Homology Modelling

Thomas Holberg Blicher
NNF Center for Protein Research
University of Copenhagen
Why are Protein Structures so Interesting?

• They provide a detailed picture of interesting biological features, such as active site, substrate specificity, allosteric regulation etc.

• They aid in rational drug design and protein engineering.

• They can elucidate evolutionary relationships undetectable by sequence comparisons.

• They can be used to put mutations in the proper structural context.
Learning Objectives

• Outline the basic steps in comparative protein structure modelling.

• Explain how structure models can be used to support biological hypotheses.

• Perform simple homology modelling using web servers and evaluate the results.
The Protein Data Bank

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray</td>
<td>13116</td>
<td>30860</td>
<td>74567</td>
</tr>
<tr>
<td>NMR</td>
<td>2451</td>
<td>5368</td>
<td>9605</td>
</tr>
<tr>
<td>Other</td>
<td>338</td>
<td>200</td>
<td>674</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>15905</strong></td>
<td><strong>36428</strong></td>
<td><strong>84846</strong></td>
</tr>
</tbody>
</table>

The PDB also contains nucleotide and nucleotide analogue structures.
Growth of Sequences

http://www.kurzweilai.net/dna-sequencing-data
No Structure = No Go?

• Do structure modelling.
  – Use known protein structures to make models.
    • Comparative modeling (Easy)
    • Fold recognition (Difficult)
    • New/unknown fold (Ab initio methods – very difficult)

• Validate the model.

• Adjust expectations and use accordingly!
Do We Need Homology Modelling?

• *Ab Initio* protein folding (random sampling):
  – 100 aa, 3 conf./residue gives approximately $10^{48}$ different overall conformations!

• Random sampling is *NOT feasible*, even if conformations can be sampled at picosecond ($10^{-12}$ sec) rates.
  – Levinthal’s paradox

• Do homology modelling instead.
How Is It Possible?

• The structure of a protein is uniquely determined by its amino acid sequence (but sequence is sometimes not enough):
  – prions
  – pH, ions, cofactors, chaperones

• Structure is conserved much longer than sequence in evolution.
  – Structure > Function > Sequence
How Often Can We Do It?

- Currently 85000 structures in the PDB
  - Reduces to 20000 structures (chains) <30 % identical (sequence) with a resolution <3.0 Å.
  - These fall in ~1400 different structure classes (folds).

- ~25% of all sequences can be modelled.

- ~50% can be assigned to a fold class.
Protein Folds (SCOP) in PDB

No new folds!
Worldwide Structural Genomics

- "Fold space coverage"

- Complete genomes
  - Disease-causing organisms
  - Model organisms

- Membrane proteins

- Protein-ligand interactions

Hou et al., PNAS 2003, 100: 2386-2390
What a **single new fold** gives.

<table>
<thead>
<tr>
<th>Experimental Structure</th>
<th>Models or fold assignments</th>
<th>Models</th>
<th>Useful models</th>
<th>Less accurate models</th>
<th>Fold assignments only</th>
</tr>
</thead>
<tbody>
<tr>
<td>P005</td>
<td>537</td>
<td>345</td>
<td>53</td>
<td>292</td>
<td>192</td>
</tr>
<tr>
<td>P007</td>
<td>42</td>
<td>40</td>
<td>28</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>P008</td>
<td>31</td>
<td>29</td>
<td>24</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>P018</td>
<td>172</td>
<td>50</td>
<td>11</td>
<td>39</td>
<td>122</td>
</tr>
<tr>
<td>P100</td>
<td>185</td>
<td>70</td>
<td>11</td>
<td>59</td>
<td>115</td>
</tr>
<tr>
<td>P102</td>
<td>26</td>
<td>25</td>
<td>22</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>P111</td>
<td>46</td>
<td>44</td>
<td>23</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1039</strong></td>
<td><strong>603</strong></td>
<td><strong>172</strong></td>
<td><strong>431</strong></td>
<td><strong>436</strong></td>
</tr>
</tbody>
</table>

1. A model is counted if it is at least 60 residues long and is assessed to have >30% of its Cα atoms within 3.5 Å of their true positions. The models are subdivided into two classes. “Useful models” are defined to be based on >30% sequence identity to the known structure, while “Less accurate models” are based on <30% sequence identity. “Fold assignments only” denotes the number of proteins with a significant PSI-BLAST relationship to a known structure (E < 0.0001) that failed to produce a reliable model. The calculations were performed in August, 2000.

---

Protein Folds

http://www.jcsg.org/
How Well Can We Do It?

How Is It Done?

• Identify template(s)
  – Initial alignment
• Improve alignment

• Backbone generation

• Loop modelling
• Side chains
• Refinement

• Validation ←
Template Identification

• Search with sequence
  – Blast
  – Psi-Blast
  – Fold recognition methods
Sequence vs. Structure

• Residues in the same column in an alignment are either:
  – *Structurally* equivalent/similar
  – *Evolutionary* equivalent/related/homologous

• Different types of similarity not necessarily equivalent.

• Use biological information to guide/adjust your alignment.
  – Functional annotation in databases
  – Active site/motifs
Figure 25.3. A typical residue exchange or scoring matrix used by alignment algorithms. Because the score for aligning residues A and B is normally the same as for B and A, this matrix is symmetric.
<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>D</th>
<th>I</th>
<th>C</th>
<th>R</th>
<th>L</th>
<th>P</th>
<th>G</th>
<th>S</th>
<th>A</th>
<th>E</th>
<th>A</th>
<th>V</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>6</td>
<td>-2</td>
<td>0</td>
<td>-3</td>
<td>-2</td>
<td>2</td>
<td>-2</td>
<td>-3</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-2</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>N</td>
<td>-3</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>0</td>
<td>-2</td>
<td>-2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>-2</td>
<td>2</td>
<td>-2</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>5</td>
<td>-2</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
<td>5</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>-3</td>
<td>2</td>
<td>-2</td>
<td>-3</td>
<td>0</td>
<td>-2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>-1</td>
<td>-3</td>
</tr>
<tr>
<td>A</td>
<td>-2</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>-2</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>-3</td>
<td>5</td>
<td>-2</td>
<td>-2</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
<td>-1</td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>C</td>
<td>-3</td>
<td>-2</td>
<td>-2</td>
<td>8</td>
<td>-2</td>
<td>-3</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-2</td>
<td>-2</td>
<td>8</td>
</tr>
</tbody>
</table>

1 2 3 4 5 6 7 8 9 10 11 12 13 14

PHE ASP ILE CYS ARG LEU PRO GLY SER ALA GLU ALA VAL CYS

PHE ASN VAL CYS ARG THR PRO --- --- --- GLU ALA ILE CYS

PHE ASN VAL CYS ARG --- --- --- THR PRO GLU ALA ILE CYS
<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>D</th>
<th>I</th>
<th>C</th>
<th>R</th>
<th>L</th>
<th>P</th>
<th>G</th>
<th>S</th>
<th>A</th>
<th>E</th>
<th>A</th>
<th>V</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>6</td>
<td>-2</td>
<td>0</td>
<td>-3</td>
<td>-2</td>
<td>2</td>
<td>-2</td>
<td>-3</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-2</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>N</td>
<td>-3</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>0</td>
<td>-2</td>
<td>-2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>-2</td>
<td>2</td>
<td>-2</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>5</td>
<td>-2</td>
</tr>
<tr>
<td>C</td>
<td>-3</td>
<td>-2</td>
<td>8</td>
<td>-2</td>
<td>-3</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-2</td>
<td>-2</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
<td>5</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>-2</td>
<td>0</td>
<td>-2</td>
<td>-3</td>
<td>0</td>
<td>-2</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-3</td>
</tr>
<tr>
<td>E</td>
<td>-3</td>
<td>2</td>
<td>-2</td>
<td>-3</td>
<td>0</td>
<td>-2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>-1</td>
<td>-3</td>
</tr>
<tr>
<td>A</td>
<td>-2</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>-2</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>-3</td>
<td>5</td>
<td>-2</td>
<td>-2</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
<td>-1</td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>C</td>
<td>-3</td>
<td>-2</td>
<td>8</td>
<td>-2</td>
<td>-3</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-2</td>
<td>-2</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHE</td>
<td>ASP</td>
<td>ILE</td>
<td>CYS</td>
<td>ARG</td>
<td>LEU</td>
<td>PRO</td>
<td>GLY</td>
<td>SER</td>
<td>ALA</td>
<td>GLU</td>
<td>ALA</td>
<td>VAL</td>
<td>CYS</td>
<td></td>
</tr>
<tr>
<td>PHE</td>
<td>ASN</td>
<td>VAL</td>
<td>CYS</td>
<td>ARG</td>
<td>THR</td>
<td>PRO</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>GLU</td>
<td>ALA</td>
<td>ILE</td>
<td>CYS</td>
<td></td>
</tr>
<tr>
<td>PHE</td>
<td>ASN</td>
<td>VAL</td>
<td>CYS</td>
<td>ARG</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>THR</td>
<td>PRO</td>
<td>GLU</td>
<td>ALA</td>
<td>ILE</td>
<td>CYS</td>
<td></td>
</tr>
</tbody>
</table>
### Improving the Alignment

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHE</td>
<td>ASP</td>
<td>ILE</td>
<td>CYS</td>
<td>ARG</td>
<td>LEU</td>
<td>PRO</td>
<td>GLY</td>
<td>SER</td>
<td>ALA</td>
<td>GLU</td>
<td>ALA</td>
<td>VAL</td>
<td>CYS</td>
</tr>
<tr>
<td>PHE</td>
<td>ASN</td>
<td>VAL</td>
<td>CYS</td>
<td>ARG</td>
<td>THR</td>
<td>PRO</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>GLU</td>
<td>ALA</td>
<td>ILE</td>
<td>CYS</td>
</tr>
<tr>
<td>PHE</td>
<td>ASN</td>
<td>VAL</td>
<td>CYS</td>
<td>ARG</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>THR</td>
<td>PRO</td>
<td>GLU</td>
<td>ALA</td>
<td>ILE</td>
<td>CYS</td>
</tr>
</tbody>
</table>

From "Professional Gambling" by Gert Vriend
http://www.cmbi.kun.nl/gv/articles/text/gambling.html
Template Quality

• Selecting the best template is crucial!
• The best template may not be the one with the highest % id (best p-value…)
  – Template 1: 93% id, 3.5 Å resolution 😞
  – Template 2: 90% id, 1.5 Å resolution 😊
The Importance of Resolution

4 Å
3 Å
2 Å
1 Å

low

high

0.5 Å
Key Parameters

- **Resolution**
- **R values**
  - Agreement between data and model.
  - Usually between 0.15 and 0.25, should not exceed 0.30.
    - $R + 0.05 > R_{\text{free}} > R$
- **Ramachandran plot**
- **B factors**
  - Contributions from static and dynamic disorder
    - Well determined $\sim$10-20 Å$^2$, intermediate $\sim$20-30 Å$^2$, flexible 30-50 Å$^2$, invisible $>$60 Å$^2$. 
Template Quality – Ramachandran Plot

X-ray structure – good data.  
NMR structure – low quality data…
Error Recovery

- Errors in the model can NOT be recovered at a later step
  - The alignment can not make up for a bad choice of template.
  - Loop modeling can not make up for a poor alignment.
- The step where the errors were introduced should be redone.
Validation

• Most programs will get the bond lengths and angles right.

• Model Rama. plot ~ template Rama. plot.
  – select a high quality template!

• Inside/outside distributions of polar and apolar residues.
Model Validation – ProQ

• ProQ is a neural network-based predictor
  – Structural features $\rightarrow$ quality of a protein model.

• ProQ is optimized to find
  – correct models…
  – …NOT (necessarily) native structures.

• Two quality measures:
  – MaxSub & LGscore

Arne Elofssons group: http://www.sbc.su.se/~bjorn/ProQ/
Summary

- Successful homology modelling depends on the following:
  - Template quality
  - Alignment (add biological information)
  - Modelling program/procedure (try more than one)

- Always validate your final model!
Finding Remote Homologues (Fold Recognition)
Why %id Is a Poor Measure

1200 models sharing 25-95% sequence identity with the submitted sequences (www.expasy.ch/swissmod)

Probabilities of SWISS-MODEL accuracy for target-template identity classes.

<table>
<thead>
<tr>
<th>Percent sequence identity</th>
<th>Total number of models</th>
<th>Percent models with rmsd lower than 1 Å</th>
<th>Percent models with rmsd lower than 2 Å</th>
<th>Percent models with rmsd lower than 3 Å</th>
<th>Percent models with rmsd lower than 4 Å</th>
<th>Percent models with rmsd lower than 5 Å</th>
<th>Percent models with rmsd higher than 5 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>25–29</td>
<td>125</td>
<td>0</td>
<td>10</td>
<td>30</td>
<td>46</td>
<td>67</td>
<td>33</td>
</tr>
<tr>
<td>30–39</td>
<td>222</td>
<td>0</td>
<td>18</td>
<td>45</td>
<td>66</td>
<td>77</td>
<td>23</td>
</tr>
<tr>
<td>40–49</td>
<td>156</td>
<td>0</td>
<td>9</td>
<td>44</td>
<td>63</td>
<td>78</td>
<td>91</td>
</tr>
<tr>
<td>50–59</td>
<td>155</td>
<td>18</td>
<td>55</td>
<td>79</td>
<td>86</td>
<td>91</td>
<td>9</td>
</tr>
<tr>
<td>60–69</td>
<td>145</td>
<td>18</td>
<td>55</td>
<td>79</td>
<td>86</td>
<td>91</td>
<td>8</td>
</tr>
<tr>
<td>70–79</td>
<td>137</td>
<td>42</td>
<td>71</td>
<td>82</td>
<td>85</td>
<td>88</td>
<td>12</td>
</tr>
<tr>
<td>80–89</td>
<td>173</td>
<td>45</td>
<td>79</td>
<td>86</td>
<td>94</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>90–95</td>
<td>88</td>
<td>59</td>
<td>78</td>
<td>83</td>
<td>86</td>
<td>91</td>
<td>9</td>
</tr>
</tbody>
</table>

a: Range of sequence identity between target and template sequence.

b: Total number of models in any given class of sequence identity. The table summarizes 1301 model-template structure pairs.

c: Probability in percent that a model, sharing X% sequence identity with its template, deviates by 1 Å or less from the corresponding experimental control structure. The following columns provide these probabilities for other rms deviations.
Identification of Correct Fold

• % ID is a poor measure
  – Many evolutionarily related proteins share low sequence identity
  – A short alignment of 5 amino acids can share 100% id, but what does it mean?

• Alignment score even worse
  – Many sequences will score high against each other (especially in hydrophobic stretches)

• P-value or E-value more reliable.
What are P and E values?

- **E-value**
  - Number of expected hits in database with score higher than match
  - Depends on database size

- **P-value**
  - Probability that a random hit will have score higher than match
  - Database size independent

Score 150
10 hits with higher score (E=10)
10000 hits in database =>
P=10/10000 = 0.001
Sequence Profiles

- Not all positions in a protein are equally likely to mutate
  - Some amino acids (active sites) are highly conserved, and the score for mismatch must be very high.
  - Other amino acids can mutate almost freely, and the score for mismatch should be lower than the BLOSUM score.

- Sequence profiles can capture these differences.
Sequence Profiles

ADDGSLAFVPSEF--SISPGGEKIVFKNNAGFPHNIVFDEDSISPSGVDAKISMSSEEDLLN
TVNGAI--PGPLIAERLKGRQGQRVTNTLDDETSIHWHGLLVPGMDGVPVGSGFPG---I
-TSMAPAFGVQEFYRTVKQGDEVTVTIT-----NIDQIEd-VShtGmVvNtGvSMe---I
IE--KMkYLTPEVFYTIKAGETVYwVngEvMPhnVAFKKGIV--GEdAFrGmMmtKD---
-TsvApSFSQPSF-LTVKREGDEVTIVTNLDE-------IDDLTHGFTMGNHGVAME---V
AsAEtMVfEpdFLVrLIGpGDRVRFVTPHKS-NHAAAtIDGMVPEGVEGFkSrinDE----
TVNGQ--FPGRlAGVAREGDQVLVKGvNhVAEnITIhWHGvQltGwADGpAYVTQCPI

Matching anything but G →
large negative score

Anything can match
How to Make Sequence Profiles

PSIBLAST

• Align (BLAST) sequence against large sequence database (Swiss-Prot).

• Select significant alignments and make profile (weight matrix) using techniques for sequence weighting and pseudo counts.

• Use weight matrix to align against sequence database to find new significant hits.

• Repeat 2 and 3 (normally 3 times!).
Ab Initio Methods
No Template – No Go?

• De novo / *ab initio* / free modelling methods:
  – simulate the biological process of protein folding

• A VERY DIFFICULT task because a protein chain can fold into millions of different conformations.

• Use it **only** when no detectable homologues can be found.

• Methods can also be useful for fold recognition in cases of extremely low homology (e.g. convergent evolution).

http://cnx.org/content/m11461/latest/
Fragment-based *ab initio* modelling

- Rosetta method of the Baker group:
  - Secondary structure prediction
  - Fragments library of 3 and 9 residues from known structures
  - Link fragments together, use only backbone and CB atoms
  - Contact/pair potential
  - Energy minimization techniques (Monte Carlo optimization) to calculate tertiary structure
  - Refine structure including side chains


http://robetta.bakerlab.org/
Problems with Empirical Potentials

Fragments with correct local structure

http://www.cs.ucl.ac.uk/staff/d.jones/t42morph.html
Two-high-scoring predictions by the top groups in FR/H (top) and FR/A (bottom). The assigned z-scores are given for the top predictions (center) as well as for two average predictions (right).

G. Wang  Assessment of fold recognition predictions in CASP6, Proteins 61, S7, Pages 46-66
Human intervention

- The best groups in CASP use maximum knowledge of query proteins

- Specialists can help to find a correct template and correct alignments

Knowledge of function
- Cysteines forming disulfide bridges or binding e.g. zinc molecules
- Proteolytic cleavage sites
- Other metal binding residues
- Antibody epitopes or escape mutants
- Ligand binding
- Results from CD or fluorescence experiments
Human Intervention II

- **Fold It: The Protein Folding Game**
  - Rosetta Energy Potentials

- [http://fold.it/portal/](http://fold.it/portal/)

- Uses the HUMAN pattern recognition abilities for finding the lowest energy fold.
“How Fast-folding Proteins Fold”

Modelling Servers

• Comparative (homology) modelling:
  – CPHmodels (simple)
    • http://www.cbs.dtu.dk/services/CPHmodels/
  – SwissModel (intermediate)
    • http://swissmodel.expasy.org
  – HHpred (complex)
    • http://toolkit.tuebingen.mpg.de/hhpred

• Ab initio
  – Robetta (also comparative; intermediate)
    • http://robetta.bakerlab.org
Summary

• Methods using sequence profiles are best
• Use only *ab initio* methods if necessary and know that the quality is really low!
• Try to use as much knowledge as possible for alignment and template selections in difficult cases.
• Use meta-servers when you can.
• TRY FOLDIT!