# Computing and Updating Molecular Conformations Using the Atomgroup Local Frames Method

Ming Zhang Lydia E. Kavraki
{mzhang, kavraki}@cs.rice.edu
Department of Computer Science
Rice University, Houston TX 77005

**Abstract**

Efficiently maintaining molecular conformations is important for molecular modeling and protein engineering. This paper reviews the widely used simple rotations scheme, simple local frames method, and introduces a new atomgroup local frames method for maintaining the molecular conformation changes due to the changes of torsional angles. The simple rotations scheme applies a sequence of rotations to update all atom positions. The order of the updates is important and some bookkeeping of the atom positions is necessary. Numeric errors can accumulate as rotations around the bonds are repeated. The simple local frames method builds local frames at the bonds, and relational matrices between parents and children frames are constructed. The atom positions are computed by chaining series of such matrices. No bookkeeping is necessary and numeric errors do not accumulate upon rotations. 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 maintain molecular conformations. A single local frame is attached to each atomgroup. Bookkeeping is not necessary and numeric errors do not accumulate upon rotations. This method also provides lazy evaluations for atom positions. Thus, the conformational maintenance cost is greatly reduced, especially when many conformations are generated and updated such as in a minimization process.

## 1 Introduction

During the last two decades a lot of emphasis has been placed in studying problems that involve reasoning about three dimensional molecular conformations (e.g., docking, protein folding, molecular dynamics etc.) [6]. Little research has been done, however, to understand how to efficiently represent and maintain molecular conformations. The efficient maintenance 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 maintaining molecular conformations. Two available methods, the simple rotation scheme and the simple local frame method, are reviewed and a new method, the atomgroup local frame 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, and torsional angles. In most kinematics studies, the van der Waals radii, electric charges, bond lengths, and bond angles are considered constants, while the torsional angles are allowed to change. We follow this assumption in this paper. That is, conformational changes are solely due to the changes of torsional angles. Many kinematics problems are ultimately reduced to searching the conformational spaces for "good" conformations of molecules and proteins.

Since conformational spaces are usually high-dimensional, the conformational search is indeed a computationally expensive process. For over two decades, people have developed many algorithms to perform conformational search [1, 4, 7, 9, 10]. The common theme of these algorithms is to reduce the size of the search space and try to find as many solution candidates as possible. Even with these algorithms, still a large number of conformations are checked, analyzed; computations spent on these checking, analyzing still take very long time. It is therefore desirable to have good data structures to efficiently maintain molecular conformations so to considerably reduce the computational cost. Many applications such as ligand-receptor docking, protein folding, and molecular dynamics will benefit from the efficient maintenance of molecular conformations.

Available methods for maintaining molecular conformations include the simple rotations scheme and the simple local frames method [8]. 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 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 simple local frames method builds local frames at the bonds, and 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. 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 maintain molecular conformations. 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 conformational maintenance cost can be greatly reduced, especially when many conformations are generated and updated such as in a minimization process.

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

# 2 Simple Rotations

A molecule is often characterized by a collection of atoms and a collection of bonds between pairs of atoms. A graph can then represent the molecule where vertices denote atoms and edges denote bonds. An atom will be chosen as the anchor for the molecule — root of the corresponding graph. In this section, we assume that the underlying graphs are trees — that is, there are no rings in the molecules or 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-child relations among the atoms are unambiguous. The following figure gives an example.

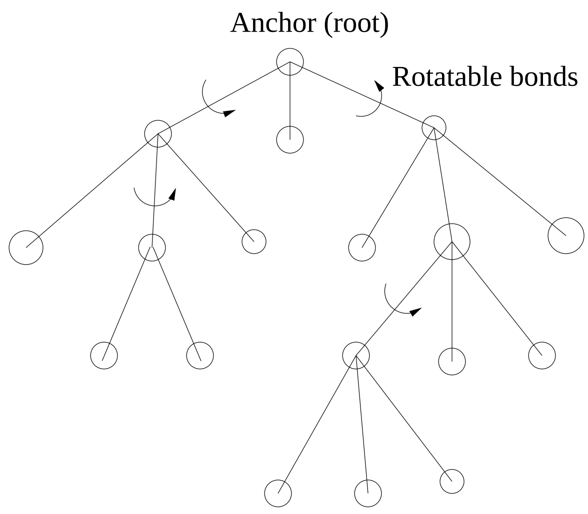


Figure 1: Tree representation of a molecule.

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 $\theta_i$, all descendant atoms of $Q_i$ rotates around $b_i$ by $\theta_i$. This rotation is determined by the two atoms $Q_{i-1}$, $Q_i$, and the angle $\theta_i$.

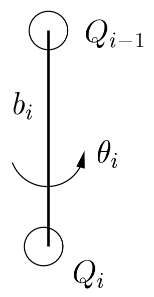


Figure 2: Bond $b_i$, its parent and child atoms $Q_{i-1}$, $Q_i$.

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 and after the rotation around

$b_i$ by angle $\theta_i$ 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:

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = Rot(b_i,\theta_i) \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \tag{1}$$

where the rotation matrix $Rot(b_i,\theta_i)$ is [5]

$$T \cdot \begin{pmatrix} v_x^2+(1-v_x^2)c\theta_i & u_xv_y(1-c\theta_i)+v_zs\theta_i & v_zv_x(1-c\theta_i)+v_ys\theta_i & 0 \\ v_xv_y(1-c\theta_i)+v_zs\theta_i & v_y^2+(1-v_y^2)c\theta_i & v_yv_z(1-c\theta_i)-v_xs\theta_i & 0 \\ v_zv_x(1-c\theta_i)-v_ys\theta_i & v_yv_z(1-c\theta_i)+v_xs\theta_i & v_z^2+(1-v_z^2)c\theta_i & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot T^{-1},$$

and $T$ is the translation matrix

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

($Q_{i,x}, Q_{i,y}, Q_{i,z}$ are the coordinates of $Q_i$; $c\theta_i$ is $cos\theta_i$ and $s\theta_i$ is $sin\theta_i$.)

Thus if $b_i$, $b_{i-1}$, $\cdots$, $b_1$ are 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 $b_1$, $\cdots$, $b_{i-1}$ rotate by angles $\theta_1$, $\cdots$, $\theta_{i-1}$ respectively) by

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = Rot(b_{i-1},\theta_{i-1}) \cdots Rot(b_1,\theta_1) \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}.$$

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. If the angles rotated around the three axes ($\{(1,0,0)^T, (0,1,0)^T, (0,0,1)^T\}$) are $\alpha,\beta,\gamma$, then the effect can be expressed in the following matrix [3]

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

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

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = Rot(b_{i-1},\theta_{i-1}) \cdots Rot(b_1,\theta_1) \cdot E(\alpha,\beta,\gamma) \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}. \tag{2}$$

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 all atoms are updated by applying a sequence of rotations. In fact, the rotation matrix for $Q_{i-1}$ is $Rot(b_{i-2}, \theta_{i-2}) \cdots Rot(b_1, \theta_1)$, and the rotation matrix for $Q_i$ is $Rot(b_{i-1}, \theta_{i-1}) \cdot Rot(b_{i-2}, \theta_{i-2}) \cdots Rot(b_1, \theta_1)$. 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 instead of applying a sequence of matrix multiplications.

The drawback of the simple rotations scheme is that a 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 $b_1, \cdots, b_{i-1}$ must be computed beforehand, because these positions are needed in the rotation matrices $Rot(b_j, \theta_j), 1 \leq j \leq i-1$. 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.

# 3 Simple Local Frames

The simple local frames method is the simplest and most widely used method for maintaining molecular conformations. This method was developed in robotics [3, 8]. In this section, we also assume that underlying graphs of molecules are trees.

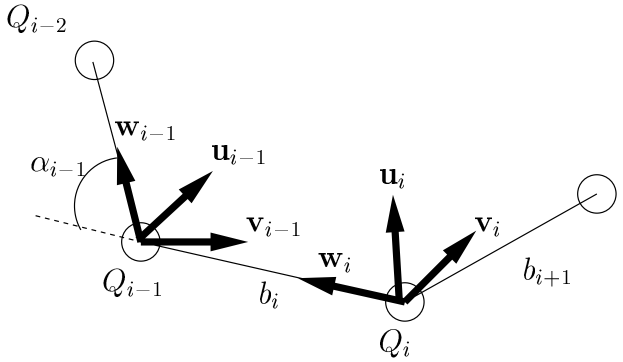


Figure 3: Local frames $F_{i-1}, F_i$, and bond angle $\alpha_{i-1}$.

In the simple 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 [3, 8]

$$Rel_i = \begin{pmatrix} c\theta_i & -s\theta_i & 0 & 0 \\ s\theta_i c\alpha_{i-1} & c\theta_i c\alpha_{i-1} & -s\alpha_{i-1} & -l_i s\alpha_{i-1} \\ s\theta_i s\alpha_{i-1} & c\theta_i s\alpha_{i-1} & c\alpha_{i-1} & l_i c\alpha_{i-1} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3}$$

where $c\theta_i = cos(\theta_i)$, $s\theta_i = sin(\theta_i)$, $c\alpha_{i-1} = cos(\alpha_{i-1})$, $s\alpha_{i-1} = sin(\alpha_{i-1})$, $l_i$ is the length of bond $b_i$, $\theta_i$ is the torsional angle of $b_i$, and $\alpha_{i-1}$ 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, so 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 have the relation

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = Rel_i \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}.$$

The root atom $Q_0$ does not have any parent atom, so we attach a local frame $F_0$ at $Q_0$ by shifting the global frame to $Q_0$. This local frame and the local frame $F_1$ (at $Q_1$) are related by a matrix $Rel_1$. Assume frame $F_1 = \{Q_1; \mathbf{u}_1^*, \mathbf{v}_1^*, \mathbf{w}_1^*\}$ before any rotation is applied to bond $b_1$, then $Rel_1$ is

$$Rel_1 = \begin{pmatrix} \mathbf{u}_1^* & \mathbf{v}_1^* & \mathbf{w}_1^* & Q_1 - Q_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} c\theta_1 & -s\theta_1 & 0 & 0 \\ s\theta_1 & c\theta_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Now the position of each atom can be obtained by chaining the matrices $Rel_i$. If an atom $Q_i$ is connected to the root atom $Q_0$ by a sequence of bonds $b_i$, $\cdots$, $b_1$, then the coordinates of $Q_i$ are

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = Rel_1 \cdots Rel_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{4}$$

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 (4) are referring to the global frame. Otherwise, the coordinates in the global frame of any atom can be computed by applying one more translation:

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = Tr(Q_0) \cdot Rel_1 \cdots Rel_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \tag{5}$$

where

$$Tr(Q_0) = \begin{pmatrix} 1 & 0 & 0 & Q_{0,x} \\ 0 & 1 & 0 & Q_{0,y} \\ 0 & 0 & 1 & Q_{0,z} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{6}$$

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

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = Tr(Q_0) \cdot E(\alpha, \beta, \gamma) \cdot Rel_1 \cdots Rel_i \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{7}$$

To update the position of an atom, the simple 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 $Rel_i$. 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 numeric error of the old position of an atom does not accumulate to the new position of the atom whenever a bond rotates — the new position does not depend on the old position.

The drawback, however, of the simple 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 child 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.

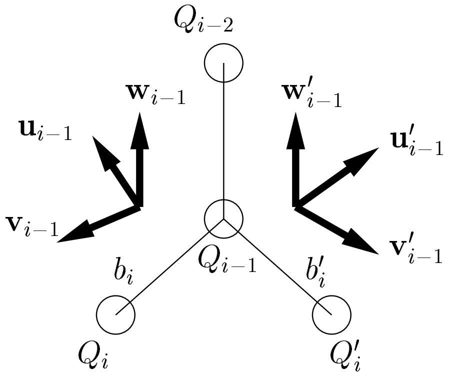


Figure 4: Multiple local frames at $Q_{i-1}$ (both have origin at $Q_{i-1}$).

Therefore the position of $Q_i$ is updated using

$$Tr(Q_0) \cdot E(\alpha, \beta, \gamma) \cdot Rel_1 \cdots Rel_{i-2} \cdot Rel_{i-1} \cdot Rel_i,$$

while the position of $Q_i'$ is updated using

$$Tr(Q_0) \cdot E(\alpha, \beta, \gamma) \cdot Rel_1 \cdots Rel_{i-2} \cdot Rel'_{i-1} \cdot Rel'_i,$$

where the last two matrices are different: $Rel_{i-1} \neq Rel'_{i-1}$, $Rel_i \neq Rel'_i$. As far as implementation is concerned, the product $Tr(Q_0) \cdot E(\alpha, \beta, \gamma) \cdot Rel_1 \cdots Rel_{i-2}$ is saved for $Q_{i-1}$, and then used to compute the position of $Q_{i-1}$ (together with $Rel_{i-1}$ or $Rel'_{i-1}$). This product is inherited by $Q_i$ and $Q'_i$ to update their positions (together with $Rel_{i-1} \cdot Rel_i$ and $Rel'_{i-1} \cdot Rel'_i$). 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 child — the child obtains its rotation matrix by two matrix multiplications.

Another drawback of the simple 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 these redundancy and inconvenience.

# 4 Atomgroup Local Frames

To speed up the computations and efficiently maintain molecular conformations, we introduce the concept of *atomgroups*. A set of atoms which are connected 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 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 still a tree, where each vertex denotes an atomgroup and each edge denotes a bond between a pair of atomgroups. One atomgroup is chosen as the root group and the parent-child relations among all atomgroups are clear. We will assume that there are no rings of atomgroups. Therefore, once the root group is chosen, the parent-child relations among the atomgroups are unambiguous.

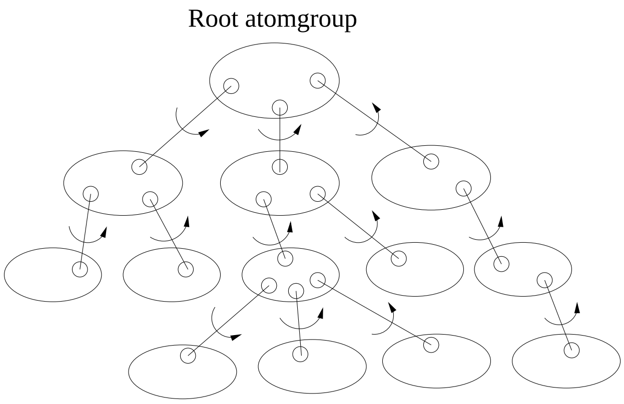


Figure 5: Atomgroups of a molecule.

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

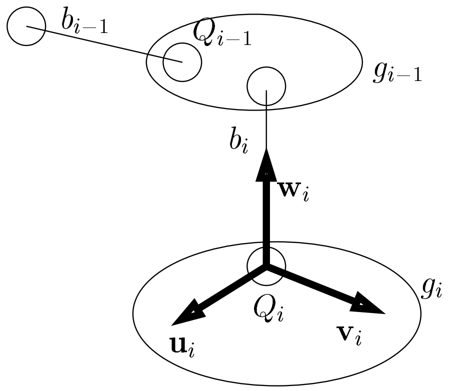


Figure 6: Local frame $F_i = \{Q_i; \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i\}$ at atomgroup $g_i$.

$\mathbf{w}_i$ = the unit vector along bond $b_i$ pointing toward $g_{i-1}$;

$\mathbf{u}_i$ = an arbitrary unit vector perpendicular to $\mathbf{w}_i$;

$\mathbf{v}_i = \mathbf{w}_i \times \mathbf{u}_i$.

The root group does not have any parent group, so we will translate the global frame to any atom in the root group — the local frame at the root group is the same as the global frame except for the origin.

## 4.2 Relational Matrices

Suppose 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 $F_{i-1}$. That is,

$$P - Q_{i-1} = \begin{pmatrix} \mathbf{u}_{i-1} & \mathbf{v}_{i-1} & \mathbf{w}_{i-1} \end{pmatrix} \begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \end{pmatrix},$$

$$P - Q_i = \begin{pmatrix} \mathbf{u}_i & \mathbf{v}_i & \mathbf{w}_i \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}.$$

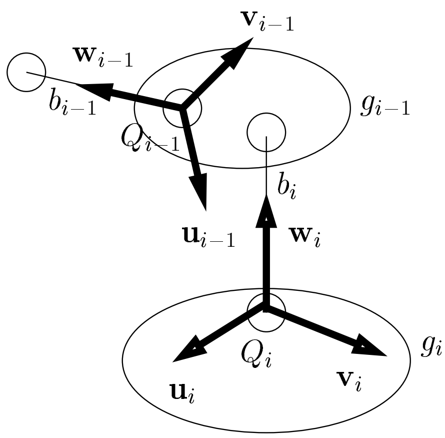


Figure 7: Frame $F_i = \{Q_i; \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i\}$ at group $g_i$ and frame $F_{i-1} = \{Q_{i-1}; \mathbf{u}_{i-1}, \mathbf{v}_{i-1}, \mathbf{v}_{i-1}\}$ at its parent group $g_{i-1}$.

It follows that

$$\begin{pmatrix} \mathbf{u}_{i-1} & \mathbf{v}_{i-1} & \mathbf{w}_{i-1} \end{pmatrix} \begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \end{pmatrix} + Q_{i-1} = \begin{pmatrix} \mathbf{u}_i & \mathbf{v}_i & \mathbf{w}_i \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + Q_i,$$

and

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

Thus

$$\begin{aligned} \begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \end{pmatrix} &= \begin{pmatrix} \mathbf{u}_{i-1}^T \\ \mathbf{v}_{i-1}^T \\ \mathbf{w}_{i-1}^T \end{pmatrix} \begin{pmatrix} \mathbf{u}_i & \mathbf{v}_i & \mathbf{w}_i \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \begin{pmatrix} \mathbf{u}_{i-1}^T \\ \mathbf{v}_{i-1}^T \\ \mathbf{w}_{i-1}^T \end{pmatrix} (Q_i - Q_{i-1}) \\ &= \begin{pmatrix} \mathbf{u}_{i-1} \cdot \mathbf{u}_i & \mathbf{u}_{i-1} \cdot \mathbf{v}_i & \mathbf{u}_{i-1} \cdot \mathbf{w}_i \\ \mathbf{v}_{i-1} \cdot \mathbf{u}_i & \mathbf{v}_{i-1} \cdot \mathbf{v}_i & \mathbf{v}_{i-1} \cdot \mathbf{w}_i \\ \mathbf{w}_{i-1} \cdot \mathbf{u}_i & \mathbf{w}_{i-1} \cdot \mathbf{v}_i & \mathbf{w}_{i-1} \cdot \mathbf{w}_i \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \begin{pmatrix} \mathbf{u}_{i-1} \cdot (Q_i - Q_{i-1}) \\ \mathbf{v}_{i-1} \cdot (Q_i - Q_{i-1}) \\ \mathbf{w}_{i-1} \cdot (Q_i - Q_{i-1}) \end{pmatrix} \end{aligned} \quad (8)$$

When bond $b_i$ rotates by an angle $\theta_i$, frame $F_i$ also rotate by angle $\theta_i$ (around $\mathbf{w}_i$ axis). Let the original frame (before $b_i$ rotates) be $F_i' = \{Q_i; \mathbf{u}_i', \mathbf{v}_i', \mathbf{w}_i\}$ (see Figure 8). By Equation (8), if a point $P$ has local coordinates $(x_i', y_i', z_i')^T$ in frame $F_i'$, then the coordinates of $P$ in $F_{i-1}$ are

$$\begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_{i-1} \cdot \mathbf{u}_i' & \mathbf{u}_{i-1} \cdot \mathbf{v}_i' & \mathbf{u}_{i-1} \cdot \mathbf{w}_i \\ \mathbf{v}_{i-1} \cdot \mathbf{u}_i' & \mathbf{v}_{i-1} \cdot \mathbf{v}_i' & \mathbf{v}_{i-1} \cdot \mathbf{w}_i \\ \mathbf{w}_{i-1} \cdot \mathbf{u}_i' & \mathbf{w}_{i-1} \cdot \mathbf{v}_i' & \mathbf{w}_{i-1} \cdot \mathbf{w}_i \end{pmatrix} \begin{pmatrix} x_i' \\ y_i' \\ z_i' \end{pmatrix} + \begin{pmatrix} \mathbf{u}_{i-1} \cdot (Q_i - Q_{i-1}) \\ \mathbf{v}_{i-1} \cdot (Q_i - Q_{i-1}) \\ \mathbf{w}_{i-1} \cdot (Q_i - Q_{i-1}) \end{pmatrix}. \quad (9)$$

It is easy to determine the relation between the frames $F_i$ and $F_i'$. In fact,

$$\begin{pmatrix} \mathbf{u}_i & \mathbf{v}_i & \mathbf{w}_i \end{pmatrix} = \begin{pmatrix} \mathbf{u}_i' & \mathbf{v}_i' & \mathbf{w}_i \end{pmatrix} \begin{pmatrix} c\theta_i & -s\theta_i & 0 \\ s\theta_i & c\theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

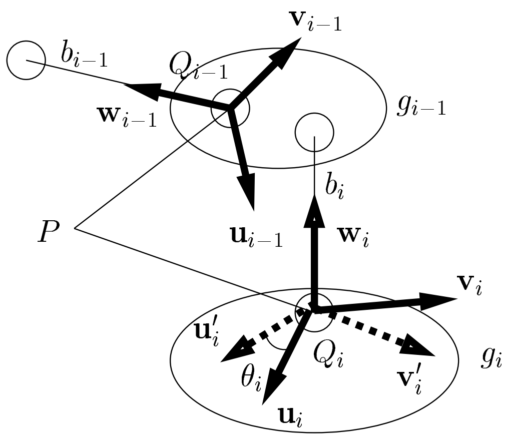


Figure 8: Local frame $F_i = \{Q_i; \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i\}$ at atomgroup $g_i$; local frame $F'_i = \{Q_i; \mathbf{u}'_i, \mathbf{v}'_i, \mathbf{w}_i\}$ which is frame $F_i$ before bond $b_i$ rotates by $\theta_i$; local frame $F_{i-1} = \{Q_{i-1}; \mathbf{u}_{i-1}, \mathbf{v}_{i-1}, \mathbf{w}_{i-1}\}$.

Therefore, for any point $P$ with coordinates $(x_i, y_i, z_i)^T$ in $F_i$, and coordinates $(x'_i, y'_i, z'_i)^T$ in $F'_i$,

$$\begin{aligned} P - Q_i &= \begin{pmatrix} \mathbf{u}'_i & \mathbf{v}'_i & \mathbf{w}_i \end{pmatrix} \begin{pmatrix} x'_i \\ y'_i \\ z'_i \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{u}_i & \mathbf{v}_i & \mathbf{w}_i \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{u}'_i & \mathbf{v}'_i & \mathbf{w}_i \end{pmatrix} \begin{pmatrix} c\theta_i & -s\theta_i & 0 \\ s\theta_i & c\theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}. \end{aligned}$$

Thus,

$$\begin{pmatrix} x'_i \\ y'_i \\ z'_i \end{pmatrix} = \begin{pmatrix} c\theta_i & -s\theta_i & 0 \\ s\theta_i & c\theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}. \tag{10}$$

It follows from Equations (9), (10) that the coordinates $(x_i, y_i, z_i)^T$ of a point $P$ in frames $F_i$ and the coordinates $(x_{i-1}, y_{i-1}, z_{i-1})^T$ in $F_{i-1}$ are related by

$$\begin{aligned} \begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \end{pmatrix} &= \begin{pmatrix} \mathbf{u}_{i-1} \cdot \mathbf{u}'_i & \mathbf{u}_{i-1} \cdot \mathbf{v}'_i & \mathbf{u}_{i-1} \cdot \mathbf{w}_i \\ \mathbf{v}_{i-1} \cdot \mathbf{u}'_i & \mathbf{v}_{i-1} \cdot \mathbf{v}'_i & \mathbf{v}_{i-1} \cdot \mathbf{w}_i \\ \mathbf{w}_{i-1} \cdot \mathbf{u}'_i & \mathbf{w}_{i-1} \cdot \mathbf{v}'_i & \mathbf{w}_{i-1} \cdot \mathbf{w}_i \end{pmatrix} \begin{pmatrix} c\theta_i & -s\theta_i & 0 \\ s\theta_i & c\theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \\ &\quad + \begin{pmatrix} \mathbf{u}_{i-1} \cdot (Q_i - Q_{i-1}) \\ \mathbf{v}_{i-1} \cdot (Q_i - Q_{i-1}) \\ \mathbf{w}_{i-1} \cdot (Q_i - Q_{i-1}) \end{pmatrix}. \end{aligned} \tag{11}$$

Equation (11) relates the coordinates of $P$ in frame $F_i$ and frame $F_{i-1}$. Note that the dot products in Equation (11) 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 group $g_i$, 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

$$\begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \\ 1 \end{pmatrix} = Rel_i \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix}, \tag{12}$$

where $Rel_i$ is

$$\begin{pmatrix} \mathbf{u}_{i-1} \cdot \mathbf{u}'_i & \mathbf{u}_{i-1} \cdot \mathbf{v}'_i & \mathbf{u}_{i-1} \cdot \mathbf{w}_i & \mathbf{u}_{i-1} \cdot (Q_i - Q_{i-1}) \\ \mathbf{v}_{i-1} \cdot \mathbf{u}'_i & \mathbf{v}_{i-1} \cdot \mathbf{v}'_i & \mathbf{v}_{i-1} \cdot \mathbf{w}_i & \mathbf{v}_{i-1} \cdot (Q_i - Q_{i-1}) \\ \mathbf{w}_{i-1} \cdot \mathbf{u}'_i & \mathbf{w}_{i-1} \cdot \mathbf{v}'_i & \mathbf{w}_{i-1} \cdot \mathbf{w}_i & \mathbf{w}_{i-1} \cdot (Q_i - Q_{i-1}) \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c\theta_i & -s\theta_i & 0 & 0 \\ s\theta_i & c\theta_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{13}$$

## 4.3 Updating the Atom Positions

Suppose $g_i, g_{i-1}, \cdots, g_0$ 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. Let $F_j$ be the frame attached to group $g_j$, $0 \le j \le i$. Below we compute the position of an atom $A$ (in group $g_i$) whose local coordinates are $(x_i, y_i, z_i)$ in $F_i$.

By Equation (12), the coordinates of atom $A$ in frame $F_{i-1}$ are

$$\begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \\ 1 \end{pmatrix} = Rel_i \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix}.$$

Similarly, the coordinates of $A$ in frame $F_{i-2}$ are

$$\begin{pmatrix} x_{i-2} \\ y_{i-2} \\ z_{i-2} \\ 1 \end{pmatrix} = Rel_{i-1} \begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \\ 1 \end{pmatrix} = Rel_{i-1} \cdot Rel_i \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix}.$$

Continuing in this fashion, the coordinates of atom $A$ in the frame at the root group are

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = Rel_1 \cdots Rel_i \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix}.$$

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 $A$ (of atomgroup $g_i$) in the global frame are

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = Tr(Q_0) \cdot E(\alpha, \beta, \gamma) \cdot Rel_1 \cdots Rel_i \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix}. \tag{14}$$

The atomgroup local frames method works with atomgroups. All atoms within an atomgroup share the same rotation matrix. Thus the number of rotation matrices (relational 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 of 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 all 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 such as collision checking. Therefore the atomgroup local frames method provides lazy evaluation for atom positions.

Comparing with the simple local frames method, the atomgroup local frames method attaches only one local frame to a rotatable bond, no matter how many child bonds there exist. The relations between the parent-child frames are easily determined by Equation (13). In the product of two matrices, the first matrix remain constant when any bond of the molecule rotates, so it can be saved when the local frames are constructed. The second matrix is very simple.

Updating the positions of the atoms in the atomgroup local frames method becomes easier than in the method of simple local frames method. In fact, all atoms in atomgroup $g_i$ share the same rotation matrix $Tr(Q_0) \cdot E(\alpha, \beta, \gamma) \cdot Rel_1 \cdots Rel_i$. The position of every atom in $g_i$ is updated by multiplying the rotation matrix to it local coordinates in frame $F_i$ — these local coordinates remain fixed. Atomgroup $g_i$ inherits the rotation matrix from $g_{i-1}$ and obtain its own rotation matrix by one more matrix multiplication: $(Tr(Q_0) \cdot E(\alpha, \beta, \gamma) \cdot Rel_1 \cdots Rel_{i-1}) \cdot Rel_i$. Therefore, every atomgroup obtains its rotation matrix by one matrix multiplication.

# 5 Comparison

In this section, we give a quantitative comparison between the three methods for maintaining 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 $Rot(b_i, \theta_i)$ from Equation (1): 24 multiplications to compute the middle matrix; 9 multiplications for multiplying $T^{-1}$; 3 multiplications, 3 divisions, and 1 square root operation to compute the unit vector $(v_x, v_y, v_z)^T$. $Rot(b_i, \theta_i)$ is then multiplied with the rotation matrix inherited from the parent atom, and this requires 36 multiplications — a $3 \times 3$ matrix times a $3 \times 4$ matrix. If bond $b_i$ does not rotate at all, then the matrix $Rot(b_i, \theta_i)$ is identity; atom $Q_i$ just pass 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 simple local frames method, depending on whether a bond has one or more child bonds, each atom needs to compute one or two matrices $Rel_i$ (Equation (3)) or $Rel_{i-1}, Rel_i$ [c.f. the end of Section 3]. Thus, it takes 6 or 12 multiplications to obtain $Rel_i$ or $Rel_{i-1}, Rel_i$. Each atom needs one or two matrix multiplications to get its rotation matrix which requires 30 or 60 scalar multiplications (each $Rel_i$ 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 the all atom positions. If we assume that half of the bonds have more than one child 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 $Rel_i$ (Equation (13)) 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 rotation scheme, simple local frame method, and the atomgroup local frame method is $102n_{rb}$, $162n_{rb}$, $75n_{rb}$ respectively.

We run the algorithms to two molecules and get the following results:

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

Table 1: Normalized time of running the three methods on two molecules.

The atomgroup local frames method is 33% and 28% faster than the simple rotations scheme, and 156% and 166% faster than the simple local frames method. The run time is measured for 10,000 rounds of random rotation and updates of all atom positions; lazy evaluation is not performed in the atomgroup local frames method yet. The two testing molecules are shown as in Figure 9.

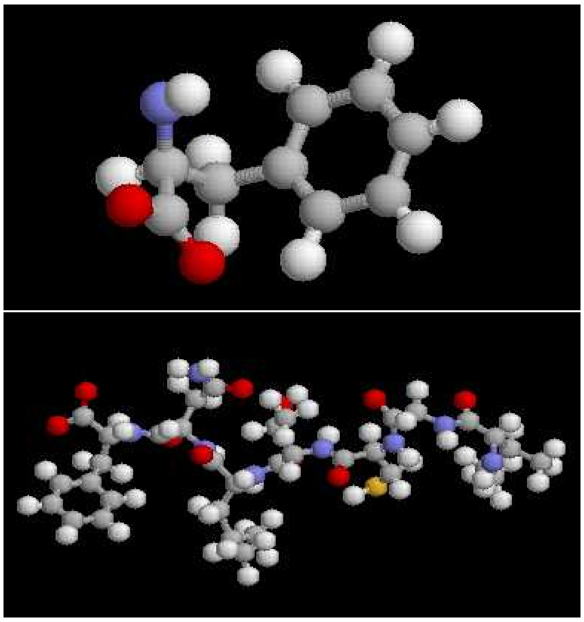

Figure 9: Two testing molecules.

We also have the following observations:

(i) From Equation (2), if bond $b_{i-1}$ does not rotate hence $Rot(b_{i-1}, \theta_{i-1})$ is the identity matrix, then the rotation matrix of atom $Q_i$ is the same as its parent atom. Thus, atom $Q_i$ does not need a matrix multiplication to obtain its rotation matrix. Therefore, in the simple rotations scheme, the number of rotation matrix updates equals the number of rotatable bonds. However, in the simple local frames method, atom positions are updated by Equation (7). Even if a bond $b_i$ does not rotate at all, the relational matrix $Rel_i$ may not necessarily be the identity matrix. Thus a matrix multiplication is still needed to pass the rotation matrices even the bond does not rotate. Therefore, the number of rotation matrix updates equals the number of (all) bonds. This explains why the simple local frames method is much slower than the simple rotation scheme.

(ii) From the above observation, it is easy to see that by grouping atoms into atomgroups, the simple rotations scheme does not improve its performance. We implemented the simple rotations scheme to the atomgroups, and the results verified this observation.

(iii) The simple rotations scheme can be modified to prevent the numeric errors from being accumulated upon the rotations around the bonds. This can be done by changing the meaning of the matrix $Rot(b_i, \theta_i)$ in Equation (1): instead of recording the torsional angle change each round, $\theta_i$ records the sum of all torsional angle changes of bond $b_i$. Then the position of any atom is updated by applying a sequence of $Rot(b_i, \theta_i)$'s to its original position. Thus numeric errors from previous rotations around a bond will not be carried to later rotations around the same bond. However, there is a tradeoff from this change. Now even bond $b_i$ does not have torsional angle change for the current round of rotation, $\theta_i$, the sum of all torsional angle changes of $b_i$, may not necessarily be zero. Hence, $Rot(b_i, \theta_i)$ may not necessarily be the identity matrix. Therefore, any atom needs a matrix multiplication to obtain its rotation matrix. Thus the number of matrix updates equals the number of total bonds instead of the number of rotatable bonds. That is, the computational complexity increases by elimination of the numeric error accumulation.

(iv) The statistics for the atomgroup local frame method do not include the lazy evaluation yet. In many applications such ligand-receptor docking, people are interested in the final positions of a small set of key atoms. The positions of other atoms do not matter as long as there is no collision. Therefore, the positions of the key atoms can be computed first. If they are in their target positions, the positions of non-key atoms will be computed and collision checking is performed. Otherwise, there is no need to compute the positions of the non-key atoms as the conformation will not be acceptable any way. The atomgroup local frame method provides the lazy evaluation mechanism: the position of any atom is not computed until it is needed. The simple rotation scheme does not have this ability since the positions of children atoms depend on the positions of ancestor atoms. In a process like the kinematics or energy minimization, the lazy evaluation will greatly reduce the maintenance cost since a large number of conformations will be generated and computed.

# 6 Conclusion

Efficiently maintaining the molecular conformations is important for molecular modeling and protein engineering. This paper reviews the widely used simple rotations scheme, simple local frames method, and introduces a new atomgroup local frames method.

The simple rotations scheme is conceptually simple and easy to implement. The position of every atom is updated by applying a sequence of rotations. However, an atom can only be updated when all its ancestor atoms have been updated. Thus some book-keeping is needed to make sure every atom is correctly updated. Another drawback of the simple rotations scheme is that the numeric errors accumulate as more and more rotations are performed around the bonds.

The simple local frames method attaches local frames to the bonds, and the relational matrices between the parent-child frames are determined. Then the position of each atom is computed by chaining a sequence of the relation matrices. The simple local frames method does not have to keep a record of the positions of all the ancestor atoms to update the position of a particular atom. Also, the numeric errors do not accumulate in the simple local frames method. The drawback of the simple local frames method is that it needs to attach multiple local frames to a bond if the bond has more than one child. Thus the rotation matrix can not be completely inherited from parent bond to the children. The simple local frames method also works only with individual atoms instead of atomgroups.

The atomgroup local frames method not only inherits the advantages of both the simple rotations scheme and simple local frames method, but also discards the disadvantages of them. The atomgroup local frames method works with atomgroups, and constructs only one local frame for each atomgroup. The rotation matrix is completely inherited by a child group from its parent group, and updated by a single matrix multiplication. To update the position of any atom, the atomgroup local frames method does not have to record the positions of its ancestor atoms or atomgroups. Moreover, the numeric errors do not accumulate in the atomgroup local frames method. Thus updating the molecular conformations becomes easier than in the simple rotations scheme or the simple local frames method.

The differences of the three methods are summarized in the table below.

| | book-keeping | atom-groups | rotation matrices | error accumulation |
|---|---|---|---|---|
| simple rotations | needed | no | completely inherited | yes |
| simple local frames | not needed | no | not completely inherited | no |
| atomgroup local frames | not needed | yes | completely inherited | no |

Table 2: Summary of three methods of maintaining molecular conformations.

### Acknowledgement

Work on this paper by Ming Zhang and Lydia Kavraki is supported in part by ATP 003604-0120-1999, NSF IRI-970228, NSF CISE SA1728-21122N, and a Sloan Fellowship awarded to Lydia Kavraki. The authors would like to thank Miguel Teodoro, Brian Chen and Andrew Ladd for useful discussions.

## References

[1] D. E. Clark, G. Jones, P. Willett, P. W. Kenny, R. C. Glen. Pharmacophoric Pattern Matching in Files of Three-Dimensional Chemical Structures: Comparison of Conformational Searching Algorithms for Flexible Searching. *Journal of Chemical Information and Computer Sciences*, 34:197–206, 1994.

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

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

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

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

[6] D. R. Henry, A. G. Ozkabak. In Encyclopedia of Computational Chemistry. Schleyer, P. v. R., Ed., Wiley, New York, 1998.

[7] G. Jones, P. Willett, R. C. Glen. A Genetic Algorithm for flexible Molecular Overlay and Pharmacophore Detection. *Journal of Computer-Aided Molecular Design*, Vol. 9, 532-549, 1995.

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

[9] M. Lipton, W. Still. The Multiple Minimum Problem in Molecular Modeling. Tree Searching Internal Coordinate Conformational Space. *Journal of Computational Chemistry*, Vol. 9, 343–355, 1988.

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