A Tale of Two Helices

A study of α-helix conformations in three-dimensional space

Robert Fraser
Hypothesis

Properties of the three-dimensional configuration of a pair of α-helices may be predicted from the contact map corresponding to the pair.
What to expect

• A tour of protein molecules
• The $\alpha$-helix
• Protein structure prediction
• Packing of helices
• Prediction of packing
• Future work
The tour!!!

• Let’s get a feel for proteins

• Thanks to Google image search
Where do proteins come from?

Amino acid residues

- Two representations of the same α-helix
  - Left helix shows residues only
  - Right helix shows all backbone atoms
α-helices

- Structures within proteins
- Hydrogen bonding provides rigidity
- Very regular
Why α-helices?

• Protein Structure Prediction
  – The 3D structure of a protein is primarily determined by groups of secondary structures
  – The α-helix is the most common secondary structure
Protein Structure Prediction

• is hard.

• There are many different approaches

• Crystallization techniques are time-consuming

• Tryptych
  – Case based-reasoning approach to prediction
  – Uses different experts as advisors
  – Works from a contact map
The Goal

- 2D representation → 3D
- Toy example:

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>4</th>
<th>5</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>
The Goal

If we apply a threshold of 4.5 units, can we still determine the shape?
Protein Structure Prediction

- Build (or predict) contact map
- Predict the three-dimensional structure from the contact map
- \(N \times N\) matrix, \(N=\#\) of amino acids in the protein
- \(C(i,j) = \text{true if the distance from amino acid } i \text{ to } j \text{ is less than a threshold (ie. 7Å)}\)
Refining a contact map

- Contact map for entire protein
- Contact map for pair of α-helices
- Interface region of contact map
Hypothesis (again)

Properties of the three-dimensional configuration of a pair of α-helices may be predicted from the contact map corresponding to the pair.
Packing of helices

- A nice simple boolean property
- Found by rotating both helices to be axis-aligned, then comparing coordinates
Prediction of packing

Need a map comparison method
Comparing maps

- Comparing two maps is not straightforward
- Reflections are insignificant
Comparing maps

- Translations are (generally) insignificant
Prediction of packing

- Classify the maps based on contact locations

![Contact Map for 6 and 4](image1.png)
![Contact Map for 5 and 4](image2.png)
![Contact Map for 2 and 1](image3.png)

- central
- edge
- corner
Packing data by class

- No clustering to be done on the central and edge classes
- Reflections of each corner map were used to build the data set

<table>
<thead>
<tr>
<th>Contact Map Class</th>
<th>Total Instances</th>
<th>Packing</th>
<th>Non-packing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central</td>
<td>112</td>
<td>112 (100%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Edge</td>
<td>535</td>
<td>531 (99.3%)</td>
<td>4 (0.7%)</td>
</tr>
<tr>
<td>Corner</td>
<td>431</td>
<td>397 (92.1%)</td>
<td>34 (7.9%)</td>
</tr>
<tr>
<td>Doubled Corner</td>
<td>862</td>
<td>794 (92.1%)</td>
<td>68 (7.9%)</td>
</tr>
<tr>
<td>All Maps</td>
<td>1078</td>
<td>1040 (96.5%)</td>
<td>38 (3.5%)</td>
</tr>
</tbody>
</table>
Clustering Corner Maps

- Each corner map is transformed to a $15 \times 15$ map
- A map corresponds to a 225 character binary string
- Distance can be measured using cosine
### Clustering Results

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Packing</th>
<th>Non-Packing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>Mixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>83</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>Non-Packing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>14</td>
</tr>
</tbody>
</table>
Clustering Results

Non-packing

Mixed

Packing
Nearest Neighbour

• Closest neighbour by cosine distance

• Of the 862 contact maps in the data set, only 9 had a packing value different from the nearest neighbour

• Almost 99% accuracy!
Hypothesis (last time, I promise)

Properties of the three-dimensional configuration of a pair of α-helices may be predicted from the contact map corresponding to the pair.
Contributions (I)

• Hippy visualization tool
  – 1\textsuperscript{st} tool for helix pairs and contact maps
  – 1\textsuperscript{st} tool to visualize relationship between contacts in map and in three-dimensional space
Contributions (II)

• Angle and packing calculations
  – Developed a novel axis calculation technique
  – 1\textsuperscript{st} method to use contact information to determine local axis
  – Created a simple geometric approach to packing determination
Contributions (III)

• Prediction of helix pair properties
  – 1st to establish that contact maps cluster
  – Developed a novel contact map classification scheme for helix pairs
  – 1st to demonstrate that helix pair properties may be predicted from contact maps
Future Work (I)

- Hippy
  - Add motion
    - User interaction
    - Energy minimization
  - Add helices, allowing triplets or more
  - Use DSSP information for secondary structures
  - Allow the visualization and alignment of multiple pairs of helices
Future Work (II)

- Implement the packing advisor (done)
- Explore other properties of α-helices that could be predicted from contact maps
  - Interhelical angle
  - Interhelical distance
Thanks

• Kelly, family, friends
• Janice & James
• School of Computing
• NSERC, PRECARN IRIS, Queen’s

• Questions?
Extra Slides!

extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
extra extra extra extra extra extra extra extra
Amino Acids
Polymerization
Levels of Structure

Primary

Secondary

Tertiary

Quaternary
α-helix properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phi</td>
<td>-57</td>
</tr>
<tr>
<td>Psi</td>
<td>-47</td>
</tr>
<tr>
<td>Omega</td>
<td>180</td>
</tr>
<tr>
<td>Rotation about axis per residue</td>
<td>100</td>
</tr>
<tr>
<td>Residues per turn</td>
<td>3.6</td>
</tr>
<tr>
<td>Translation per residue</td>
<td>1.5</td>
</tr>
</tbody>
</table>
Tryptych
Chothia packing model
Preferred packing angles
4 classes of interhelical configurations
Chothia/Walther axis calculation
Hippy’s technical details

• Implemented in OpenGL
• Platform independent, open source
• Atoms coloured using C-P-K scheme
• Atomic radii given by United Atom Radius
• Opens any PDB file or one with similar data structure
Quaternions

\[ r' = \left( \cos^2 \frac{\theta}{2} - \sin^2 \frac{\theta}{2} \right) r - \sin \theta (n \times r) + 2 \sin^2 \frac{\theta}{2} \frac{\theta}{2(n \cdot r) n} \]

\[ r = (x, y, z) \]

\[ r' = (x', y', z') \]

\[ m \tan \frac{\theta}{2} (z + z') - n \tan \frac{\theta}{2} (y + y') = (x' - x) \]

\[ -i \tan \frac{\theta}{2} (z + z') + n \tan \frac{\theta}{2} (x + x') = (y' - y) \]

\[ i \tan \frac{\theta}{2} (y + y') - m \tan \frac{\theta}{2} (x + x') = (z' - z) \]
Chothia angle calculation

\[ t_k = b_k + S_k (e_k - b_k), \]

\[ d^2 = (t_k - t_j) \cdot (t_k - t_j) = (t_k - t_j)^2 \]

\[ d^2 = (b_k + S_k (e_k - b_k) - b_j - S_j (e_j - b_j))^2 \]

\[ t_k = b_k + S_k (e_k - b_k) \]

\[ t_j = b_j + S_j (e_j - b_j) \]

\[ S_k = 0 = \frac{(b_k - b_j - S_j (e_j - b_j)) \cdot (e_k - b_k)}{(e_k - b_k)^2} \]

\[ S_j = 0 = \frac{(b_k + S_k (e_k - b_k) - b_j) \cdot (e_j - b_j)}{(e_j - b_j)^2} \]
Algorithm 1 Iteratively finding the closest points

if \( t_k \) is outside helix then
    move \( t_k \) to the nearest endpoint of helix \( k \)
    recompute \( t_j \)
    if \( t_j \) is outside helix then
        move \( t_j \) to the nearest endpoint of helix \( j \)
        recompute \( t_k \)
        if \( t_k \) is outside helix then
            move \( t_k \) to the nearest endpoint of helix \( k \)
        end if
    end if
end if
Packing model

\[
\text{if } (d_a \times d_b) \cdot (b_0 - a_0) < 0 \\
\text{then the top vector is rotated clockwise} \\
\text{else the top vector is rotated counter-clockwise}
\]
Distance Metrics

\[ d_E = \sqrt{\sum_{i=1}^{n} (C_1(i) - C_2(i))^2} \]

\[ d_J = \frac{C_{11}}{C_{11} + C_{10} + C_{01}} \]

\[ d_{\cos} = \frac{C_1 \cdot C_2}{||C_1|| ||C_2||} \]
Forces equation

\[ \nu (r^N) = \sum_{\text{bonds}} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 + \sum_{\text{torsions}} \frac{V_n}{2} (1 + \cos (n\omega - \gamma)) \]

\[ + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left( 4\varepsilon_{ij} \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) + \frac{qq_{ij}}{4\pi\varepsilon_0 r_{ij}} \right) \] (2.1)

where

\( \nu (r^N) \) is the potential energy of the system
\( r \) is the position of the atoms
\( N \) is the total number of atoms
\( k \) is the stretching constant (experimentally determined)
\( l_i \) is the length of the bond
\( l_{i,0} \) is the ideal length of the bond
\( \theta_i \) is the bond angle
\( \theta_{i,0} \) is the ideal bond angle
\( V_n \) is the barrier height (experimentally determined)

\( n \) is the multiplicity (the number of minimum energy angles)
\( \gamma \) is the phase factor (at which angle the minimum energy exists)
\( \omega \) is the torsion angle
\( \varepsilon \) is the well depth (experimentally determined)
\( \sigma \) is the collision diameter (experimentally determined)
\( q \) is the charge of the atom
\( \varepsilon_0 \) is the permittivity of space