

CS 273

April 10th, 2005

Lecturer: Jean-Claude Latombe

Scribe: Kyle Bruck

Lecture 2: Protein Geometric and Kinematic Models

1.0 Introduction and Definitions

What is Kinematics?

Kinematics is the study of movement independent of the forces that cause them.

Protein is a sequence of amino acids attached by peptide bonds. The peptide bond initially creates a chain of repeated atoms. Each **amino acid** has a different R side chain. There are 20 amino acids.

The protein will fold from a chain into a compacted structure that has some movement but the structure is fairly unique.

2.0 Role and Issues of Geometric/Kinematic models

2.1 Role of Models

1. Represent the possible shape of proteins (Compare and classify shapes or find motifs)
2. Answer proximity queries such as which atoms are close to another given atom? (Utilizes the computation of energy)
3. Compute the surface area of proteins (Useful for how they interact with solvents)
4. Find shape features such as cavities (Useful for binding ligand binding sites)

2.2 Issues with Models

1. Large number of atoms (Combinatorial problems)
2. Large number of degrees of freedom (Large-dimensional conformation space_
3. Need to efficiently update information during simulation (surface area, proximity among atoms):
 - a. What is the position of every atom in some given coordinate system?
 - b. Which atoms intersect a given atom?
 - c. What atoms are within some distance range from another one?
4. Complex metric in conformational space
5. Issues with Shape matching

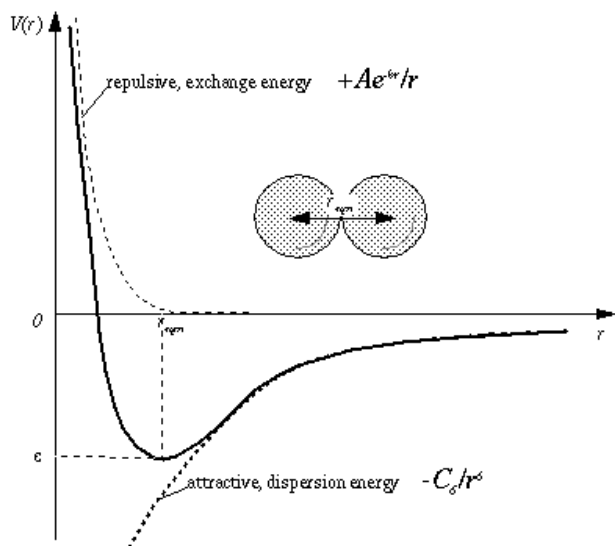
3.0 Geometric Models of Bio Molecules

Geometric Models are generally very easy. Model atoms as spheres called the Hard-sphere model. Programmer can choose the radius of the sphere and most people use the van der Waals radii which creates a Van der Waals surface around the hard spheres.

3.1 Van der Waals Potential

The surface is determined by the Van der Waals potential. The two atoms will get closer and closer and the potential will decrease. As the distance goes to 0, the repulsive energy goes to infinite. As the distance gets larger, the attractive energy gets closer to 0. The Van der Waals potential is the sum of these two equations.

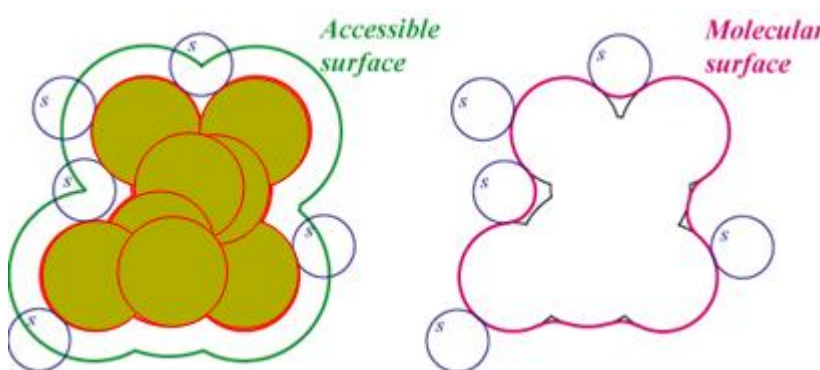
The important point is the tangent point which creates a trough called epsilon. The tangent point is the most stable point between the two atoms. There may be overlap because of a global interaction.



3.2 Other Models being used

Solvent-Accessible surface is defined by a solvent like a water molecule that is modeled by a sphere and roll that sphere around the hard surface of the molecule. The center of the sphere traces the accessible surface. This defines where the solvent has access to.

Molecular Surface is similar to the accessible surface. It defines the minimum surface that the molecule actually touches. It depends on the radius of the sphere S .



3.3 Computation of Hard-Sphere Surface

(Grid Method [Halperin and Shelton, 97])

3.3.1 Properties of Computation

- Each sphere intersects $O(1)$ spheres
- Computing each atom's contribution to molecular surface takes $O(1)$ time

- Computation of molecular surface takes $\Theta(n)$ time

3.3.2 Computation Algorithm

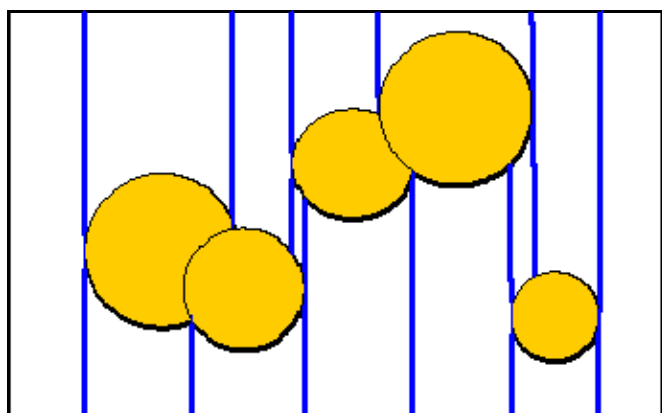
- Each sphere intersects a constant number of spheres independent of the size of the protein because you cannot have all of the atoms intersect at the same place. So there is a limited number of overlap that can occur.
 - Kdtree versus grid. Kdtree will not allow for linear time like grid cannot and thus is not better.
- For any given atom, take the center and see what cube that the center lies in. Because the cell size is roughly the size of the Van der Waals radii, have to examine all cells in the area as well as all atoms in those cells.
- Compute each atom's contribution to molecular surface takes constant time. Create a 3d grid that would represent where in space an atom will be located. Each cell is a cube in 3d that has a size slightly bigger than the Van Der Waals diameter. If one wanted to know what atoms overlap with a particular atom, look in that cell and see what other atoms are in that cell or around that cell and calculate the Van Der Waals potential around them to see where the overlap is.

4.0 Removing Surface of Overlap.

Trapezoidal Decomposition is by the process by which the area of the white space on 2d plane is calculated around obstacles. Sweep a line across the space and create a line where it hits a vertex. Then calculate the area of all of the trapezoids created.

For circles create a line at each intersection or tangency point. Then compute the area of the arched trapezoids.

For spheres create a line that goes from one pole to another and sweep the line across the phase. This is similar if one took the sphere surface and flattened it into a 2d rectangle and removed circular regions as in the original description of the process above.



5.0 Visualization Models

Stick (bond) model is the model where bonds are sticks

Small-sphere model is the model where each atom is a small sphere connected by a stick representing the bond

Ribbon Model is the model where motifs such as alpha helices and beta sheets are represented as ribbons rather than individual atoms.

6.0 Kinematics of Biomolecules

What is Kinematics? Study of mobility independent of the cause of the motion. Specifically of a molecule independent of the energy field. Dynamics has to do with the cause such as inertia, energy, and the other forces applied to create that motion.

6.1 Kinematic Models

Make assumption of the structure of the molecule such that some movement within the molecule becomes intrinsic to the model and thus do not have to be calculated.

6.1.1 Atomistic Model

The position of each atom is defined by its coordinates in 3d space. P atoms have 3p parameters. But this does not take into account the bonds between the atoms and thus the length between the atoms have to be enforced in another way. Peptide bonds make proteins into long kinematic chain, however the atomistic model does not encode this kinematic structure. The torsional angles can vary but the angles of the bond between the atoms cannot.

6.1.2 Linkage model

The kinematics is defined by internal coordinates (bond lengths and angles, and torsional angles around bonds). Encode a data structure that inherently enforces the constraints of the bond and the angles. This is better than the atomistic model where the data structure does not inherently enforce this constraint.

Issues

- More difficult to update the position of each atom in the world coordinate system.
- Determine which pairs of atoms are within some given distance (topological proximity along chain → spatial proximity but the reverse is not true)
- Typical model consists of atoms with angles and magnitudes to the next atom.

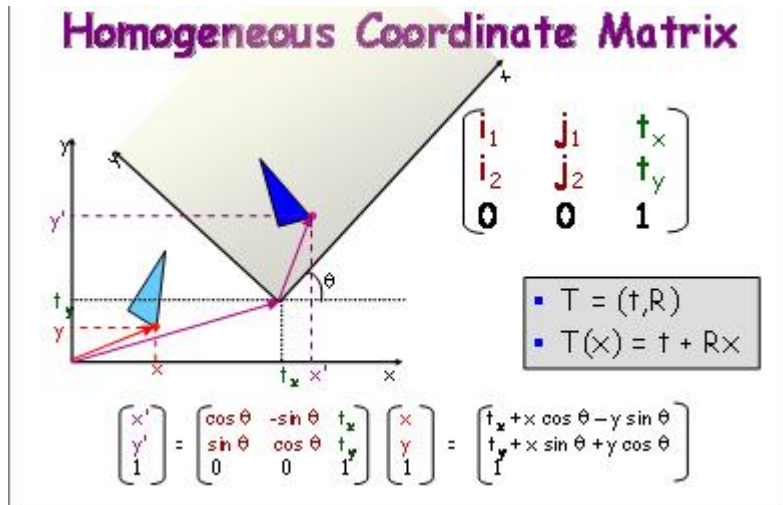
7.0 Rigid-Body transform

The rigid-body transform is a mapping of every point in an object to a new coordinate the still represents the original object but transforms the rotation and translation

2d Case takes a plane and shifts the plane into place and then rotates the plane until in place.

A rotation matrix in 2d is defined as
$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

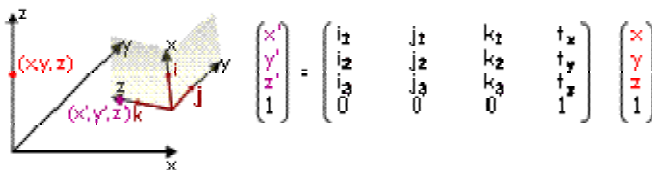
This rotation matrix can be expanded to take advantage of the Homogeneous Coordinate Matrix. This matrix maps all coordinates in one plane to that of another plane with regards to an origin point.



The 3d Homogeneous Coordinate Matrix is an expansion of the original 2d version to include the extra points. However the 3d Matrix must also conform to the constraints:

- $i_1^2 + i_2^2 + i_3^2 = 1$
- $i_1 j_1 + i_2 j_2 + i_3 j_3 = 0$
- $\det(R) = +1$
- $R^{-1} = R^T$

Homogeneous Coordinate Matrix in 3-D

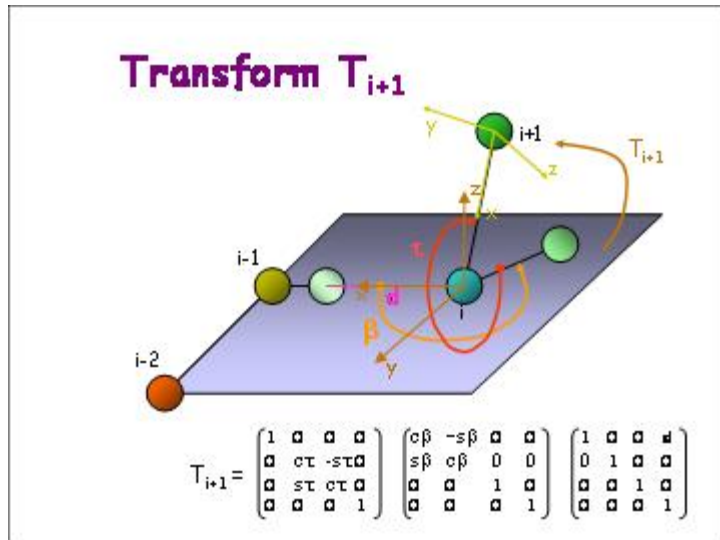


Composition of two transforms represented by matrices T_1 and T_2 : $T_2 \times T_1$

Potential problem with homogeneous coordinate matrix? There are 9 parameters. If take the two constraints could reduce the matrix down to 3 parameters.

8.0 Building a serial Linkage Model

1. Build the assembly of the first 3 atoms:
 - a. Place 1st atom anywhere in space.
 - b. Place 2nd atom anywhere at bond length.
 - c. Place the 3rd atom anywhere at bond length with the certain bond angle.
 - d. Need 3 parameters. Position of 1st and 2nd atom, angle of 3rd atom that will put it on a plane.
 - e. Define a coordinate frame that is relative to the 3rd atom where the origin is at the center of the 3rd atom and the x axis is along the bond angle.
2. Add each other atom
 - a. Need 3 parameters



The Update between atoms follows:

Given:

$$T_k^{(i)} = T_k \dots T_{i+2} T_{i+1}$$

Then the atom j between i and k is represented as

$$T_k^{(i)} = T_j^{(i)} T_{j+1} T_k^{(j+1)}$$

Thus A parameter between j and j+1 is changed

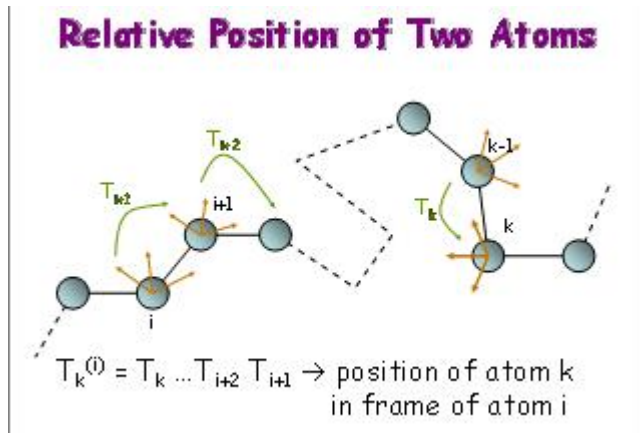
$$T_{j+1} \rightarrow \mathcal{T}_{j+1}$$

Yielding

$$T_k^{(i)} \rightarrow \mathcal{T}_k^{(i)} = T_j^{(i)} \mathcal{T}_{j+1} T_k^{(j+1)}$$

8.1 Relative Position of Two Atoms

The relative position of two atoms can be calculated as by progressing through all of the various position matrices to get to the final atom. $T_k^{(i)} = T_k \dots T_{i+2} \cdot T_{i+1} \rightarrow$ Position of atom k in frame of atom i.



9.0 Tree-Shaped Linkage

Root group of 3 atoms. It follows the same basic technique of the peptide chain above but all of the positions are relative to the starting 3 atom group.

P atoms $\rightarrow$ 3p-6 parameters because position all of the atoms with respect to the first 3 atoms.