

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.....  
S G F R K M A E P S G K V E G C M V Q V T C G T T I N G L W L D D T V Y C P R H V I C T A B D M L  
.....60.....70.....80.....90.....100.....  
N F N Y E D L L I R K S N H S F L V Q A G N - V Q L R V I G H S M O N C L L R L K V D T S N F K T P
```

Computational
methods

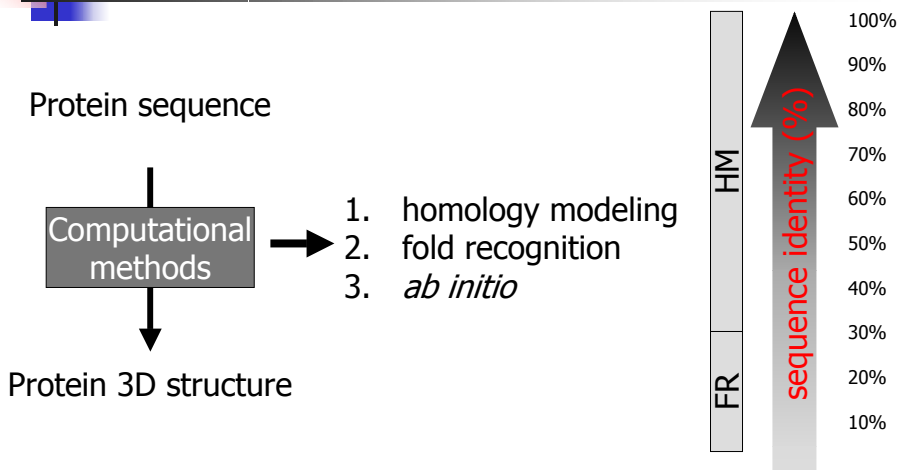
Protein 3D structure



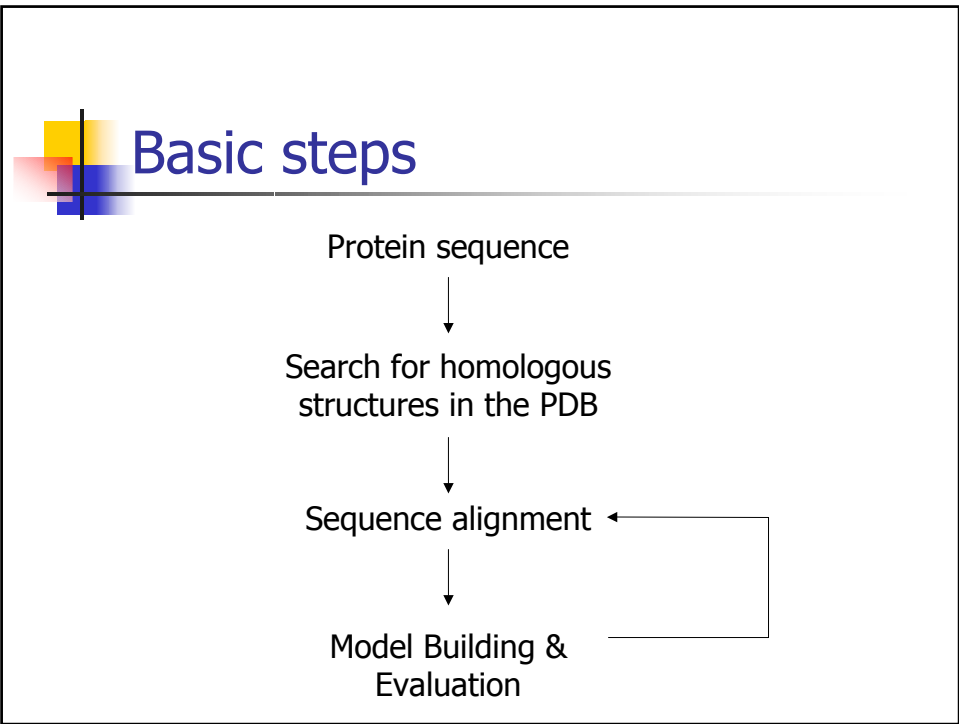
Why?


- Demand
 - Determination of protein sequences far outpaces structure determination
- Useful
 - Structure-function studies
 - Drug design
 - Protein Engineering

Methods available




Homology Modeling





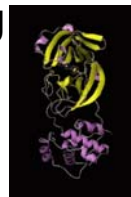
Model building

- SWISS-MODEL
(<http://swissmodel.expasy.org>)
- MODELLER
(<http://salilab.org/modeller/>)

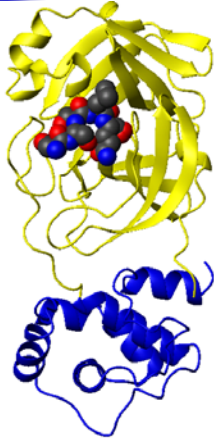


Example 1

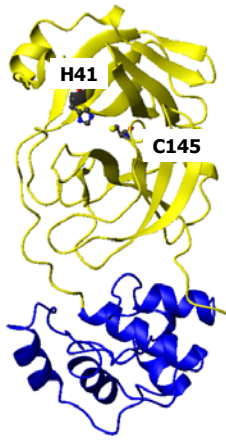
SARS Main Protease
A tutorial on homology modeling
using SWISS-MODEL



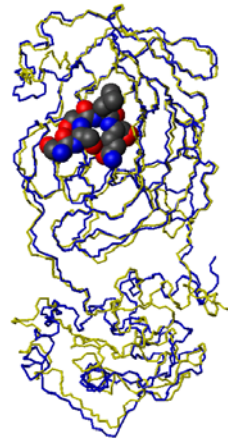
The predicted structure is similar to the structure determined experimentally



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



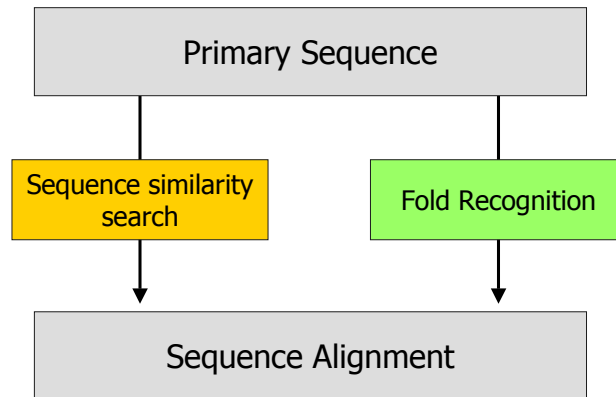
Predicted structure



Fold recognition

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

Homology modeling vs Fold recognition



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

```
s1 1 -----DRCRFDVDDVAHNYDHTSSNFGVYVFD-----EIFSDTLRLTLDFLFPYENVYGFHINHGKNPVI 65
1q9d 1 PVTINNFVNDPINDNNIIMHEHFAAGTGRVYKAFKITEHMLIPERVYFGYKPEDFNKSSGIFNBSVGCYDYDFVYLNINDKKNIFLQMIKILNRIKS 100

s1 66 PFKDGIYFAATEKSNVVFQVFGSTNNNKSQSVIIINSTNVVIRACNFELCDNPFFAVSEKQHGQTQHTHIFDNAFNCFEYISDAFSLDVSEKSNQFKH 165
1q9d 101 KPLGKLEMIINGIPYLDGRRVPLEFFNINIASVIVKLSINPGEVERKKGIFANLILPQGGVNLNENETIDKIGONFASREGFGGIMONKFCPEYVVS 200

s1 166 LREHFRVRSQFLVYVYRQGGIDVWRDLPSSGFNFK-----PFFHFFHGHINLIDRALILAFSFKODIYQSTSAAAAYFVQYKTTFFHLHYDEN-----G 254
1q9d 201 VFNWYDNRKSAHFNRRGQFSDPALILHHELIHIVLHGLYGIKVDLGGVFNERRFPHQSDAICDEELYTFGGQDFSEITTSFDDKSEYDKVLCNFRGLV 300

s1 255 TITDANRSCNPLAELRCVKSSEIDKGIYQISNFRVPSQDVVRFPNITNLGFRSEVDNAIKFSPVYANERKKKSNQVADYSVYVNSIFSTFKCYLVMS 354
1q9d 301 RNMKVLVQISDPPNINIIYKNGKDKYKVFERSQKYSIDVESFDKLYKSLFQVETNLDENYKIKIRASYFSSLSLPPVKIKMLDNEIYITIEGFNDS 400

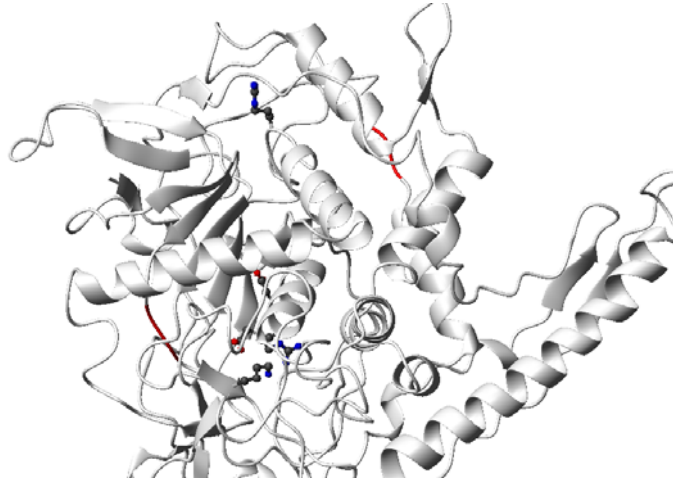
s1 355 ATKINDLCFS-----NVVDSFVRSQDDVRIAPGQGTGLADYNYKLDDEHGCVLAMVTPNIDATSTGNINNYKRYVLR-HGKLRPFPHSRNVVPSFDG 448
1q9d 401 DRDHEKEYRGNRAIKKCYEEERHSLAVYKIQCKSGVICIDVDNEDLGFYADKNSFSDLLSKNERIEHFNQSNVHENDFFINELIILDLEISKIQLP 500

s1 449 KPCFPPALNCVYPLNDSFYTTITGLQFQYRNVVLSFELLNRAAVCCGPRSEIDLIKNCUNFNFNGLTGVLTSSRFQFQFQSRDVSDFDQVSD 548
1q9d 501 SEMESLTFNVDVYVVERKQPAIKKIFIDENITFQYLYSQTFEDDIRDISLSTSFDDALLFENKRVYFFSHDYIKVAMQVWEAGLAFAGVVKQIVADPQTE 600

s1 549 PKTSEILDISPCSFGGQSVHPTPGNASSQVNLVYQVNCIDVSTLHADQLTEAUNFYSKCNVVFQTCAGCLIGAEHVDTSYECDFIIGAGICASYHIYS 648
1q9d 601 ANKENTPKIADISLIPYHGLAVNGEHTKGNFEN-AFELAGSILLEFSEILLFPVWGAFLLESYIDNKN---KIKITIDNALTKRNEKWSDMYGLI 696

s1 649 LLRSTQKSIIVAYTMS
1q9d 697 VAQLSEIVNTOFYTK
```


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