Conformations of Proteins on Lattice Models

Jiangbo Miao
Natalie Kantz
Lattice Model

The lattice model offers a discrete space that limits the infinite number of protein conformations to the lattice space.

The flexibility of the model is defined by the rotational flexibility offered by the number of neighboring nodes.

The fit of the protein backbone to the lattice nodes, measured by the RMS value is determined by the flexibility of the model.
Three Lattice Models

- **Simple Cubic (SC)**
  - bond length: 3.8 Å
  - bond angles: 90, 180
  - torsional angles: 0°, 90°, 180°, 270°

- **Body-Centered Cubic (BCC):**
  - bond lengths: 3.8 Å, 3.3 Å
  - bond angles: additional 4
  - torsional angles: additional 8
Three Lattice Models

- **Face-Centered Cubic (FCC):**
  
  - bond lengths: 3.8 Å, 2.68 Å, 4.65 Å
  - bond angles: additional 4
  - torsional angles: additional 16
Simple Cubic Lattice

Figure 1:

- **neighbors**
- **Lattice point occupied by \( C_\alpha \)**
Face Centered Cubic Lattice
Extract $C_\alpha$ from PDB file

Create Lattice

Fit $C_\alpha$'s to lattice points

Find Best Fit for the first $C_\alpha$

Fit Remaining $C_\alpha$'s

Calculate new coordinates according to a 3-way rotation

Output of $C_\alpha$ coordinates at the lowest RMS value
Simple Cubic Latice Model Implemented with a 2-way rotation

**User Input:**
1. PDB file
2. Number of amino acids
3. Number of iterations for the phi
4. Number of iterations for the theta

**Some Important Program Variables:**
6 arrays of the C
1. Cartesian coordinates
2. Polar coordinates
3. Rotated Polar Coordinates
4. Rotated Cartesian Coordinates
5. Shifted Cartesian Coordinates
6. Type of Amino Acid
Algorithm:

**Step 1: Lattice Setup**

1. Shift Cartesian to get positive values
2. Convert the Shifted Cartesian to Polar Coordinates
3. Initialize the lattice by creating a 4-D array of a given size with indices values set to zero

**Step 2: Lattice Fit**

** as each residue is placed on the lattice the RMS value is calculated

1. Fit the first residue
2. Fit the remaining residues by checking a neighbor one less than the current node value or one more than the current node value according to the bounds of the lattice. Then checking to see if these neighbors are occupied
Algorithm:

Step 3: Rotations

1. For each Phi angle the RMS value given by all of the theta rotations are calculated from the following loop
   a. polar coordinates are rotated with values incrementing values of phi and theta
   b. polar coordinates are converted to cartesian coordinates
   c. Cartesian coordinates are shifted
   d. lattice is reset to empty or zero
   e. Step 2 : Lattice Fit is repeated
Output

1. Minimum RMS value
2. The Phi and Theta angles for the minimum RMS value
3. The names of the residues with their cartesian coordinates for the minimum RMS value.
## Results:

<table>
<thead>
<tr>
<th>Protein</th>
<th># Residues</th>
<th>SC 3-way</th>
<th>SC 2-way</th>
<th>FCC 3-way</th>
<th>FCC 2-way</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ppt</td>
<td>36</td>
<td>1.869</td>
<td>2.460</td>
<td>1.046</td>
<td>1.136</td>
</tr>
<tr>
<td>1crn</td>
<td>46</td>
<td>2.250</td>
<td>2.950</td>
<td>1.100</td>
<td>1.244</td>
</tr>
<tr>
<td>8pit</td>
<td>58</td>
<td>2.176</td>
<td>2.914</td>
<td>1.162</td>
<td>1.312</td>
</tr>
<tr>
<td>1nxb</td>
<td>62</td>
<td>2.525</td>
<td>3.540</td>
<td>1.134</td>
<td>1.306</td>
</tr>
<tr>
<td>1noa</td>
<td>113</td>
<td>3.037</td>
<td>3.531</td>
<td>1.244</td>
<td>1.288</td>
</tr>
<tr>
<td>1rat</td>
<td>124</td>
<td>2.956</td>
<td>3.126</td>
<td>1.238</td>
<td>1.313</td>
</tr>
<tr>
<td>1mba</td>
<td>147</td>
<td>2.596</td>
<td>2.850</td>
<td>1.262</td>
<td>1.347</td>
</tr>
<tr>
<td>2ptn</td>
<td>223</td>
<td>3.232</td>
<td>3.785</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5cpa</td>
<td>307</td>
<td>3.116</td>
<td>3.324</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3app</td>
<td>323</td>
<td>3.297</td>
<td>3.857</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2apr</td>
<td>325</td>
<td>3.370</td>
<td>3.531</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>2.766</strong></td>
<td><strong>3.261</strong></td>
<td><strong>1.169</strong></td>
<td><strong>1.278</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: RMS values of SC and FCC models with 2-way and 3-way rotations with step size between 100-300 for the first round and between 30-50 for the second round and 10-20 for the third round.
Results:

<table>
<thead>
<tr>
<th>Step Size</th>
<th>RMS</th>
<th>$\Phi$</th>
<th>$\Theta$</th>
<th>$\psi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50/ $\pi$</td>
<td>1.083990</td>
<td>1.389200</td>
<td>0.956930</td>
<td>2.605000</td>
</tr>
<tr>
<td>100/ $\pi$</td>
<td>1.081790</td>
<td>0.062832</td>
<td>1.162389</td>
<td>0.141372</td>
</tr>
<tr>
<td>150/ $\pi$</td>
<td>1.070178</td>
<td>0.519996</td>
<td>1.773785</td>
<td>2.049994</td>
</tr>
<tr>
<td>250/ $\pi$</td>
<td>1.070213</td>
<td>3.640729</td>
<td>1.772487</td>
<td>2.050015</td>
</tr>
</tbody>
</table>

Table 2: Minimum RMS values for $\Phi$, $\Theta$, $\psi$ for step sizes less than 300
Question

Finding Global Minimum:

• Newton’s Method will not work because it may result in a local minima

• A derivative-based method will not work because the Eularian rotations are cos and sin based.
As expected the lower RMS values of the FCC 3-way rotation model in Table 1 indicate a best fit of the C and the nodes on the lattice model. These figures demonstrate that larger number of neighbor nodes allows for greater flexibility in the rotational angles and a more precise placement of the actual protein backbone on the lattice nodes.

As indicated by Table 2, the location of the RMS value is determined by the step size of the exhaustive search. Smaller RMS values were found in smaller step size increments. The incremental value of the three angles determines the comprehensive level of the search.
Future Research

Using the conformations of from the FCC model to do percolation with the $C_\alpha$ of the proteins conformation acting as the lattice to define hydrophic clusters.
Future Research

Using the conformations of from the FCC model to do percolation with the C of the proteins conformation acting as the lattice to define hydrophic clusters.