Reading instructions for: "Practical lessons from protein structure prediction", Ginalski et al. 2005

By Anne Molgaard and Thomas Blicher

The idea about including this paper in the course is to give a general overview of methods used in protein fold recognition. In addition, a number of good take-home messages are mentioned in the last three chapters of the paper. You should go more into detail when reading these chapters ("Results of evaluation experiments"-"Conclusions"), whereas the first part of the paper can be read without going too much into detail about how the methods work.

The sections “Assessment of 3D models” and “Evaluation protocols” are not part of the reading assignment. In addition, the paper includes a number of tables describing servers and validation measures and you are not required to read these.

Here are some keywords which can be useful for guiding you through the text:

**Introduction**
Motives for structure prediction
Historical overview of the structure prediction field

**Sequence similarity-based methods**
BLAST and FASTA
PSI-BLAST
HMM based methods
Profile-profile comparison methods

**Threading methods**
Residue contact potentials
Fitness of sequence to known structure

**Hybrid sequence similarity and threading methods**
Mutation preferences
Frozen approximation

**Ab initio methods**
Latice based
Fragment library based
Quality

**Meta-predictors**
Clustering of models
Structural comparison
Assembly of different initial models

**Evaluation of prediction methods**
CASP
CAFASP
LiveBench

**Results of evaluation experiments**
Short segments of significant similarity
Conclusions:
- Individual threading/hybrid methods are not better than good sequence-profile methods
Meta-predictors out-perform single methods.
Domain definitions are important as most methods cannot handle multi-domain proteins.
A high meta-method score is most reliable when several independent methods recognize the same fold for a sequence.
In fold recognition human intervention can be fruitful if additional knowledge such as similar function and conservation of essential amino acids is used with expertise to chose templates and correct alignments.
Ab initio methods are in general of low quality and should only be used if everything else fails. Some biological hints can be obtained from these models.
It is helpful to estimate which parts of a model is most reliable.
Many on-line servers are unfortunately too slow for high-throughput methods and therefore in-house installations should be used for large-scale analysis.

Utility of low resolution protein structure predictions
- Low resolution models can help to define domain boundaries and thereby lead to better constructs that facilitate experimental procedures.
- Analysis of variant active sites and guidance of mutational studies.
- Function prediction.
- Aid in experimental structure determination processes (X-ray or NMR).