Protein Structure Prediction

Protein Bioinformatics Workshop
Croucher Foundation ASI

Presented by Kambo Wong
Department of Biochemistry
The Chinese University of Hong Kong

What is Structure Prediction?

Protein sequence

```
              10        20        30        40        50
....|....|....|....|....|....|....|....|....|....|
SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDML
              60        70        80        90       100
....|....|....|....|....|....|....|....|....|....|
NPNYEDLLIRKSNHSFLVQAGN-VQLRVIGHSMQNCLLRLKVDTSNPKTP
```

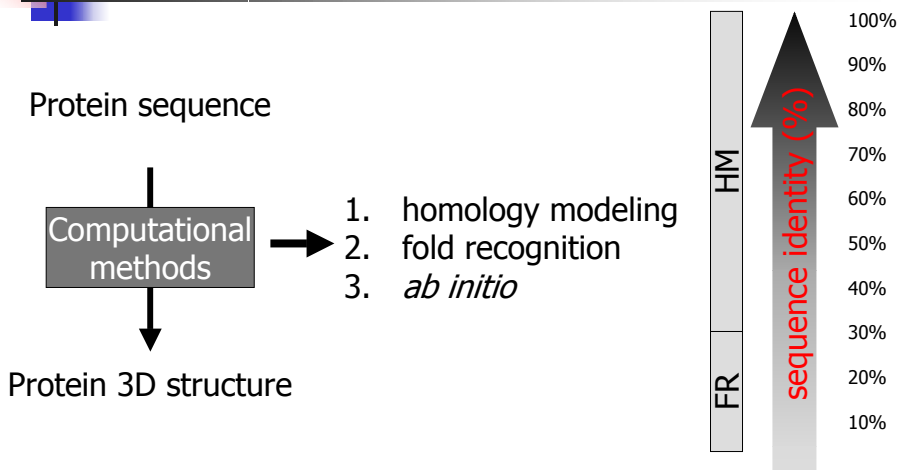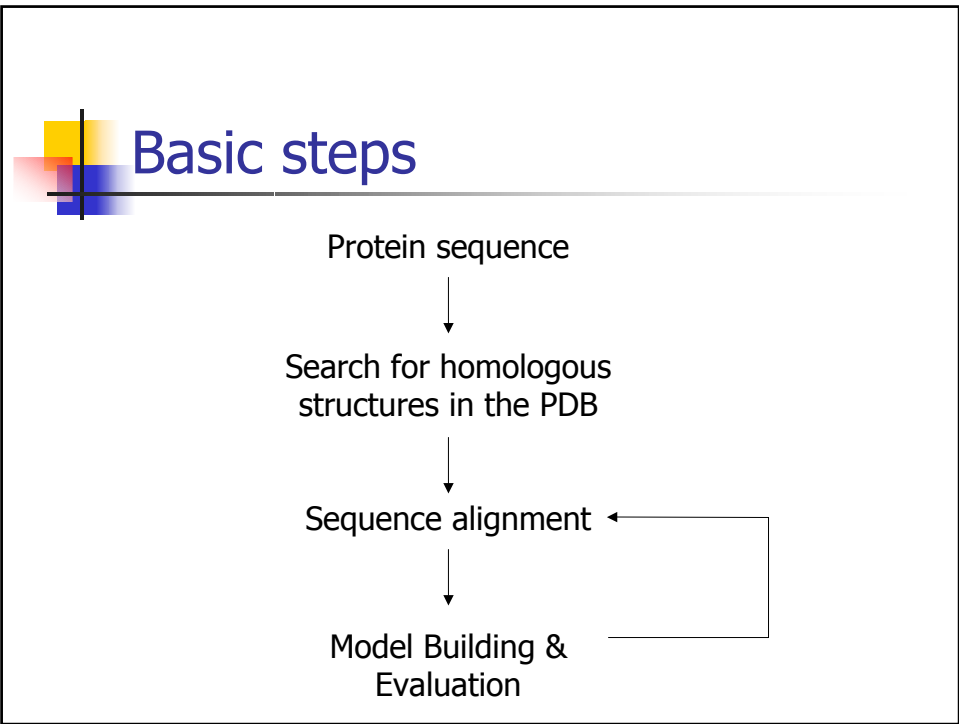Computational
methods

Protein 3D structure

Why?

- Demand
  - Determination of protein sequences far outpaces structure determination

- Useful
  - Structure-function studies
  - Drug design
  - Protein Engineering

Methods available

Protein sequence

↓

Computational methods → 1. homology modeling
                        2. fold recognition
                        3. *ab initio*

↓

Protein 3D structure

HM

FR

sequence identity (%)

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%

Homology Modeling

Basic steps

Protein sequence

↓

Search for homologous
structures in the PDB

↓

Sequence alignment

↓

Model Building &
Evaluation

Model building

- SWISS-MODEL
  (http://swissmodel.expasy.org)

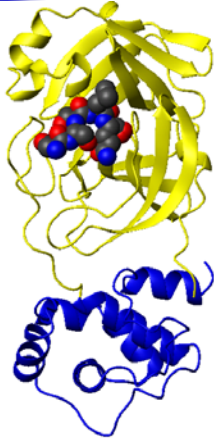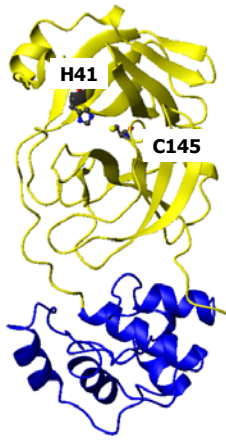- MODELLER
  (http://salilab.org/modeller/)

Example 1

SARS Main Protease
A tutorial on homology modeling
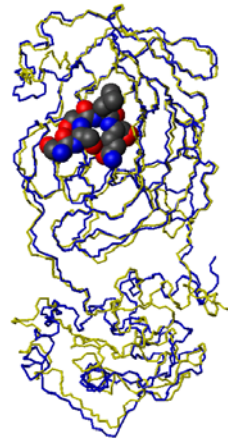using SWISS-MODEL
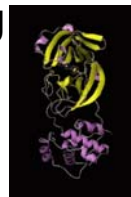
The predicted structure is similar to the structure determined experimentally

H41

C145

Crystal structure of M^{pro} solved by Rao ZH's laboratory (pdb: 1uk4)
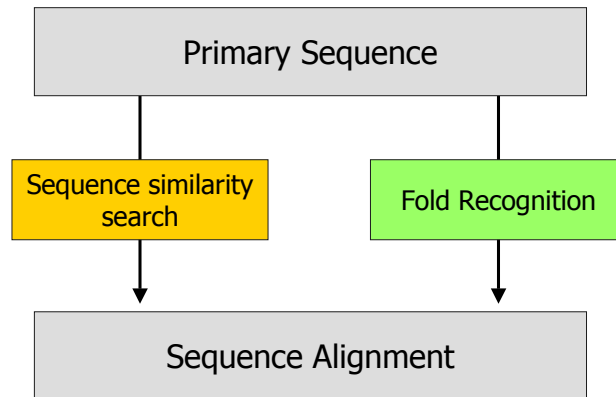
Predicted structure

Fold recognition

when sequence identity is too low
to be detectable by simple
sequence similarity search

Homology modeling vs Fold recognition

```
        ┌─────────────────────────────────┐
        │        Primary Sequence         │
        └─────────────────────────────────┘
            │                        │
            ▼                        ▼
  ┌──────────────────┐    ┌──────────────────┐
  │ Sequence similarity│   │  Fold Recognition │
  │      search       │    │                  │
  └──────────────────┘    └──────────────────┘
            │                        │
            ▼                        ▼
        ┌─────────────────────────────────┐
        │        Sequence Alignment        │
        └─────────────────────────────────┘
```

A list of servers for structure prediction

- **BIOINGU** (http://www.cs.bgu.ac.il/~bioinbgu/form.html)
- **SAM-T02** (http://www.soe.ucsc.edu/research/compbio/HMM-apps/T02-query.html)
- **RAPTOR** (http://www.cs.uwaterloo.ca/~j3xu/RAPTOR_form.htm)
- **FFAS03** (http://ffas.ljcrf.edu/ffas-cgi/cgi/ffas.pl)
- **ORFeus** (http://basic.bioinfo.pl/)
- **3DPSSM** (http://www.sbg.bio.ic.ac.uk/~3dpssm/)
- **GenThreader** (http://bioinf.cs.ucl.ac.uk/psipred/)
- **FUGUE** (http://www-cryst.bioc.cam.ac.uk/~fugue/)

Which server is the best?

- Critical Assessment of Techniques for Protein Structure Prediction (CASP). (http://predictioncenter.llnl.gov/)
- LiveBench (http://bioinfo.pl/LiveBench/)

- The meta-server approach.
  - e.g. The bioinfo.pl meta-server (http://bioinfo.pl/meta)

Example 2

SARS nsp9

Example 3: SARS S protein

Some mistakes to avoid …

Sequence alignment

Buried charge residues

References

- Schwede T, Kopp J, Guex N, and Peitsch MC (2003) **SWISS-MODEL: an automated protein homology-modeling server.** *Nucleic Acids Research* 31, 3381-3385.

- Ginalski K, Elofsson A, Fischer D, and Rychlewski L (2003) **3D-Jury: a simple approach to improve protein structure predictions.** *Bioinformatics* 19, 1015-1018.