Protein structure analysis
and verification

Risto Laakso
risto.laakso@hut.fi

23rd November 2004
Overview

1. Proteins analyzed

2. Structure similarity analysis

3. Stereochemical analysis

4. Structure verification
Proteins 1HLB and 1HLM

1HLB HEMOGLOBIN (SEA CUCUMBER)

1HLM HEMOGLOBIN (CYANO-MET) (SEA CUCUMBER)
Structure similarity analysis

- Structures of two proteins can be compared for similarity.

- Root-mean-square deviation calculates deviation between \(C_\alpha\)-atom positions.

- Distance-matrix alignment (DALI) method calculates residue-residue \((C_\alpha - C_\alpha)\) distance matrices.
Root mean square deviation

- Calculates squareroot of average of distance between $C_\alpha$-atom positions squared, i.e.

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (r_{i}^{model} - r_{i}^{real})^2}$$
RMSD=0.81157
RMS-deviation for 1HLB and 1HLM

Root-mean-square deviation:

0.0 for 1HLB vs. 1HLB (identical)

2.9 for 1HLM vs. 1HLB (quite similar)
Distance-matrix alignment (DALI)

- As proteins evolve, their structures change.
- Patterns of contacts between residues tend to remain similar.
- Distantly-related proteins can be identified by detecting conserved contact patterns.
http://www2.ebi.ac.uk/dali/

<table>
<thead>
<tr>
<th>Chain</th>
<th>raw-score</th>
<th>Z-score</th>
<th>%id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1HLB</td>
<td>1932.5</td>
<td>30.7</td>
<td>100</td>
</tr>
<tr>
<td>1HIM</td>
<td>1036.6</td>
<td>15.5</td>
<td>58</td>
</tr>
</tbody>
</table>

Raw-score is “raw DALI-score”,

while Z-score is normalized (e.g. \( \frac{\text{score} - \text{mean}}{\text{deviation}} \)).
Stereochemical analysis

- All protein structures have common polypeptide chain.

- Rotation is permitted around $N - C_\alpha$ and $C_\alpha - C$ single bonds of all residues.

- Angles $\phi$ and $\psi$ around these bonds, and angle $\omega$ around the peptide bond define the conformation of a residue.
Ramachandran plot

- Ramachandran plot shows $\phi - \psi$ torsion angles for all residues in the protein structure.

- The peptide bond tends to be planar, with two allowed states: trans, $\omega = 180^\circ$ (usually), and cis, $\omega = 0^\circ$ (rarely).
Ramachandran plot angles
Ramachandran plot of 1HLB and 1HLM
Protein structure verification

- PDB-file consistency checks (chain naming, atom weights, etc).
- Geometry checks (chirality, bond lengths, bond angles, torsion angles, etc).
- Verification for both PDB-structure depositors and users.
WHAT_IF / WHAT_CHECK

- Various verification checks.


Verification log for 1HLB

Most checks OK, some warnings, e.g.:

Warning: Chirality deviations detected

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>153</td>
<td>LEU (153)</td>
<td>C</td>
<td>6.1</td>
<td>10.8</td>
</tr>
<tr>
<td>154</td>
<td>VAL (154)</td>
<td>C</td>
<td>8.5</td>
<td>16.0</td>
</tr>
<tr>
<td>154</td>
<td>VAL (154)</td>
<td>CB</td>
<td>-7.7</td>
<td>-47.7</td>
</tr>
</tbody>
</table>

Full verification log 17 pages long.
The end.