Announcements

Videos: YouTube / www.rostlab.org/talks
THANKS:

EXERCISES:

Special lectures:
• Mikal Boden UQ Brisbane

No lecture:
• 04/26 Security check Rostlab (exercise WILL be)
• 05/01 May Day (also no exercise)
• 05/08 Student representation (SVV) - exercise WILL happen
• 05/10 Ascension Day (also no exercise)
• 05/22 Whitsun holiday (also no exercise)
• 05/31 Corpus Christi (also no exercise)
• 06/21 no lecture (but exercise)

LAST lecture: bef: Jul 12

Examen:
• Makeup: Jul 12 18-20:00 (room TBA)

CONTACT: pp1ex@rostlab.org

Lothar Richter

Dmitrij Nechaev

Michael Heinzinger

Your Name

next
Recap: multiple alignments
Notation: protein structure 1D, 2D, 3D

| P | PP P | 128 | 110 |
| Q | QQQY | 175 | 97 |
| I | FFQVI | 70 E 60 |
| T | SSIVR | 77 E 69 |
| L | LLSTL | 120 E 14 |
| W | WWQED | 238 E 81 |
| Q | RKQAK | 169 E 97 |
| R | RRPFQ | 200 E 62 |
| P | PPPPP | 24 E 48 |
| L | VVTKF E | 71 E 59 |
| V | VVLI E | 14 E 0 |
| T | TTKEK E | 74 E 69 |
| I | AALIV E | 0 E 0 |
| K | HYKPF E | 90 E 73 |
| I | IILVE | 4 E 0 |
| G | EENGG | 46 E 41 |
| G | GGGTG | 62 E 53 |
| Q | QQKRR | 68 E 71 |
| L | PLLMW E | 118 E 59 |
| K | VVFKV E | 31 E 73 |
| E | EESKK E | 124 E 95 |
| A | VVGLG E | 1 E 0 |
| L | LLIL E | 29 E 0 |
| L | LLLVV E | 24 E 0 |
| D | DDDDD | 49 E 58 |
| T | TTTEE | 72 E 51 |
| G | GGGGG | 62 E 30 |
| A | AAAAA | 17 E 0 |
| D | DDDDD | 102 E 79 |
| D | DDAKE | 69 E 58 |
| T | SSTTV | 1 E 69 |
| V | IIIVIV E | 14 E 0 |
| L | VVIVL E | 0 E 0 |

1D 2D 3D
Pairwise alignments: sequence-sequence

YDFHGVGE{D}DISIKRG

YHDHGVAE{Q}QLLLLKKA

Generic scoring matrix
(here BLOSUM62)

### Sequence-profile comparison

**YDFHGVGEBDDISIKRG**

<table>
<thead>
<tr>
<th>PS-position specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>fyn_human VTLYFALDY EARTEDDSL HKGEKPQILN SSEGDWAER SLLTGETGQY</td>
</tr>
<tr>
<td>yrk_chick VTLYFALDY EARTEDDSL QKGEKEHIIN NTGREWAEAR SLSSSGATQY</td>
</tr>
<tr>
<td>fgr_human VTLYFALDY EARTEDDSL TKGEKHPILN NTGREWAEAR SLSSGRTQCI</td>
</tr>
<tr>
<td>yes_chick VTVFALDY EARTDDDSL HKGERFQIIN NTGREWAEAR SIATGKTYQI</td>
</tr>
<tr>
<td>src_avis2 VTVFALDY EARTDDDSL HKGERFLQIIN NTGREWAEAR SLTTGQTYQI</td>
</tr>
<tr>
<td>src_avisis VTVFALDY ESRTEEDDSL HKGERFQIIN NTGREWAEAR SLTTGQTYQI</td>
</tr>
<tr>
<td>src_avisor VTVFALDY ESRTEDDSL HKGERFQIIN NTGREWAEAR SLTTGQTYQI</td>
</tr>
<tr>
<td>src_chick VTVFALDY ESRTEDDSL HKGERFQIIN NTGREWAEAR SLTTGQTYQI</td>
</tr>
<tr>
<td>stk_hydat VTVFALDY EARISSDSL HKGERFQIIN TADGWVAR SLITNSEQI</td>
</tr>
<tr>
<td>src_rsvpa ........ ESRIETDSL SSRRRQIIN NTGREWAEAR SLTTGQTYQI</td>
</tr>
<tr>
<td>hck_human IVVALDY EALHEDSLQ QDGMVYLE ES GEWKR SLATTKEQYI</td>
</tr>
<tr>
<td>bsk_mouse FVVALDY AAVNRDLQVL KGEKEQILR .STGWLAR SLVTGREGQY</td>
</tr>
<tr>
<td>hck_mouse TIIVLALDY EALHEDSLQ QDGMVYLE .EAGEEWKR SLATKEQYI</td>
</tr>
<tr>
<td>lyn_human IVVALPY DGIHDDSLK KGEKMKVLQ .EHGEWAK SSLKKKEQFI</td>
</tr>
<tr>
<td>lck_human LVIALSY EPHSGDLGF KEGEQIRILE QS GEWKAQ SLTTGQEFI</td>
</tr>
<tr>
<td>ss8_l_yeast ....... ALPY DADDDeeISF EONELIQVS .IEGRWKAR R.ANGETEQI</td>
</tr>
<tr>
<td>abl1 mouse .LFVLYDF VAGINTLSI TKEKIRVQG YnnGEWCEAQ ..TKNGQQV</td>
</tr>
<tr>
<td>abl1_human .LFVLYDF VAGINTLSI TKEKIRVQG YnnGEWCEAQ ..TKNGQQV</td>
</tr>
<tr>
<td>src_l_drome .LVYLDY KSRDEDSLK MKGERMEID DTSEDVRV NLTRQEQLI</td>
</tr>
<tr>
<td>mys_d_dicdi ....... ALYDF DAESSMELSF KEQDIITVD QSSGDWDAE L..KGRKkv</td>
</tr>
<tr>
<td>yfj4_yeast ....... ALYSF AEGESDLPF RKGVTILK ksQNDWQGR V..NGRCEF</td>
</tr>
<tr>
<td>abl2_human .LFALYDF VAGINTLSI TKEKIRVQG YNNQNGEVR RSKNGQOQV</td>
</tr>
<tr>
<td>tec_human .EVALVYDF QAAEHDRLR EQGELYILE KNVHWWAR D.KYGNEQYI</td>
</tr>
<tr>
<td>abl1_cael .LFALYDF HGVGEQSLQ RKGQVRILG YKNHENEC GRLRGEI</td>
</tr>
<tr>
<td>tkx_human ....... ALYSF LPERPCLAL RAEELILE KYPNHWVAR D.RLGNEQGL</td>
</tr>
<tr>
<td>yha2_yeast VRRVVALDY TTNEPDSLK RKGVTILVLQ YVRDYNLGA L..RGNMCEF</td>
</tr>
<tr>
<td>abpl_sacex ....... AEYDY EAGEDNLTFL AENKIINIE FVDDWALGE LETTGQKGF</td>
</tr>
</tbody>
</table>

**PSI-BLAST** SF Altschul 1997 *Nucl Acids Res* 25 3389-3402

© Burkhard Rost

ROSTLAB. TUM
Profile-profile comparison

1  50  

fyn_human  VTLFAVLYD  EARTEDLSF  HKGEEKQILN  SSVGDOMVR  SLTGETQYI  
yrk_chick  VTLFAVLYD  EARTEDLSF  KQGERGPIIN  NTEDGWEAR  SSLSGATQI  
grf_human  VTLFAVLYD  EARTDDLSF  TGKTERKILN  NTEDGWEAR  SLSSTGKTCI  
yes_chick  VTLFAVLYD  EARTEDLSF  KQKERQPIIN  NTEDGWEAR  SIAEKTQI  
src_avisi2  VTLFAVLYD  ESRTEDLSF  KGERQIRIQV  NTEDGWEAR  SLTGETQYI  
src_avisi2  VTLFAVLYD  ESRTEDLSF  KGERQIRIQV  NTEDGWEAR  SLTGETQYI  
src_avisi  VTLFAVLYD  ESRTEDLSF  KGERQIRIQV  NTEDGWEAR  SLTGETQYI  
src_avisi  VTLFAVLYD  ESRTEDLSF  KGERQIRIQV  NTEDGWEAR  SLTGETQYI  
stk_hydat  VTLFAVLYD  EARISEDLSF  KGERQIRIQV  NTEDGWEAR  SLTGETQYI  
stk_hydat  VTLFAVLYD  EARISEDLSF  KGERQIRIQV  NTEDGWEAR  SLTGETQYI  
hck_human  ..IYALVYD  EALIHEDESFP  KQGDQMVE  ES.GEWMAR  SALTREKFI  
bik_mous  ..FVALFALAY  AVNHRDLQV  LQGEKIQVLR  .STGDGEAR  SVTGETQKV  
hck_mous  ..TIVALFDF  EAIIHERDSFP  KQGDQMVLE  .AGEDMAR  SLTREKFI  
lyn_mous  ..LIVALFDF  DGINPFDLSF  KQGEKQVLE  .ERHGEWMAR  SALTREKFI  
lck_mous  ..LIVALFDF  EPSMLQDLF  KKQIERKQL  ES.GEWMAR  SLTQEGFI  
s81_yeas  ..ALYFP  DDADDEISFP  ENEIQIQVSD  .IEGEWMAR  R.ANGETQI  
abl_mous  ..LFALFDF  VASGINTLSL  TKQKIEKLV  YnGEGEMQ  .TKQOQQV  
abl_mous  ..LFALFDF  VASGINTLSL  TKQKIEKLV  YnGEGEMQ  .TKQOQQV  
srl1_drome  ..VVALFDF  KSRDDSDELSP  KQGDQMEIVD  DTEGWEAR  NLTRIKQELI  
myd_didi  ..ALYDF  DAASMLSF  KQDGEIILV  .SSGDMAVDE  LS.KQGKVY  
yfj4_yeas  ..ALYFP  AGESGDFLDF  KQGDVTRKLV  kQGNDGWGR  V..NGERGF  
abl2_human  ..LFALFDF  VASGINTLSL  TKQKIEKLV  YQNGEMQ  RSKG.QKV  
tec_human  ..EIVVQFD  QAADGDLAL  ERQKYLLE  KNDHVEAN  D.KYNEQGI  
tsk_hum  ..ALYDF  LPERCNLAL  RAEEYSLIE  KYNHPWARM  D.RLNGELI  
haa2_yeasvRVARVLYD  TTNEMDLSP  RKGDGTVLR  YQVRDMQVA  L..RNGMZF  
abpl_sacex  ..ALYDY  EAGEINETF  AHNDKINIE  FVDDWALGE  LETTGQKLF  

© Burkhard Rost

ROSTLAB

7/96
3D prediction from first principles?
Goal of structure prediction

Epstein & Anfinsen, 1961: sequence uniquely determines structure

- **INPUT:** sequence
- **OUTPUT:** 3D structure and function
protein folding from first principles should then be possible
Protein structure prediction problem* solved!

- 60s - Washington Post
- 70s - New York Times
- 90s - Washington Post

*Problem:
“predict the 3D structure of a protein from sequence alone”
How would you assess prediction performance?
CASP: how it works

- Critical Assessment of Structure Prediction
- April-May (Organizers):
  collect experimental structures
  (since 2004 from structural genomics)
- June-August: Prediction season
  deadline: predictions in before experimental
  structures are published
- September-November: Assessors divine
- December: Meeting to discuss results
Protein Structure Prediction

- Only homology modeling good
- No general prediction of 3D from sequence, yet
- Important improvements in many fields!
Servers, META-servers, META-META, …
3D from experimental co-ordinates
3D details - 3D cartoon
Structure by experiment
Experiments determine protein structure

<table>
<thead>
<tr>
<th>Method</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>120,817</td>
<td>100</td>
</tr>
<tr>
<td>Xray</td>
<td>109,064</td>
<td>90</td>
</tr>
<tr>
<td>NMR</td>
<td>10,366</td>
<td>9</td>
</tr>
<tr>
<td>EM (Electron Microscopy)</td>
<td>1,098</td>
<td>1</td>
</tr>
</tbody>
</table>

PDB (Protein Data Bank) 2017/05
Helen Berman (Rutgers Univ, New Brunswick) & Phil Bourne (UCSD San Diego)
Protein structure by X-ray crystallography

Myoglobin structure
* THIS 1mbo: SE Philips JMB 142:531-54
(image Wikipedia Aza Toth)
(Hemoglobin: Max Perutz 1959)

<table>
<thead>
<tr>
<th>Method</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>120,817</td>
<td>100</td>
</tr>
<tr>
<td>X-ray</td>
<td>109,064</td>
<td>90</td>
</tr>
<tr>
<td>NMR</td>
<td>10,366</td>
<td>9</td>
</tr>
<tr>
<td>EM</td>
<td>1,098</td>
<td>1</td>
</tr>
</tbody>
</table>

© Wikipedia
Protein structure by NMR* spectroscopy

* NMR: Nuclear Magnetic Resonance

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>120,817</td>
<td>100</td>
</tr>
<tr>
<td>Xray</td>
<td>109,064</td>
<td>90</td>
</tr>
<tr>
<td>NMR</td>
<td>10,366</td>
<td>9</td>
</tr>
<tr>
<td>EM</td>
<td>1,098</td>
<td>1</td>
</tr>
</tbody>
</table>

© Wikipedia
Protein structure by NMR* spectroscopy

* NMR: Nuclear Magnetic Resonance

<table>
<thead>
<tr>
<th>Method</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>120,817</td>
<td>100%</td>
</tr>
<tr>
<td>Xray</td>
<td>109,064</td>
<td>90%</td>
</tr>
<tr>
<td>NMR</td>
<td>10,366</td>
<td>9%</td>
</tr>
<tr>
<td>EM</td>
<td>1,098</td>
<td>1%</td>
</tr>
</tbody>
</table>

900 MHz NMR machine

NYSBC - City College New York City


© Wikipedia
Protein structure by cryo-EM

* EM: Cryo-Electron Microscopy

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>120,817</td>
<td>100</td>
</tr>
<tr>
<td>Xray</td>
<td>109,064</td>
<td>90</td>
</tr>
<tr>
<td>NMR</td>
<td>10,366</td>
<td>9</td>
</tr>
<tr>
<td>EM</td>
<td>1,098</td>
<td>1</td>
</tr>
</tbody>
</table>

4 Ångstrøm  
8Å  
16Å  
32Å  

GroEL - J Wang & DC Bosvert (2004) 1j4z  

© Wikipedia
Protein structure by cryo-EM

* EM: Cryo-Electron Microscopy

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>120,817</td>
<td>100</td>
</tr>
<tr>
<td>Xray</td>
<td>109,064</td>
<td>90</td>
</tr>
<tr>
<td>NMR</td>
<td>10,366</td>
<td>9</td>
</tr>
<tr>
<td>EM</td>
<td>1,098</td>
<td>1</td>
</tr>
</tbody>
</table>

Computer-Aided Design
39:352-60
Structure resolution

<table>
<thead>
<tr>
<th>Method</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>120,817</td>
<td>100</td>
</tr>
<tr>
<td>Xray</td>
<td>109,064</td>
<td>90</td>
</tr>
<tr>
<td>NMR</td>
<td>10,366</td>
<td>9</td>
</tr>
<tr>
<td>EM</td>
<td>1,098</td>
<td>1</td>
</tr>
</tbody>
</table>
Notation: protein structure 1D, 2D, 3D

| P | PP | P | 128 | 110 |
| Q | QQQ | Q | 175 | 97 |
| I | FFQVI | I | 70 | 60 |
| T | SSIVR | T | 77 | 69 |
| L | LLSTL | L | 120 | 14 |
| W | WWQED | W | 238 | 81 |
| Q | RKQAK | Q | 169 | 97 |
| R | RRRFPQ | R | 200 | 62 |
| P | PPPPP | P | 24 | 48 |
| L | VVTKF | L | 71 | 59 |
| V | VVLII | V | 14 | 0 |
| T | TTKKEK | T | 74 | 69 |
| I | AALIV | I | 0 | 0 |
| K | HYKKF | K | 90 | 73 |
| I | IILVI | I | 4 | 0 |
| G | EENGG | G | 46 | 41 |
| G | GGGTG | G | 62 | 53 |
| Q | QQKRR | Q | 68 | 71 |
| L | PPLNW | L | 118 | 59 |
| K | VVFKV | K | 31 | 73 |
| E | EESK | E | 124 | 95 |
| A | VVGLG | A | 1 | 0 |
| L | LLIII | L | 29 | 0 |
| L | LLLVV | L | 24 | 0 |
| D | DDDDD | D | 49 | 58 |
| T | TTTT | T | 72 | 51 |
| G | GGGGG | G | 62 | 30 |
| A | AAAAA | A | 17 | 0 |
| D | DDDDD | D | 102 | 79 |
| D | DDAKE | D | 69 | 58 |
| T | SSTTV | T | 1 | 69 |
| V | IVIVI | V | 14 | 0 |
| L | VVIVL | L | 0 | 0 |

1D 2D 3D

kcal/mol
Secondary structure stabilized by hydrogen bonds
Hydrogen-bond formation

concept introduced by Linus Pauling (Nobel Prize in 1954)

helix

strand


© Wikipedia

© Burkhard Rost
3D details - 3D cartoon
3D details - 3D cartoon
Pauling Nobel Prize 1954 – first protein structure when?
First protein structures

John Kendrew


Max Perutz


myoglobin

hemoglobin
Secondary structure from 3D structure
3D details - 3D cartoon
How to “annotate” 1D-secondary structure from 3D co-ordinates?
Secondary structure assignment

Different evaluation criteria applied:

- Assignment coverage: **DEFINE**
  Geometry (fitting ideal sec str segments)

- Enthalpic energy: **DSSP**
DSSP: Coulomb

\[ E = f\delta^+\delta^- \left( \frac{1}{r_{NO}} + \frac{1}{r_{HC'}} + \frac{1}{r_{HO}} + \frac{1}{r_{NC'}} \right) \]  \hspace{1cm} (17.2)

Figure 17.1. Distances used to calculate the Coulomb hydrogen bond.

Pauling’s H-bond pattern used in DSSP

L Pauling & RB Corey (1953) PNAS 39:247-252
L Pauling, RB Corey & HR Branson (1951) PNAS 37:205-234
Comparative modeling 1
(homology modeling)
Background

similar sequences ->
similar 3D
Goal of protein prediction

Epstein & Anfinsen, 1961: sequence uniquely determines structure

 INPUT: protein sequence
 OUTPUT: 3D structure and function
Amino acid sequence determines protein 3D structure

Christian Anfinsen
Nobel Prize in Chemistry 1972
Oncogene K-Ras

structure (PDB id 4lpk):
Oncogene K-Ras: single G12C mutation

rainbow: K-Ras 4lpk WT
red: K-Ras 4l8g G12C

structure (PDB id 4lpk-rainbow/4l8g-red):
Oncogene K-Ras / Rash

85% PIDE
(pairwise identical residues)

green: 3gft K-Ras - human
lime: 3lbn Rash - human

3gft: Y Tong et al. & H Park (unpublished)
human: K-Ras/Rash & fly: Rab6

PIDE:
pairwise identical residues

green: 3gft K-Ras - human
lime: 3lbn Rash - human 85%
orange: 2y8e Rab6 - fly 28%
human: K-Ras/Rash & fly: Rab6

green: 3gft K-Ras - human
lime: 3lbn Rash - human
orange: 2y8e Rab6 - fly

$\approx 85\%$ $\approx 28\%$

PIDE:
pairwise identical residues

3gft: Y Tong et al. & H Park (unpublished)
human - fly - bacteria

green: 3gft K-Ras - human
lime: 3lbn Rash - human
orange: 2y8e Rab6 - fly
purple: 2y8e hydroxylase
        P putida

PIDE:
pairwise identical residues

3gft: Y Tong et al. & H Park (unpublished) / 4IW3: JS Scotti (unpublished)
Comparative modeling task

• start with sequence U
• find structure in PDB that matches

Christine Orengo 1997 *Structures* 5 1093-1108
How to use this to predict 3D structure?
How to answer the question?

How to use this to predict 3D structure?

form groups and answer together in 3 min
Comparative modeling: rough idea

Query  protein sequence
Comparative modeling: rough idea

Query: protein sequence

PDB
Comparative modeling: rough idea

<table>
<thead>
<tr>
<th>Query</th>
<th>Protein sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>Protein sequence</td>
</tr>
</tbody>
</table>

Comparative modeling: rough idea

Query  protein sequence

PDB  protein sequence

structure (PDB id 4lpk):
“Reach” of comparative modeling
whole pie=all protein sequences known

- Experimental: 13%
- HoMo: 22%
-FoRc: 20%
- 1D: 44%
Comparative modeling: concepts
Comparative modeling: words

- Comparative modeling vs. Homology modeling

Lingo:

- Target: protein to model
- Template: protein to model from
Comparative modeling: steps

- Identify template database search (one or many)
How to identify template?
Comparative modeling: steps

☐ Identify template
database search (one or many)

☐ Align target/template
Comparative modeling: steps

- Identify template
  database search (one or many)
- Align target/template
- Build model
Comparative modeling: steps

- Identify template
  - database search (one or many)
- Align target/template
- Build model
- Assess model
How to assess model?
Assessing the model

cannot stretch the green to the orange
Comparative modeling: steps

- Identify template
database search (one or many)
- Align target/template
- Build model
- Assess model
- (refine)
Refine the model

if green were stretched: move it back, or …
Comparative modeling: steps

- Identify template
  - database search (one or many)
    - PSI-BLAST
    - HHblits
    - threading
- Align target/template
- Build model
- Assess model
- (refine)
Comparative modeling: quality

**Percentage of pairwise identical residues**
- 100%
- 75%
- 50%
- 25%
- 0%

**Limiting factor in homology modelling**
- SPEED of modelling
- QUALITY of model
- ALIGNMENT accuracy
- DETECTION of homology

Increasing accuracy
Increasing coverage
Comparative modeling: steps

☐ Identify template /database search
  • PSI-BLAST / HHblits / threading

☐ Align target/template
  • dynamic programming
  • structural alignment of targets to template-relatives
  • threading-like
  • profile-profile

☐ Build model
☐ Assess model
☐ (refine)
Comparative modeling: State-of-the-art methods
Modeller
Sali lab
UCSF
Comparative modeling: MODELLER

MODELLER: overview

N Eswar et al. & A Sali (2008) Methods Mol Biol 426: 145-59 (Fig. 1)
MODELLER: constraint satisfaction

1. Align sequence with structures
2. Extract spatial restraints
3. Satisfy spatial restraints

Source: MODELLER manual

MODELLER: constraint satisfaction: fits

\[ p(x_1 \leq x < x_2) = \int_{x_2}^{x_1} p(x) \, dx \]

with

\[ \int p(x) \, dx = 1 \]

\[ p(x) > 0 \]

Source: MODELLER manual

MODELLER: constraint satisfaction

- Find the model with highest probability

Variable Target function:
- Start model close to the template conformation
- First only local constraints
- Minimize using conjugate gradient optimization
- Repeat, introduce more and more long-range constraints
Run optimization repeatedly
Starting point: template coordinates with random fluctuations
Explore different local minima

MODELLER: multiple models

N Eswar et al. & A Sali (2008) Methods Mol Biol 426: 145-59 (Fig. 3)
MODELLER: typical errors

side chain packing

mis-alignment

wrong template

N Eswar et al. & A Sali (2006) Current Protocols in Bioinformatics: Chapter 5 - Unit 5.6.1-30 (Fig. 5.6.12)
MODELLER: Identify best models

DOPE score
Discrete Optimized Protein Energy

Based on knowledge-based pair potentials

What if the loop is missing?

EEE B B B B EEEEE EEEEEE EEEEEEEHHHEEE

1shf 100% VTLFVALYDYEARTEDDLFSHKGEKFQIGINSEGDEWAVEARSLTGETGYIPSNYAPVD
1srm 78% VTTFVALYDYESRTETDLSFKGERLQIVNTEGDDWLAHLTTGQTGYIPSNYVAPSD
1sem 39% ....VAEHDFQAGSPDELSFKRGNLKVKNKEDPHWYKAEL.DGNEGFIPLSNYIRMTE
Similar loops

http://www.cmbi.ru.nl/~hvensela/EGFR-verslag/spitz.html

blue: EGF (human)
red: Spitz (fly)
What if loop completely missing?

What if loop completely missing?

Molecular dynamics
MODELLER: loop modeling

Andras Fiser, Richard Kinh Gian Do & Andrej Sali (2000) Protein Science 9:1753-73: Fig. 9

Fig. 9. Accuracy of loop modeling in the correct environment as a function of loop length. Models were calculated for 40 loops at each length from 1 to 14 residues, as described in Theory and algorithms. Fifty independent optimizations were used to make each prediction. Average accuracy and the standard deviation of the accuracy are shown for each length for (A) local and (B) global superposition.
Comparative modeling methods

☐ MODELLER
   lots of whistles and bells
   downloadable
   very accurate

☐ SWISS-MODEL
SWISS-MODEL
Schwede lab
Basel
Comparative modeling: SWISS-MODEL

Comparative modeling: SWISS-MODEL

- Underlying “philosophy”:
  fully automated
  for non-expert users/experimental biologists
  do less -> you do fewer mistakes

- original:
  1. alignment by BLAST/PSI-BLAST
  2. copy co-ordinates
  3. end

Comparative modeling: SWISS-MODEL

Things got more complicated ...

Lorenza Bordoli, Florian Kiefer, Konstantin Arnold, Pascal Benkert, James Battey & Torsten Schwede (2009) Nature Protocols doi: 10.1038/nprot.2008.197; Fig. 2
Comparative modeling methods

□ MODELLER
lots of whistles and bells, downloadable, very accurate

□ SWISS-MODEL
automated, increasingly comprehensive and flexible
Show me the mistake of your method

René Magritte (1889-1967)

Margritte: Treachery of Images, 1929

This is not a pipe.

Portrait of Magritte by Lothar Wolleh, 1967
Models and reality

“A Model must be wrong, in some respects, else it would be the thing itself. The trick is to see where it is right.” (Henry A. Bent)

“A model is a tool that helps to interpret biochemical data.” (Torsten Schwede)

René Margritte (1889-1967)
# Lecture plan (CB1 structure: INF)

<table>
<thead>
<tr>
<th>Date</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
</tr>
</thead>
<tbody>
<tr>
<td>01/04/10</td>
<td></td>
<td>No lecture</td>
<td></td>
</tr>
<tr>
<td>02/04/12</td>
<td></td>
<td>No lecture</td>
<td></td>
</tr>
<tr>
<td>03/04/17</td>
<td></td>
<td>No lecture</td>
<td></td>
</tr>
<tr>
<td>04/04/19</td>
<td>Intro 1: organization of lecture: intro into cells &amp; biology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>05/04/24</td>
<td>Intro 2: amino acids, protein structure (comparison), domains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>06/04/26</td>
<td></td>
<td>No lecture</td>
<td></td>
</tr>
<tr>
<td>07/05/01</td>
<td></td>
<td>SKIP: May Day</td>
<td></td>
</tr>
<tr>
<td>08/05/03</td>
<td></td>
<td>Alignment 1</td>
<td></td>
</tr>
<tr>
<td>09/05/08</td>
<td></td>
<td>SKIP: Student Representation (SVV)</td>
<td></td>
</tr>
<tr>
<td>10/05/10</td>
<td></td>
<td>SKIP: Ascension Day</td>
<td></td>
</tr>
<tr>
<td>11/05/15</td>
<td></td>
<td>Alignment 2</td>
<td></td>
</tr>
<tr>
<td>12/05/17</td>
<td></td>
<td>Comparative modeling &amp; exp structure determination &amp; secondary structure assignment</td>
<td></td>
</tr>
<tr>
<td>13/05/22</td>
<td></td>
<td>SKIP: Whitsun holiday</td>
<td></td>
</tr>
<tr>
<td>14/05/24</td>
<td></td>
<td>1D: Secondary structure prediction 1</td>
<td></td>
</tr>
<tr>
<td>15/05/29</td>
<td></td>
<td>1D: Secondary structure prediction 3</td>
<td></td>
</tr>
<tr>
<td>16/05/31</td>
<td></td>
<td>SKIP: Corpus Christi</td>
<td></td>
</tr>
<tr>
<td>17/06/05</td>
<td></td>
<td>1D: Transmembrane structure prediction 1</td>
<td></td>
</tr>
<tr>
<td>18/06/07</td>
<td></td>
<td>1D: Transmembrane structure prediction 2 / Solvent accessibility prediction</td>
<td></td>
</tr>
<tr>
<td>19/06/12</td>
<td></td>
<td>1D: Transmembrane structure prediction 3 / Solvent accessibility prediction</td>
<td></td>
</tr>
<tr>
<td>20/06/14</td>
<td></td>
<td>1D: Disorder prediction</td>
<td></td>
</tr>
<tr>
<td>21/06/19</td>
<td></td>
<td>2D prediction / 3D prediction</td>
<td></td>
</tr>
<tr>
<td>22/06/21</td>
<td></td>
<td>No lecture</td>
<td></td>
</tr>
<tr>
<td>23/06/26</td>
<td></td>
<td>recap 1</td>
<td></td>
</tr>
<tr>
<td>24/06/28</td>
<td></td>
<td>recap 2</td>
<td></td>
</tr>
<tr>
<td>25/07/03</td>
<td></td>
<td>TBA</td>
<td></td>
</tr>
<tr>
<td>26/07/05</td>
<td></td>
<td>TBA</td>
<td></td>
</tr>
<tr>
<td>27/07/10</td>
<td></td>
<td>TBA</td>
<td></td>
</tr>
<tr>
<td>28/07/12</td>
<td></td>
<td>TBA</td>
<td></td>
</tr>
</tbody>
</table>