Protein Structure Analysis

http://binf.gmu.edu/vaisman/binf731/

Iosif Vaisman
2012

Secondary Structure Conformations

<table>
<thead>
<tr>
<th>Structure</th>
<th>φ</th>
<th>ψ</th>
</tr>
</thead>
<tbody>
<tr>
<td>α helix</td>
<td>-57</td>
<td>-47</td>
</tr>
<tr>
<td>α-L</td>
<td>57</td>
<td>47</td>
</tr>
<tr>
<td>3-10 helix</td>
<td>-49</td>
<td>-26</td>
</tr>
<tr>
<td>π helix</td>
<td>-57</td>
<td>-80</td>
</tr>
<tr>
<td>type II helix</td>
<td>-79</td>
<td>150</td>
</tr>
<tr>
<td>β-sheet parallel</td>
<td>119</td>
<td>113</td>
</tr>
<tr>
<td>β-sheet antiparallel</td>
<td>-139</td>
<td>135</td>
</tr>
</tbody>
</table>

Secondary Structure Assignment

- Sequential number, including chain breaks as extra residues
- Amino acid sequence in one letter code
- α-helix
- β-turn/helix
- β-sheet label
- β-sheet partner resnum
- Beta sheet label
- Solvent accessibility

DSSP  STRIDE  DEFINE

helix  strand  coil

Secondary Structure Assignment

<table>
<thead>
<tr>
<th>RESIDUE AA</th>
<th>STRUCTURE</th>
<th>DPS</th>
<th>RZ</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 47 50</td>
<td>K T Y</td>
<td>S</td>
<td>3</td>
<td>50</td>
</tr>
<tr>
<td>36 37 40</td>
<td>S T S</td>
<td>S</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>38 39 40</td>
<td>N S T</td>
<td>S</td>
<td>5</td>
<td>34</td>
</tr>
<tr>
<td>39 41 42</td>
<td>W E N</td>
<td>K</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>40 41 42</td>
<td>S T W</td>
<td>K</td>
<td>3</td>
<td>36</td>
</tr>
</tbody>
</table>

Adapted from Zvelebil, Baum, 2008

Adapted from Zvelebil, Rosa, 2008
Secondary Structure Assignment

<table>
<thead>
<tr>
<th>RESIDUE</th>
<th>STRUCTURE</th>
<th>SEQ</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 A T</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2 A T</td>
<td>-5</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>4 A E</td>
<td>-5</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>5 A B</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>6 A A S</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>7 A A S</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>8 A A S</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>9 A A S</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>10 A A S</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Secondary Structure Prediction

Three-state model: helix, strand, coil

Given a protein sequence:
- NWVLSTAADMQGVVTDGMASGLDKD...

Predict a secondary structure sequence:
- LLEEEELLLLHHLHHHLHHL...

Methods:
- statistical
- stereochemical

Accuracy: 50-85%

Statistical Methods

Residue conformational preferences:
Glul, Ala, Leu, Met, Gln, Lys, Arg - helix
Val, Ile, Tyr, Cys, Trp, Phe, Thr, Thr - strand
Gly, Asn, Ser, Lys, Asp - turn

Chou-Fasman algorithm:
Identification of helix and sheet "nuclei"
Propagation until termination criteria met
Chou-Fasman Algorithm

Identification of helix and sheet "nuclei"
- helix - 4 out of 6 residues with high helix propensity (P > 100)
- sheet - 3 out of 5 residues with high sheet propensity (P > 100)

Propagation until termination criteria met

Turn prediction
1) \( p(t) > 0.000075 \)
2) \( P(\text{turn}) > 1.00 \)
3) \( P(a) < P(\text{turn}) < P(b) \)

where \( p(t) = f(j)f(j+1)f(j+2)f(j+3) \)

Garnier - Osguthorpe - Robson (GOR) Algorithm

Likelihood of a secondary structure state depends on the neighboring residues:

\[ L(S_j) = \sum (S_j; R_{j+m}) \]

Window size - \([j-8; j+8]\) residues

Accuracy for a single sequence - 60%
Accuracy for an alignment - 65%

Evolutionary Information

Evolutionary Methods

Taking into account related sequences helps in identification of “structurally important” residues.

Algorithm:
- find similar sequences
- construct multiple alignment
- use alignment profile for secondary structure prediction

Additional information used for prediction
- mutation statistics
- residue position in sequence
- sequence length

Adapted from: Zvelebil, Baum, 2008
Evolutionary Methods
Neural Networks

Perceptron

\[ Y = \begin{cases} 
1 \text{ if } \sum w_i i > \Theta \\ 
0 \text{ otherwise} 
\end{cases} \]

Learning process: \( \Delta w_i = (T_p - Y_p) i_p \)

Stereochemical Methods

Patterns of hydrophobic and hydrophilic residues in secondary structure elements:

- segregation of hydrophobic and hydrophilic residues
- hydrophobic residues in the positions 1-2-5 and 1-4-5
- oppositely charged polar residues in the positions 1-5 and 1-4 (e.g. Glu (i), Lys (i+4))

Definitions of hydrophobic and hydrophilic residues (hydrophobicity scales) are ambiguous

Stereochemical Methods

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( i, i+2 )</th>
<th>( i, i+3 )</th>
<th>( i, i+4 )</th>
<th>( i, i+5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F-F )</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>-</td>
</tr>
<tr>
<td>( F-L )</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>*</td>
</tr>
<tr>
<td>( L-F )</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>*</td>
</tr>
<tr>
<td>( L-L )</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( i, i+1 )</th>
<th>( i, i+2 )</th>
<th>( i, i+3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F-F )</td>
<td>-</td>
<td>*</td>
<td>-</td>
</tr>
<tr>
<td>( F-L )</td>
<td>*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( L-F )</td>
<td>-</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>( L-L )</td>
<td>-</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
Jnet Algorithm

sequence-to-structure network
1st profile 2nd profile 3rd profile alignment

structure-to-structure network

Adopted from Zvelebil, Baum, 2008

Accuracy of prediction

Adopted from Zvelebil, Baum, 2008

Accuracy of prediction

Adopted from Zvelebil, Baum, 2008

Chemotaxis protein CheY
Residues 81–88
Hydrophobic face

Hydrophobic face

Accuracy of prediction

Adopted from Zvelebil, Baum, 2008

EV A (http://cubic.bioc.columbia.edu/eva/)

Adopted from Zvelebil, Baum, 2008
Accuracy of Prediction

\[ Q_3 = \frac{PH + PE + PC}{N} \]

\[ W = \log \frac{TP \times TN}{FP \times FN} \]

Range: 50-85%