A New Method for Fast and Accurate Derivation of Molecular Conformations

Ming Zhang[1] Lydia E. Kavraki[1,2,*]
{mzhang, kavraki}@cs.rice.edu
[1]Department of Computer Science
[2]Department of Bioengineering
Rice University, Houston TX 77005 U.S.A.
[*]Corresponding Author: tel 713-348-5737 kavraki@cs.rice.edu

Abstract

During molecular simulations, three dimensional conformations of biomolecules are calculated from the values of their bond angles, bond lengths and torsional angles. In this paper we study how to efficiently derive three dimensional molecular conformations from the values of torsional angles. This case is of broad interest as torsional angles greatly affect molecular shape and are always taken into account during simulations. We first review two widely-used methods for deriving molecular conformations, the simple rotations scheme and the Denavit-Hartenberg local frames method. We discuss their disadvantages which include extensive bookkeeping, accumulation of numerical errors and redundancies in the local frames used. Then we introduce a new, fast and accurate method called the atomgroup local frames method. This new method not only eliminates the disadvantages of earlier approaches, but also provides lazy evaluation of atom positions and reduces the computational cost. Our method is especially useful in applications where many conformations are generated or updated such as in energy minimization and conformational search.

1 Introduction

During the last two decades a lot of emphasis has been placed in studying problems that involve reasoning about three dimensional (3D) molecular conformations in simulation (e.g., docking, protein folding, molecular dynamics, etc.) [3, 4, 6, 7, 8, 9, 10, 11, 12]. Little research has been done, however, to understand how to efficiently represent, derive and update molecular conformations from molecular data. The efficient derivation of molecular conformations can greatly impact the performance of conformational search procedures, energy minimization procedures, and all computations that involve large molecules and require frequent recalculation of conformations. This paper is a systematic investigation of the efficiency of different representations for the derivation of molecular conformations. Two available methods, the simple rotations scheme and the Denavit-Hartenberg local frames method, are reviewed and a new improved method, the atomgroup local frames method, is introduced in this paper.

For computational purposes, a molecule is often represented as a collection of atoms and a collection of bonds between pairs of atoms. Additional information is associated with each of the atoms and bonds, such as van der Waals radii, electric charges, bond lengths, bond angles, torsional angles and others [4, 7]. In most kinematics studies, bond lengths and bond angles are considered constant, while the torsional angles are allowed to change [9, 11] (see also Figure 1). That is why in this paper we assume that conformational changes are solely due to the changes of torsional angles and investigate how to efficiently derive and update conformations under this assumption.

A direct application of our work is in improving the running time of procedures that generate and/or update a large number of conformations. Take conformational search procedures as an example [6, 8, 9, 10, 11]. The goal of conformational search is to find a set of low energy conformations of a molecule. External constraints (such as specific distances between two atoms in the molecule) may also be required [4]. Unfortunately, the conformational spaces of even small molecules are high-dimensional (10–15 torsional angles). As a result, conformational search is a computationally expensive procedure and many researchers have developed a variety of algorithms to perform this procedure as efficiently as possible [6, 8, 9, 10, 11]. The common theme of these algorithms is to reduce the size of the search space and try to find, in the reduced space, as many solution candidates as possible. Still, a large number of conformations are checked and analyzed. It is therefore desirable to have good computer representations for molecules and good algorithms for finding the Cartesian positions of their atoms so as to minimize the cost of the evaluation of each conformation and speed up the entire conformational search procedure. Other applications that can benefit from our work in a similar way are ligand-receptor docking, protein folding, molecular dynamics and many more.

Available methods for deriving molecular conformations from the values of torsional angles include the simple rotations scheme and the Denavit-Hartenberg local frames method [9]. The simple rotations scheme applies a sequence of rotations to update all atom positions, where each rotation is determined by two points and an angle. The order of the updates of the atoms is important – an atom can only be updated after all its ancestor atoms have been updated. Thus some bookkeeping of the atom positions is necessary. The Denavit-Hartenberg local frames method builds local frames at the bonds [2, 9]. The relations between local frames at parents and children are updated when a rotation is applied. The atom positions are computed by series of matrix multiplications. An advantage of the Denavit-Hartenberg local frames method is that no bookkeeping is needed. The drawback of this method is that multiple local frames are needed at a bond if it has more than one child.

This paper introduces a new atomgroup local frames method to efficiently derive molecular conformations from the values of torsional angles. In this method, a single local frame is attached to each rotatable bond, and the position of each atom is updated by a single matrix multiplication. This method provides accumulated, lazy evaluations for atom positions. Thus the computational cost can be greatly reduced, especially when many conformations are generated and updated (e.g., during a minimization procedure). Another advantage of the new method can be seen when selective atom positions need to be calculated to check, for example, if they satisfy distance constraints to fit a pharmacophoric pattern [4, 9]. Our approach easily permits the evaluation of the position of any atom in the molecule without computing the positions of the rest of the atoms.

The rest of this paper is organized in the following way. Section 2 describes the simple rotations scheme for calculating the positions of atoms in a molecule. Section 3 discusses the Denavit-Hartenberg local frames method. In Section 4, we introduce the atomgroup local frames method. We give some comparison results between the three methods and some discussion in Section 5 and summarize our results in Section 6.

2 Simple Rotations

A molecule is characterized by a collection of atoms and a collection of bonds between pairs of atoms [9]. A graph can then represent the molecule with vertices denoting atoms and edges denoting bonds. An atom is chosen as the anchor for the molecule — root of the corresponding graph. (The choice of the anchor atom is arbitrary.) In this section, we assume that the underlying graphs are trees— that is, there are no rings in the molecules. Another way to view this is that rings are considered rigid and have been replaced by special single "atoms". Therefore, for any atom in the graph, there is only one unique way to traverse back to the root (anchor) atom. Thus, the parent-children relations among the atoms are unambiguous. Figure 1 gives an example.

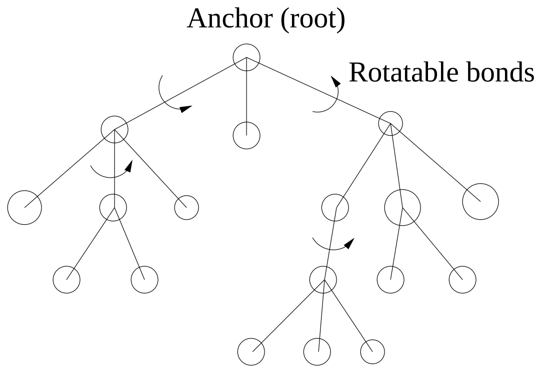

Figure 1: Tree representation of a molecule.

Method In the simple rotations method, all atom positions are updated by series of rotations. Below we show how to compute the new position $(x', y', z')^t$ of an atom initially at $(x, y, z)^t$ when the values of the torsional angles change. Here the superscript t denotes the vector transpose.

Let Q_{i-1} and Q_i be the parent and child atoms of a bond b_i. When bond b_i rotates by an angle θ_i, all descendant atoms of Q_i rotates around b_i by θ_i. This rotation is determined by (the centers of) the two atoms Q_{i-1}, Q_i, and the angle θ_i.

Let $(v_x, v_y, v_z)^t$ be the unit vector along bond b_i (from Q_i to Q_{i-1}). Assume P is an arbitrary point in space. Let the coordinates of P before

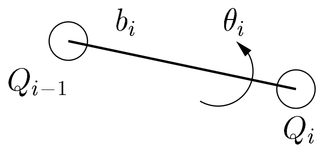

Figure 2: Bond b_i, atoms Q_{i-1}, Q_i, torsional angle θ_i.

and after the rotation around b_i by angle θ_i be $(x, y, z)^t$ and $(x', y', z')^t$ respectively. Then the coordinates $(x', y', z')^t$ and $(x, y, z)^t$ have the following relation:

$$\begin{pmatrix} x' & y' & z' & 1 \end{pmatrix}^t = R_i \cdot \begin{pmatrix} x & y & z & 1 \end{pmatrix}^t, \tag{1}$$

where the matrix R_i is given by [5]

$$T \cdot \begin{pmatrix} v_x^2 + (1 - v_x^2)c\theta_i & u_x v_y(1 - c\theta_i) + v_z s\theta_i & v_z v_x(1 - c\theta_i) + v_y s\theta_i & 0 \\ v_x v_y(1 - c\theta_i) + v_z s\theta_i & v_y^2 + (1 - v_y^2)c\theta_i & v_y v_z(1 - c\theta_i) - v_x s\theta_i & 0 \\ v_z v_x(1 - c\theta_i) - v_y s\theta_i & v_y v_z(1 - c\theta_i) + v_x s\theta_i & v_z^2 + (1 - v_z^2)c\theta_i & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot T^{-1}, \tag{2}$$

and T is the translation matrix

$$T = \begin{pmatrix} 1 & 0 & 0 & Q_{i,x} \\ 0 & 1 & 0 & Q_{i,y} \\ 0 & 0 & 1 & Q_{i,z} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In the above, $Q_{i,x}, Q_{i,y}, Q_{i,z}$ are the coordinates of Q_i; $c\theta_i = cos\theta_i$ and $s\theta_i = sin\theta_i$.

Thus, if b_i, b_{i-1}, …, b_1 is a sequence of bonds in the path from atom Q_i to the root (anchor) atom Q_0, the position of atom Q_i is updated (after the bonds b_1, …, b_{i-1} rotate by angles θ_1, …, θ_{i-1} respectively) by

$$\begin{pmatrix} x' & y' & z' & 1 \end{pmatrix}^t = R_{i-1} \dots R_1 \cdot \begin{pmatrix} x & y & z & 1 \end{pmatrix}^t.$$

Note that the rotation around bond b_i does not change the position of Q_i.

It is necessary to add the effect of the Euler angles since the whole molecule can rotate along the x, y, z axes. Let the Euler angles be α, β, γ, then the effect can be expressed in the following matrix [2]

$$E = \begin{pmatrix} c\alpha c\beta & c\alpha s\beta s\gamma - s\alpha c\gamma & c\alpha s\beta c\gamma + s\alpha s\gamma & 0 \\ s\alpha c\beta & s\alpha s\beta s\gamma + c\alpha c\gamma & s\alpha s\beta c\gamma - c\alpha s\gamma & 0 \\ -s\beta & c\beta s\gamma & c\beta c\gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

Taking the effect of Euler angles into account, the final position of atom Q_i is

$$(\, x' \quad y' \quad z' \quad 1\,)^t = R_{i-1} \dots R_i \cdot E \cdot (\, x \quad y \quad z \quad 1\,)^t. \tag{4}$$

Advantages The simple rotations scheme is conceptually straightforward. The coordinates of all atoms are referring to the global frame (world coordinate system). The implementation is also obvious: the coordinates of each atom are updated by applying a sequence of rotations. In fact, the rotation matrix for Q_{i-1} is $R_{i-2} \dots R_1 \cdot E$, and the rotation matrix for Q_i is $R_{i-1} \cdot R_{i-2} \dots R_1 \cdot E$. Therefore, if the rotation matrix is saved for atom Q_{i-1} and passed to Q_i, then Q_i can obtain its rotation matrix by one matrix multiplication $R_{i-1} \cdot (R_{i-2} \dots R_1 \cdot E)$ instead of a sequence of matrix multiplications.

Disadvantages The drawback of the simple rotations scheme is that bookkeeping of atom positions is necessary to update the coordinates of all atoms. Indeed, to compute the position of Q_i, the positions of all atoms at the path $b_1, \dots, b_{i-1}$ must be computed beforehand, because these positions are needed in matrices $R_j, 1 \le j \le i-1$. Thus, an atom can only be updated after all its ancestor atoms have been updated.

Moreover, errors accumulate in the simple rotations scheme. Whenever a bond rotates, the new position of an atom (after the rotation) depends on its old position (before the rotation). Therefore, any numerical error of the old position is accumulated to the new position. This accumulated error may grow beyond tolerance when many rotations are performed around the bonds. The choice of the anchor atom also affects accuracy: an anchor atom at the very end of a molecule chain will generate large errors at the other end.

3 Denavit-Hartenberg(DH) Local Frames

The Denavit-Hartenberg(DH) local frames method is a simple and widely used method for deriving molecular conformations. The theory behind it was initially developed in robotics. By considering molecules as tiny robots, one can transfer much of this theory to molecular modeling [2, 9]. In this section, we also assume that the underlying graphs of molecules are trees.

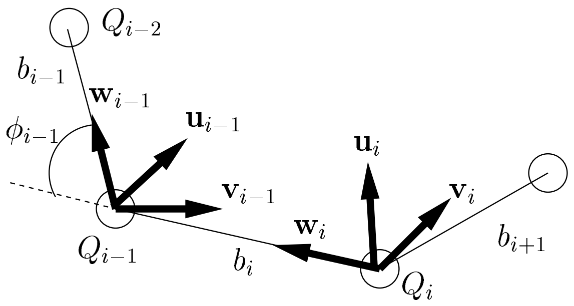

Figure 3: Local frames F_{i-1}, F_i, and bond angle ϕ_{i-1}.

Method In the DH local frames method, each local frame is attached at the end of a bond (the center of the child atom of the bond). Let b_i be a bond following bond b_{i-1}. A local frame $F_{i-1} = \{Q_{i-1}; \mathbf{u}_{i-1}, \mathbf{v}_{i-1}, \mathbf{w}_{i-1}\}$ is attached to bond b_{i-1} as follows: $\mathbf{w}_{i-1}$ has the direction of bond b_{i-1}; $\mathbf{u}_{i-1}$ is perpendicular to both b_{i-1} and b_i; $\mathbf{v}_{i-1}$ is perpendicular to both $\mathbf{u}_{i-1}$ and $\mathbf{w}_{i-1}$. Similarly, a local frame $F_i = \{Q_i; \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i\}$ is attached to bond b_i. The two local frames are related by the following matrix [2, 9]

$$R_i = \begin{pmatrix} c\theta_i & -s\theta_i & 0 & 0 \\ s\theta_i c\phi_{i-1} & c\theta_i c\phi_{i-1} & -s\phi_{i-1} & -l_i s\phi_{i-1} \\ s\theta_i s\phi_{i-1} & c\theta_i s\phi_{i-1} & c\phi_{i-1} & l_i c\phi_{i-1} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5}$$

where $c\theta_i = cos(\theta_i)$, $s\theta_i = sin(\theta_i)$, $c\phi_{i-1} = cos(\phi_{i-1})$, $s\phi_{i-1} = sin(\phi_{i-1})$, l_i is the length of bond b_i, θ_i is the torsional angle of b_i, and ϕ_{i-1} is the bond angle between bond b_{i-1} and b_i.

Note that local frames rotate along with the bonds. That is, frame F_i rotates along with bond b_i. Moreover, since all descendant atoms of Q_i rotate when bond b_i rotates, the local frames attached with these atoms will rotate along with bond b_i, too.

Assume P is an arbitrary point in space, and let the coordinates of P in frame F_{i-1} be $(x', y', z')^t$ and the coordinates of P in frame F_i be $(x, y, z)^t$. Then the two sets of coordinates are related as

$$\begin{pmatrix} x' & y' & z' & 1 \end{pmatrix}^t = R_i \cdot \begin{pmatrix} x & y & z & 1 \end{pmatrix}^t.$$

Now the position of any atom can be obtained by chaining the matrices R_i. If an atom Q_i is connected to the root atom Q_0 by a sequence of bonds

b_i, ..., b_1, then the coordinates of Q_i are

$$(\,x \quad y \quad z \quad 1\,)^t = R_1 \ldots R_i \cdot (\,0 \quad 0 \quad 0 \quad 1\,)^t. \tag{6}$$

If the root atom Q_0 lies at the origin of the global frame, i.e., the local frame F_0 is the same as the global frame, then the coordinates obtained from Equation (6) are referring to the global frame. Otherwise, the coordinates in the global frame of any atom can be computed by adding one translation

$$(\,x \quad y \quad z \quad 1\,)^t = Tr \cdot R_1 \ldots R_i \cdot (\,0 \quad 0 \quad 0 \quad 1\,)^t, \tag{7}$$

where

$$Tr = \begin{pmatrix} 1 & 0 & 0 & Q_{0,x} \\ 0 & 1 & 0 & Q_{0,y} \\ 0 & 0 & 1 & Q_{0,z} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{8}$$

and $Q_{0,x}, Q_{0,y}, Q_{0,z}$ are the coordinates of Q_0.

Combining the effect of Euler angles, the final position of Q_i in the global frame is

$$(\,x \quad y \quad z \quad 1\,)^t = Tr \cdot E \cdot R_1 \ldots R_i \cdot (\,0 \quad 0 \quad 0 \quad 1\,)^t. \tag{9}$$

Advantages To update the position of an atom, the DH local frames method does not need to keep the positions of all its ancestor atoms — no bookkeeping is necessary. Instead, the method records the angles of the bonds in the matrices R_i. Thus multiple rotations around a bond can be combined into one rotation. Once the rotation matrix of an atom is known, the final position of the atom is just the last column of the matrix. Moreover, the numerical error of the old position of an atom does not propagate to the new position of the atom when a bond rotates — the new position does not depend on the old position. Finally, note that the accuracy of calculation is independent of the choice of the anchor atom.

Disadvantages The drawback, however, of the DH local frames method is that multiple local frames are needed if an atom has more than one child. For example, if atom Q_{i-1} has two children atoms Q_i and Q'_i, two local frames will be attached to Q_{i-1} since bonds b_i and b'_i are different (see Figure 4).

Therefore, the position of Q_i is updated using

$$Tr \cdot E \cdot R_1 \ldots R_{i-2} \cdot R_{i-1} \cdot R_i,$$

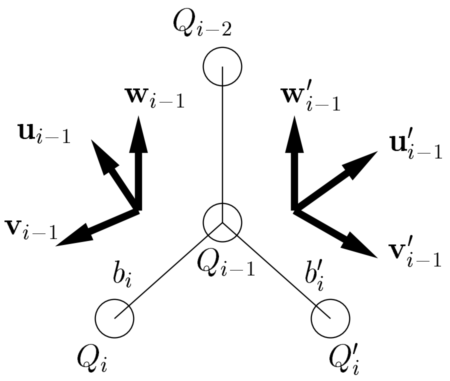

Figure 4: Multiple local frames at Q_{i-1} (both have origin at Q_{i-1}).

while the position of Q'_i is updated using

$$Tr \cdot E \cdot R_1 \ldots R_{i-2} \cdot R'_{i-1} \cdot R'_i,$$

where the last two matrices are different: $R_{i-1} \neq R'_{i-1}$, $R_i \neq R'_i$. As far as implementation is concerned, the product $Tr \cdot E \cdot R_1 \ldots R_{i-2}$ is saved for Q_{i-1}, and then used to compute the position of Q_{i-1} (together with R_{i-1} or R'_{i-1}). This product is inherited by Q_i and Q'_i to update their positions (together with $R_{i-1} \cdot R_i$ and $R'_{i-1} \cdot R'_i$). Thus, if an atom has only one child, then its rotation matrix can be completely inherited by the child; the child obtains its rotation matrix by one matrix multiplication. Otherwise, the rotation matrix is only partially inherited by the children; the children obtain their rotation matrices by two matrix multiplications.

Another drawback of the DH local frames method is that local frames are attached to every bond, whether a bond rotates or not. Since the non-rotatable bonds just provide the spatial structure, carrying local frames (especially multiple local frames) is unnecessary. The atomgroup local frames method in Section 4 will eliminate this redundancy and inconvenience.

4 Atomgroup Local Frames

To speed up the computations and efficiently calculate molecular conformations, we introduce the concept of *atomgroups*. A set of connected atoms are in one atomgroup if none of the bonds between these atoms rotates. Equivalently, each rotatable bond separates two distinct atomgroups. It is easy to see that the number of atomgroups is one more than the number of rotatable bonds.

We will assume in this section that all atoms are grouped into atomgroups. Therefore, all atoms of a ring are in one atomgroup; there is no need to replace rings by special atoms. Thus, the underlying graph of a molecule is a tree with each vertex denoting an atomgroup and each edge denoting a (rotatable) bond between a pair of atomgroups (see Figure 5). Once an atomgroup is chosen as the root group, the parent-children relations among all atomgroups are well defined.

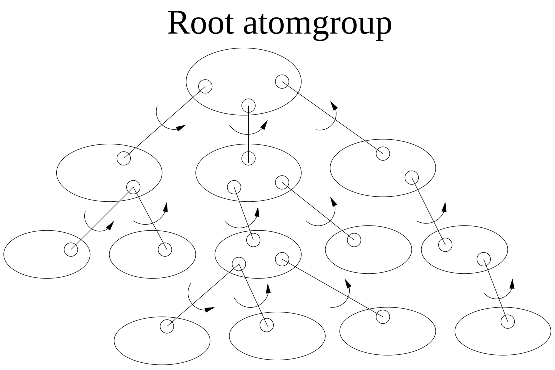

Figure 5: Atomgroups of a molecule.

Below, we introduce a new scheme to attach local frames to atomgroups. For each atomgroup, only one local frame is needed no matter how many children atomgroups there are.

4.1 Attaching Local Frames

A local frame $F_i = \{Q_i; \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i\}$ is attached to atomgroup g_i as follows (Figure 6):

> Q_i is the child atom (in g_i) of bond b_i; $\mathbf{w}_i$ is the unit vector along bond b_i pointing toward g_{i-1}; $\mathbf{u}_i$ is an arbitrary unit vector perpendicular to $\mathbf{w}_i$; $\mathbf{v}_i$ is perpendicular to $\mathbf{w}_i$ and $\mathbf{u}_i$.

Note that the construction of local frame F_i depends on the atomgroup g_i and its parent group g_{i-1}; no information about g_i's children groups is used. Therefore, there is only one local frame F_i constructed for g_i no matter how many children groups g_i may have.

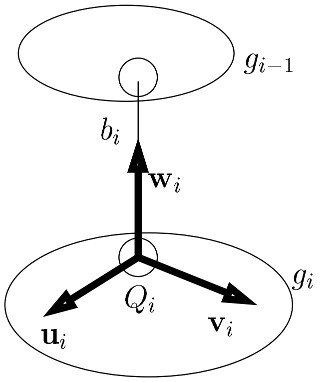

Figure 6: Local frame $F_i = \{Q_i; \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i\}$ at atomgroup g_i.

The root group does not have any parent group, so we will translate the global frame to any atom in the root group.

4.2 Relational Matrices

Suppose that the frame at atomgroup g_i is $F_i = \{Q_i; \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i\}$ and the frame at its parent group g_{i-1} is $F_{i-1} = \{Q_{i-1}; \mathbf{u}_{i-1}, \mathbf{v}_{i-1}, \mathbf{w}_{i-1}\}$. Below we derive the relation between F_i and F_{i-1}.

Let P be any point in space. Suppose P has coordinates $(x_i, y_i, z_i)^t$ in frame F_i, and coordinates $(x_{i-1}, y_{i-1}, z_{i-1})^t$ in frame F_{i-1}. When bond b_i rotates by an angle θ_i, frame F_i also rotates by angle θ_i (around $\mathbf{w}_i$ axis) (see Figure 7). Let the original frame (before b_i rotates) be $F_i' = \{Q_i; \mathbf{u}_i', \mathbf{v}_i', \mathbf{w}_i\}$. Let the local coordinates of P be $(x_i', y_i', z_i')^t$ in frame F_i', then

$$\begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_{i-1}\cdot\mathbf{u}_i' & \mathbf{u}_{i-1}\cdot\mathbf{v}_i' & \mathbf{u}_{i-1}\cdot\mathbf{w}_i \\ \mathbf{v}_{i-1}\cdot\mathbf{u}_i' & \mathbf{v}_{i-1}\cdot\mathbf{v}_i' & \mathbf{v}_{i-1}\cdot\mathbf{w}_i \\ \mathbf{w}_{i-1}\cdot\mathbf{u}_i' & \mathbf{w}_{i-1}\cdot\mathbf{v}_i' & \mathbf{w}_{i-1}\cdot\mathbf{w}_i \end{pmatrix} \cdot \begin{pmatrix} x_i' \\ y_i' \\ z_i' \end{pmatrix} + \begin{pmatrix} \mathbf{u}_{i-1}\cdot(Q_i - Q_{i-1}) \\ \mathbf{v}_{i-1}\cdot(Q_i - Q_{i-1}) \\ \mathbf{w}_{i-1}\cdot(Q_i - Q_{i-1}) \end{pmatrix}. \tag{10}$$

It is easy to determine that

$$\begin{pmatrix} x_i' \\ y_i' \\ z_i' \end{pmatrix} = \begin{pmatrix} c\theta_i & -s\theta_i & 0 \\ s\theta_i & c\theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}. \tag{11}$$

It follows from Equations (10), (11) that the coordinates $(x_i, y_i, z_i)^t$ of a point P in frame F_i and the coordinates $(x_{i-1}, y_{i-1}, z_{i-1})^t$ of P in F_{i-1} are related

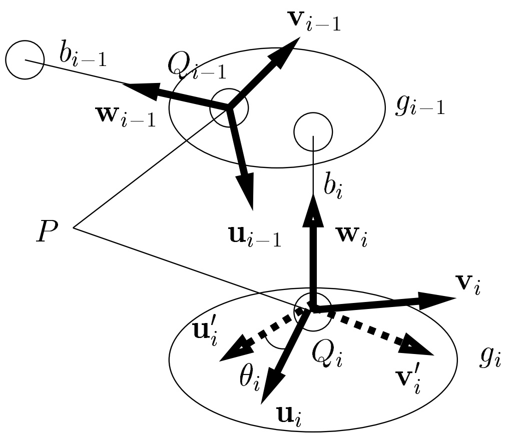

Figure 7: Local frame $F_i = \{Q_i; \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i\}$ at atomgroup g_i; local frame $F'_i = \{Q_i; \mathbf{u}'_i, \mathbf{v}'_i, \mathbf{w}_i\}$ is frame F_i before bond b_i rotates by angle θ_i; local frame $F_{i-1} = \{Q_{i-1}; \mathbf{u}_{i-1}, \mathbf{v}_{i-1}, \mathbf{w}_{i-1}\}$.

by

$$\begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_{i-1} \cdot \mathbf{u}'_i & \mathbf{u}_{i-1} \cdot \mathbf{v}'_i & \mathbf{u}_{i-1} \cdot \mathbf{w}_i \\ \mathbf{v}_{i-1} \cdot \mathbf{u}'_i & \mathbf{v}_{i-1} \cdot \mathbf{v}'_i & \mathbf{v}_{i-1} \cdot \mathbf{w}_i \\ \mathbf{w}_{i-1} \cdot \mathbf{u}'_i & \mathbf{w}_{i-1} \cdot \mathbf{v}'_i & \mathbf{w}_{i-1} \cdot \mathbf{w}_i \end{pmatrix} \begin{pmatrix} c\theta_i & -s\theta_i & 0 \\ s\theta_i & c\theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \\ + \begin{pmatrix} \mathbf{u}_{i-1} \cdot (Q_i - Q_{i-1}) \\ \mathbf{v}_{i-1} \cdot (Q_i - Q_{i-1}) \\ \mathbf{w}_{i-1} \cdot (Q_i - Q_{i-1}) \end{pmatrix}. \tag{12}$$

Note that the dot products in Equation (12) all remain unchanged when bond b_i rotates. (Of course, these dot products do not change when any other bonds rotates.) Thus these dot products can be computed and saved for each atomgroup once the local frames are constructed. Therefore, for each atom A in atomgroup g_i, the coordinates $(x_i, y_i, z_i)^t$ in F_i and the coordinates $(x_{i-1}, y_{i-1}, z_{i-1})^t$ in F_{i-1} are related by

$$(\, x_{i-1} \quad y_{i-1} \quad z_{i-1} \quad 1\,)^t = R_i \cdot (\, x_i \quad y_i \quad z_i \quad 1\,)^t, \tag{13}$$

where the relational matrix R_i is

$$\begin{pmatrix} \mathbf{u}_{i-1}\cdot\mathbf{u}'_i & \mathbf{u}_{i-1}\cdot\mathbf{v}'_i & \mathbf{u}_{i-1}\cdot\mathbf{w}_i & \mathbf{u}_{i-1}\cdot(Q_i-Q_{i-1}) \\ \mathbf{v}_{i-1}\cdot\mathbf{u}'_i & \mathbf{v}_{i-1}\cdot\mathbf{v}'_i & \mathbf{v}_{i-1}\cdot\mathbf{w}_i & \mathbf{v}_{i-1}\cdot(Q_i-Q_{i-1}) \\ \mathbf{w}_{i-1}\cdot\mathbf{u}'_i & \mathbf{w}_{i-1}\cdot\mathbf{v}'_i & \mathbf{w}_{i-1}\cdot\mathbf{w}_i & \mathbf{w}_{i-1}\cdot(Q_i-Q_{i-1}) \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c\theta_i & -s\theta_i & 0 & 0 \\ s\theta_i & c\theta_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{14}$$

4.3 Updating the Atom Positions

Suppose $g_i, g_{i-1}, \ldots, g_0$ is a sequence of groups, where g_j is the parent group of g_{j+1}, $0 \le j \le i-1$, and g_0 is the root group. Then the coordinates in the local frame at the root group of atom $A \in g_i$ are

$$(\,x \quad y \quad z \quad 1\,)^t = R_1 \ldots R_i \cdot (\,x_i \quad y_i \quad z_i \quad 1\,)^t,$$

where (x_i, y_i, z_i) are the coordinates of A in the local frame at atomgroup g_i. Taking the effect of the Euler angles into account, and also considering that the local frame at the root group may be different from the global frame, the coordinates of any atom A ($\in g_i$) in the global frame are

$$(\,x \quad y \quad z \quad 1\,)^t = Tr \cdot E \cdot R_1 \ldots R_i \cdot (\,x_i \quad y_i \quad z_i \quad 1\,)^t. \tag{15}$$

In the atomgroup local frames method, all atoms within an atomgroup share the same rotation matrix. Thus the number of rotation matrices is greatly reduced from the number of total bonds to the number of rotatable bonds — the former is generally several times larger than the latter. The atomgroup local frames method records the rotations around the bonds in the relational matrices. Thus, it can combine multiple rotations around one bond into one rotation. Moreover, to update the position of an atom, the method does not have to keep records of the positions of the ancestor atoms. In fact, the method does not need the position of any other atom to update the position of an atom. The update of any atom position can be delayed until it is necessary, for example when its position is needed to check if it collides with another atom or structure in a docking simulation experiment. Therefore, the atomgroup local frames method provides lazy evaluation for atom positions.

Comparing with the DH local frames method, the atomgroup local frames method attaches only one local frame to each rotatable bond, no matter how many children bonds there exist. The relations between the parent-children frames are easily determined by Equation (14). In the product of two matrices, the first matrix remains constant when any bond of the molecule rotates, so it can be saved when the local frames are constructed.

Updating the positions of the atoms in the atomgroup local frames method becomes easier than in the method of DH local frames method. In fact, all atoms in atomgroup g_i share the same rotation matrix $Tr \cdot E \cdot R_1 \dots R_i$. The position of every atom in g_i is updated by multiplying the rotation matrix to its local coordinates in frame F_i — these local coordinates remain fixed. Atomgroup g_i inherits the rotation matrix from g_{i-1} and obtains its own rotation matrix by one matrix multiplication: $(Tr \cdot E \cdot R_1 \dots R_{i-1}) \cdot R_i$.

5 Comparison and Discussion

In this section, we give a quantitative comparison of the three methods for deriving the molecular conformations presented in this paper.

Let the number of atoms be n_a and the number of rotatable bonds be n_{rb}. Note that the number of atomgroups is $n_{rb} + 1$. Below we give the number of (scalar) multiplications needed to update the positions of all atoms in a molecule. (Here we ignore the cost of computing the sine and cosine functions of the torsional angles since the cost is the same for all three methods.)

In the simple rotations scheme, if bond b_i rotates, 39 multiplications are required to obtain the matrix R_i from Equation (2): 24 multiplications to compute the middle matrix; 9 multiplications for multiplying T^{-1}; 3 multiplications, 1 square root operation, and 3 divisions to compute the unit vector $(v_x, v_y, v_z)^t$. R_i is then multiplied with the rotation matrix inherited from the parent atom, and this requires 36 multiplications — a 3×3 matrix times a 3×4 matrix. If bond b_i does not rotate at all, then the matrix R_i is identity; atom Q_i just passes its rotation matrix to its child(ren). Since there are n_{rb} rotatable bonds, $(39 + 36) * n_{rb} = 75n_{rb}$ multiplications are needed to update the rotation matrices. Each atom needs 9 additional multiplications to obtain its coordinates. Therefore, the total number of multiplications to update all atom positions in a molecule is $75n_{rb} + 9n_a$.

In the DH local frames method, depending on whether a bond has one or more children bonds, each atom needs to compute one or two matrices R_i (Equation (5)) or R_{i-1}, R_i [c.f. the end of Section 3]. Thus, it takes 6 or 12 multiplications to obtain R_i or R_{i-1}, R_i. Each atom needs one or two matrix multiplications to get its rotation matrix which requires 30 or 60 scalar multiplications (each R_i has two zeroes in the first row). The coordinates of each atom are simply the last column of the rotation matrix — no additional multiplications needed. Thus $(6+30)*n_a = 36n_a$ to $(12+60)*n_a = 72n_a$ multiplications are required to compute all atom positions. If we assume that half of the bonds have more than one children bonds, then the total number of multiplications needed to update atom positions in a molecule is $(36n_a + 72n_a)/2 = 54n_a$.

In the atomgroup local frames method, all atoms within an atomgroup share the same rotation matrix. For each atomgroup, it takes 12 multiplications to obtain the matrix R_i (Equation (14)) and 36 additional multiplications to compute the rotation matrix. Each atom within the atomgroup still needs 9 multiplications to compute its coordinates. Therefore, the number of multiplications needed to update the atom positions in a molecule is $(12+36)*n_{rb} + 9*n_a = 48n_{rb} + 9n_a$.

Usually the number of atoms is much bigger than the number of rotatable bonds. If we assume that $n_a = 3*n_{rb}$, then multiplications needed by simple rotations scheme, DH local frames method, and the atomgroup local frames method is $102n_{rb}$, $162n_{rb}$, $75n_{rb}$ respectively.

We run all three algorithms using two molecules shown in Figure 8.

The running time is measured for 10,000 rounds of random rotations and updates of all atom positions for each round of rotation. The atomgroup local frames method is 33% and 28% faster than the simple rotations scheme, and 156% and 166% faster than the DH local frames method. The results are shown in Table 1.

	molecule 1	molecule 2
simple rotations	1	1
DH local frames	1.92	2.07
atomgroup local frames	0.75	0.78

Table 1: Normalized running time of the three methods.

We also have the following observations about the three methods.

(i) From Equation (2), if bond b_{i-1} does not rotate, then R_{i-1} is the identity matrix; hence atom Q_i directly inherits its rotation matrix from its parent. Therefore, in the simple rotations scheme, the number of

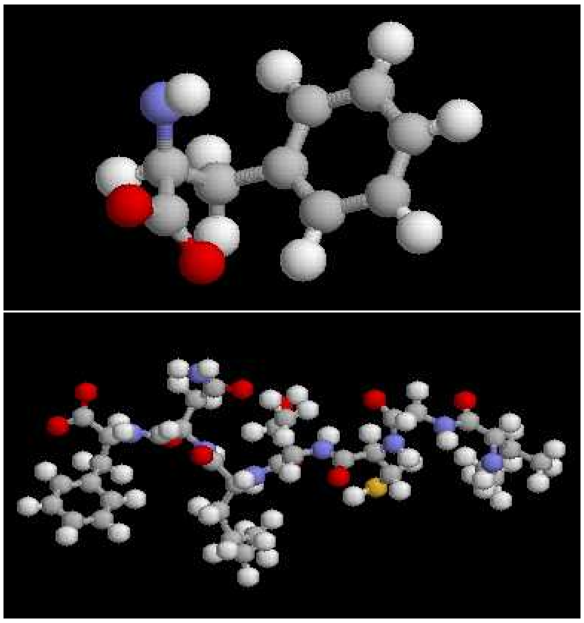

Figure 8: Two testing molecules.

rotation matrix updates equals the number of rotatable bonds. However, in the DH local frames method, even if a bond b_i does not rotate, the relational matrix R_i [c.f. Equation (5)] may not necessarily be the identity matrix. Thus a matrix multiplication is still needed for Q_i to obtain its rotation matrix. Therefore, the number of rotation matrix updates equals the number of all bonds. This explains why the DH local frames method is much slower than the simple rotations scheme. It is also easy to see that by grouping atoms into atomgroups, the simple rotations scheme does not improve its performance. We implemented this idea and our results verified this observation.

(ii) In the DH local frames method, the value of each torsional angle θ_i has a difference (by construction) from the value of θ_i in the simple rotations methods or the atomgroup local frames method. These differences can be calculated and saved as the DH local frames are constructed; they will be added later to report the correct torsional angles for the rotatable bonds.

(iii) The simple rotations scheme can be modified to prevent the numerical errors from accumulating upon the rotations around the bonds. This can be done by changing the meaning of the θ_i: instead of recording the torsional angle change each round, θ_i records the sum of all torsional angle changes of bond b_i. There is a disadvantage for this change: now the number of matrix updates equals the number of total bonds instead of the number of rotatable bonds. That is, the computational complexity increases at the tradeoff of eliminating numerical error accumulation.

(iv) The statistics for the method of atomgroup local frames in Table 1 do not include potential time savings that can result from lazy evaluations. In many applications such as ligand-receptor docking, the ligand-receptor complex is represented by a *moving shell*, including atoms of the ligand, contact residues from the protein and perhaps water molecules. This moving shell is surrounded by a *frozen shell* comprised of the rest of the protein structure and the *outer shell* comprised of the solvent. The conformational changes affect only the moving shell and the rest of the complex can be ignored to reduce the computational burden tremendously. To further reduce the computational complexity, we can focus on the final positions of a small set of key atoms (features). The positions of other atoms in the moving shell are less important as long as there is no collision. Therefore, it is not necessary to compute the positions of the non-key atoms in the moving shell until the key atoms are in target positions. The atomgroup local frames method thus provides a lazy evaluation mechanism — the position of any atom is not computed until it is needed. This mechanism can be exploited to reduce computational time and is not available in the simple rotations scheme since the positions of children atoms depend on the updated positions of ancestor atoms.

The above observations reveal the superiority of the atomgroup local frames method over the simple rotations scheme and the DH local frames method, but further work is needed when bond angles, bond lengths as well as the torsional angles change. The relational matrix R_i [c.f. Equation (14)] in the atomgroup local frames method is decomposed into two matrices where the first one remains constant and the second is very simple and sparse (many zero entries). This is because of the assumption that all conformational changes are due to the changes of the torsional angles. If the bond angles also change, then the second matrix in Equation (14) is no longer simple or sparse. Therefore, the execution of the algorithm will be considerably slower. A new method which has no bookkeeping, no numeric error accumulation, and minimal local frames is still under investigation.

6 Conclusion

The atomgroup local frames method is superior to earlier approaches for representing, deriving and updating molecular conformations as it eliminates bookkeeping, error accumulation, and provides complete inheritance of rotation matrices. The usage of this method in pharmacophore constrained conformational search is currently under investigation by the authors. A comparative evaluations of the three methods are summarized in Table 2.

	bookkeeping needed	rot. mat. complete inheritance	errors accumulation
simple rotations	yes	yes	yes
DH local frames	not	not	no
atomgroup local frames	not	yes	no

Table 2: Summary of three methods for deriving molecular conformations.

Acknowledgment Work on this paper by Ming Zhang and Lydia Kavraki is supported in part by ATP 003604-0120-1999, NSF CISE SA1728-21122N, the Whitaker Foundation, NSF EIA-0072743, and a Sloan Fellowship awarded to Lydia Kavraki. The authors would like to thank Miguel Teodoro, Brian Chen and Andrew Ladd for useful discussions. The authors also thank the referee for the insightful comments and suggestions.

References

[1] Cox, D.; Little, J.; O'Shea, D. *Using Algebraic Geometry.* Springer-Verlag New York, Inc., 1998.

[2] Craig, J. J. *Introduction to Robotics.* Addison-Wesley, Reading, MA, 1989.

[3] Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation.* Research Studies Press Ltd., 1988.

[4] Finn, P. W.; Kavraki, L. E. Computational Approaches to Drug Design. *Algorithmica*, 1999, 25, 347–371.

[5] Foley, J. D.; van Dam, A.; Feiner, S. K.; Hughes, J. F. *Computer Graphics: Principles and Practice*. Addison-Wesley, Reading, MA, 1990.

[6] Gardiner, E.J.; Willett, P.; Artymiuk, P. J. Graph-theoretic techniques for macromolecular docking, *Journal of Chemical Information and Computer Sciences*, 2000, 40, 273-279.

[7] Henry, D. R.; Ozkabak, A. G. Conformational Flexibility in 3D Structure Searching. *Encyclopedia of Computational Chemistry*. Schleyer, P. v. R., Ed., Wiley, New York, 1998.

[8] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Further development of a genetic algorithm for ligand docking and its application to screening combinatorial libraries. *ACS Symposium Series (Rational Drug Design: Novel Methodology and Practical Applications)*, 1999, 719, 271-291.

[9] Lavalle, S. M.; Finn, P. W.; Kavraki, L. E.; Latombe, Jean-Claude. A Randomized Kinematics-based Approach to Pharmacophore-Constrained Conformational Search and Database Screening. *Journal of Computational Chemistry*, 2000, Vol. 21, No. 9, 731–747.

[10] Lipton, M.; Still, W. The Multiple Minimum Problem in Molecular Modeling. Tree Searching Internal Coordinate Conformational Space. *Journal of Computational Chemistry*, 1998, Vol. 9, 343–355.

[11] Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of Conformational Coverage. 1. Validation and Estimation of Coverage. *Journal of Chemical Information and Computer Sciences*, 1995, Vol. 35, 285–294.

[12] Song, G.; Amato, N. M. Using Motion Planning to Study Protein Folding Pathways. *Proceedings of the Fifth Annual International Conference on Computational Biology*, 2001, 287-296.