(5)-C(6)-H(10)} = t_2 - \frac{2\pi}{3},$$

$$\phi_{C(8)-C(5)-C(6)-H(11)} = t_2,$$

$$\phi_{C(4)-C(5)-C(7)-H(12)} = t_3,$$

$$\phi_{C(4)-C(5)-C(7)-H(13)} = t_3 + \frac{2\pi}{3},$$

$$\phi_{C(4)-C(5)-C(7)-H(14)} = t_3 - \frac{2\pi}{3},$$

$$\phi_{C(6)-C(5)-C(7)-H(12)} = t_3 - \frac{2\pi}{3},$$

$$\phi_{C(6)-C(5)-C(7)-H(13)} = t_3,$$

$$\phi_{C(6)-C(5)-C(7)-H(14)} = t_3 + \frac{2\pi}{3},$$

$$\phi_{C(8)-C(5)-C(7)-H(12)} = t_3 + \frac{2\pi}{3},$$

$$\phi_{C(8)-C(5)-C(7)-H(13)} = t_3 - \frac{2\pi}{3},$$

$$\phi_{C(8)-C(5)-C(7)-H(14)} = t_3,$$

$$\phi_{C(4)-C(5)-C(8)-H(15)} = t_4,$$

$$\phi_{C(4)-C(5)-C(8)-H(16)} = t_4 + \frac{2\pi}{3},$$

$$\phi_{C(4)-C(5)-C(8)-H(17)} = t_4 - \frac{2\pi}{3},$$

$$\phi_{C(6)-C(5)-C(8)-H(15)} = t_4 - \frac{2\pi}{3},$$

$$\phi_{C(6)-C(5)-C(8)-H(16)} = t_4,$$

$$\phi_{C(6)-C(5)-C(8)-H(17)} = t_4 + \frac{2\pi}{3},$$

$$\phi_{C(7)-C(5)-C(8)-H(15)} = t_4 + \frac{2\pi}{3},$$

$$\phi_{C(7)-C(5)-C(8)-H(16)} = t_4 - \frac{2\pi}{3},$$

$$\phi_{C(7)-C(5)-C(8)-H(17)} = t_4.$$

The solutions generated by MINOS 5.3[108] are summarized in Table VIII. It is interesting to note that only 10% of the randomly distributed initial points generated paths leading to the global minimum solution.

However, the proposed global optimization algorithm with $\alpha = 5$ converged to the global minimum solution

$$t_1^* = 299.92°, \quad t_2^* = 59.59°, \quad t_3^* = 59.87°,$$

$$t_4^* = 60.18° \rightarrow V^* = -7.121\,635 \text{ kcal/mol}$$

in all runs requiring about 1000 iterations with an average CPU time of 4000 s.

VII. SUMMARY AND CONCLUSIONS

In this paper a deterministic branch and bound type global optimization algorithm was presented for finding the most stable conformations of molecules based solely on the energetics of nonbonded interactions. First, explicit relations for the nonbonded atom Euclidean distances as function of bond lengths, covalent bond angles, and dihedral (torsion) angles were derived. Then, the problem was formulated as an unconstrained nonconvex optimization problem on a set of independent dihedral angles. Based on an eigenvalue analysis, a convex lower bounding function $\mathscr{L}$ of the total potential energy function V was defined involving a number of important properties. These properties of function $\mathscr{L}$ coupled with an efficient partitioning scheme enabled us to construct a global optimization algorithm guaranteed to always ϵ converge to the global minimum conformation through the solution of a series of convex nonlinear optimization problems. Finally, the proposed approach was applied successfully to a number of acyclic molecular conformation problems, six of which were included in this paper. It should be emphasized that the proposed global optimization algorithm is applicable to other classes of problems (e.g., microclusters). Also, the presented analysis can be extended to more complicated interaction fields involving complex potential interaction terms (e.g., bending, stretching, torsion, bending–stretching, and out-of-plane distortion). Work in this direction is currently in progress.

ACKNOWLEDGMENTS

[1] L. T. Wille and J. Vennik, J. Phys. A **18**, L419 (1985).
[2] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, San Francisco, 1979).
[3] J. M. Troyer and F. E. Cohen, *Simplified Models for Understanding and Predicting Protein Structure* (VCH, New York, 1991).
[4] M. Levitt, Annu. Rev. Biophys. Bioeng. **11**, 251 (1982).
[5] A. J. Hopfinger, *Conformational Properties of Macromolecules* (Academic, New York, 1973).
[6] F. A. Momany, L. M. Carruthers, R. F. McGuire, and H. A. Scheraga, J. Phys. Chem. **78**, 1595 (1974a).
[7] F. A. Momany, L. M. Carruthers, and H. A. Scheraga, J. Phys. Chem. **78**, 1621 (1974b).
[8] F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, J. Phys. Chem. **79**, 2361 (1975).
[9] N. L. Allinger, J. Am. Chem. Soc. **99**, 8127 (1977).
[10] G. Némethy, M. S. Pottle, and H. A. Scheraga, J. Phys. Chem. **89**, 1883 (1983).
[11] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **8**, 132 (1983).
[12] S. Weiner, P. Kollmann, D. Nguyen, and D. Case, J. Comput. Chem. **7**, 230 (1986).
[13] W. F. van Gunsteren and H. J. C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual* (Biomos, Groningen, The Netherlands, 1987).
[14] N. L. Allinger, Y. H. Yuh, and J.-H. Lii, J. Am. Chem. Soc. **111**, 8551 (1089).
[15] G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga, J. Phys. Chem. **96**, 6472 (1992).
[16] S. W. Fesik, T. J. O'Donnell, R. T. Gampe, and E. T. Olejniczak, J. Am. Chem. Soc. **83**, 3801 (1986).
[17] A. T. Brünger, G. M. Clore, A. M. Gronenborn, and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **83**, 3801 (1986).
[18] S. W. Fesik, G. Bolis, H. L. Sham, and E. T. Olejniczak, Biochemistry **26**, 1851 (1987).
[19] H. Kessler, C. Griesinger, J. Lautz, A. Müller, W. F. van Gunsteren, and H. J. C. Berendsen, J. Am. Chem. Soc. **110**, 3393 (1988).
[20] L. Nilsson, G. M. Clore, A. M. Gronenborn, A. T. Brünger, and M. Karplus, J. Mol. Biol. **188**, 455 (1986).
[21] R. Elber and M. Karplus, J. Am. Chem. Soc. **112**, 9161 (1990).
[22] C. L. Brooks III, M. Karplus, and B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, Adv. Chem. Phys. LXXI (Wiley, New York, 1988).
[23] J. A. McCammon and M. Karplus, Acc. Chem. Res. **16**, 187 (1983).
[24] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[25] E. R. P. Zuiderweg, R. M. Scheek, R. Boelens, W. F. van Gunsteren, and R. Kaptein, Biochimie **67**, 707 (1985).
[26] J. Lautz, H. Kessler, R. Kaptein, and W. F. van Gunsteren, J. Comp.-Aided Mol. Design **1**, 219 (1987).
[27] R. C. van Schaik, W. F. van Gunsteren, and H. J. C. Berendsen, J. Comp.-Aided Mol. Design **6**, 97 (1992).
[28] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[29] Z. Li and H. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **84**, 6611 (1987).
[30] Z. Li and H. Scheraga, J. Phys. Chem. **92**, 2633 (1988).
[31] D. R. Ripoll and H. A. Scheraga, Biopolymers **27**, 1283 (1988).
[32] D. R. Ripoll and H. A. Scheraga, J. Protein Chem. **8**, 263 (1989).
[33] D. R. Ripoll and H. A. Scheraga, Biopolymers **30**, 165 (1990).
[34] D. R. Ripoll, M. Vasquez, and H. A. Scheraga, Biopolymers **31**, 319 (1991).
[35] G. H. Paine and H. A. Scheraga, Biopolymers **24**, 1391 (1985).
[36] G. H. Paine and H. A. Scheraga, Biopolymers **25**, 1547 (1986).
[37] G. H. Paine and H. A. Scheraga, Biopolymers **26**, 1125 (1987).
[38] G. Chang, W. C. Guida, and W. C. Still, J. Am. Chem. Soc. **11**, 4379 (1989).
[39] D. M. Jones and J. M. Goodfellow, J. Comput. Chem. **14**, 127 (1993).
[40] S. Kirkpatrick, C. D. Gelatt, Jr., M. P. Vecchi, Science **220**, 671 (1983).
[41] S. Wilson and W. Cui, Tetrahedron Lett. **29**, 4373 (1988).
[42] S. Wilson and W. Cui, Biopolymers **29**, 225 (1990).
[43] R. A. Donnely, Chem. Phys. Lett. **136**, 274 (1987).
[44] R. G. Juárez, L. B. Morales, and F. P. Neri, J. Mol. Struct. **208**, 279 (1990).
[45] B. M. Pettitt, T. Matsunaga, F. Al-Obeidi, C. Gehrig, V. J. Hruby, and M. Karplus, Biophys. J. **60**, 1540 (1991).
[46] R. E. Bruccoleri and M. Karplus, Biopolymers **26**, 137 (1987).
[47] R. E. Bruccoleri and M. Karplus, Biopolymers **29**, 1847 (1990).
[48] M. Saunders, J. Comput. Chem. **10**, 203 (1989).
[49] M. Saunders and N. Krause, J. Am. Chem. Soc. **112**, 1791 (1990).
[50] M. Saunders, J. Comput. Chem. **12**, 645 (1991).
[51] B. Robson and J. Garnier, *Introduction to Proteins and Protein Engineering* (Elsevier, New York, 1986).
[52] E. O. Purisima and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **83**, 2782 (1986).
[53] E. O. Purisima and H. A. Scheraga, J. Mol. Biol. **196**, 697 (1987).
[54] G. M. Crippen, J. Comput. Chem. **3**, 471 (1982).
[55] G. M. Crippen, J. Comput. Chem. **5**, 548 (1984).
[56] G. M. Crippen, J. Phys. Chem. **91**, 6341 (1987).
[57] G. M. Crippen and T. F. Havel, *Distance Geometry and Molecular*

Conformation, Chemometrics Research Studies Series (Research Studies, Wiley, New York, 1988).

[58] G. M. Crippen and T. F. Havel, J. Chem. Inf. Comput. Sci. **30**, 222 (1990).

[59] G. M. Crippen, *Distance Geometry and Conformational Calculations* (Research Studies, Wiley, New York, 1981).

[60] I. D. Kuntz, Protein Eng. **1**, 147 (1987).

[61] J. Lautz, H. Kessler, J. M. Blaney, R. M. Scheek, and W. F. van Gunsteren, Int. J. Peptide Protein Res. **33**, 281 (1989).

[62] P. L. Weber, R. Morrison, and D. Hare, J. Mol. Biol. **203**, 483 (1988).

[63] T. F. Havel, Prog. Biophys. Mol. Biol. **56**, 43 (1991).

[64] M. Billeter, T. F. Havel, and K. J. Wüthrich, J. Comput. Chem. **8**, 132 (1987).

[65] M. Billeter, A. E. Howard, I. D. Kuntz, and P. A. Kollman, J. Am. Chem. Soc. **110**, 8385 (1988).

[66] B. van der Graaf and J. M. A. Baas, Rev. Trav. Chim. Pays-Bas **99**, 327 (1980).

[67] M. Levitt, J. Mol. Biol. **170**, 723 (1983).

[68] J. Moult and M. N. G. James, Proteins: Struct., Funct., Genet. **1**, 146 (1986).

[69] B. C. Wilkes and P. W. Schiller, Biopolymers **26**, 1431 (1987).

[70] M. Lipton and W. C. Still, J. Comput. Chem. **9**, 343 (1988).

[71] T. Ryhänen, F. J. Bermejo, J. Santoro, M. Rico, Comput. Chem. **11**, 13 (1987).

[72] D. N. J. White and D. H. Kitson, J. Mol. Graphics **4**, 112 (1986).

[73] D. P. Dolata, A. R. Leach, K. Prout, J. Comput.-Aided Mol. Design **1**, 73 (1987).

[74] K. D. Gibson and H. A. Scheraga, J. Comput. Chem. **8**, 826 (1987).

[75] M. Vásquez and H. A. Scheraga, Biopolymers **24**, 1437 (1985).

[76] M. Vásquez and H. A. Scheraga, J. Biomol. Stuct. Dyn. **5**, 705 (1988a).

[77] M. Vásquez and H. A. Scheraga, J. Biomol. Stuct. Dyn. **5**, 757 (1988b).

[78] D. M. Ferguson and D. J. Raber, J. Am. Chem. Soc. **111**, 4371 (1989).

[79] D. M. Ferguson, W. A. Glauser, and D. J. Raber, J. Comput. Chem. **10**, 903 (1989).

[80] D. M. Ferguson and D. J. Raber, J. Comput. Chem. **11**, 1061 (1990).

[81] L. Piela, H. A. Scheraga, Biopolymers **26**, S33 (1987).

[82] L. Holley and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **86**, 152 (1989).

[83] D. Kneller, F. Cohen, and R. Langridge, J. Mol. Biol. **214**, 171 (1990).

[84] S. Le Grand and K. M. Merz Jr., J. Global Opt. **3**, 49 (1993).

[85] S. Vajda and C. Delisi, Biopolymers **29**, 1755 (1990).

[86] M. Lambert and H. Scheraga, J. Comput. Chem. **10**, 770 (1989a).

[87] M. Lambert and H. Scheraga, J. Comput. Chem. **10**, 798 (1989b).

[88] M. Lambert and H. Scheraga, J. Comput. Chem. **10**, 817 (1989c).

[89] L. Piela, J. Kostrowicki, and H. A. Scheraga, J. Phys. Chem. **93**, 3339 (1989).

[90] J. Kostrowicki, L. Piela, B. J. Cherayil, and H. A. Scheraga, J. Phys. Chem. **95**, 4113 (1991).

[91] J. Kostrowicki and H. A. Scheraga, J. Phys. Chem. **96**, 7442 (1992).

[92] K. R. Acharya, D. I. Stuart, D. C. Phillips, and H. A. Scheraga, J. Protein Chem. **9**, 549 (1990).

[93] M. J. Dudek and H. A. Scheraga, J. Comput. Chem. **11**, 121 (1990).

[94] R. L. Somorjai, J. Phys. Chem. **95**, 4141 (1991).

[95] M. Sylvain and R. L. Somorjai, J. Phys. Chem. **95**, 4147 (1991).

[96] K. A. Olszewski, L. Piela, and H. A. Scheraga, J. Phys. Chem. **96**, 4672 (1992).

[97] U. Burkert and N. L. Allinger, *Molecular Mechanics* (American Chemical Society, Washington, D.C., 1982).

[98] C. D. Maranas and C. A. Floudas, J. Chem. Phys. **97**, 10 (1992).

[99] C. D. Maranas and C. A. Floudas, Ann. Operation Res. **42**, 85, (1993).

[100] C. D. Maranas and C. A. Floudas, J. Global Opt. **4**, 3 (1994).

[101] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms* (McGraw-Hill, New York, 1989).

[102] H. B. Thompson, J. Chem. Phys. **47**, 3407 (1969).

[103] C. A. Floudas and V. Visweswaran, Comput. Chem. Eng. **14**, 1397 (1990).

[104] V. Visweswaran and C. A. Floudas, Comput. Chem. Eng. **14**, 1419 (1990).

[105] V. Visweswaran and C. A. Floudas, Proceedings of Process Systems Engineering, PSE '91, I.6.1 (1991).

[106] C. A. Floudas and V. Visweswaran, J. Opt. Theory Appl. **78**, Iss. 2 (1993).

[107] V. Visweswaran and C. A. Floudas, J. Global Opt. **3**, Iss. 4 (1993).

[108] B. A. Murtagh and M. A. Saunders, *MINOS 5.3 User's Guide* (Systems Optimization Laboratory, Department of Operations Research, Stanford University, 1987).

[109] R. Susnow, R. B. Nachbar, C. Schutt, and H. Rabitz, J. Phys. Chem. **95**, 8585 (1991).