title: 1D: beta membrane, accessibility

short title: cb1_1d_tmb_acc

lecture: Computational Biology 1 - Protein structure (for Informatics) - TUM summer semester
Videos: YouTube / www.rostlab.org

THANKS:

Special lectures:
- 06/20 Michael Bernhofer
- 07/xx Predrag Radivojac - Indiana Univ.
- 06/xx Yana Bromberg - Rutgers Univ.

No lecture:
- 05/09 no lecture
- 05/23 Student assembly (SVV)
- 05/25 Ascension day
- 06/06 Whitsun holiday
- 06/15 Corpus Christi

LAST lecture:
bef: Jul 11
after: Jul 28

Examen:
- Makeup: WEDNESDAY(!!) July 12: 18:00-19:30 TBA

EXERCISES:

CONTACT: Lothar Richter richter@rostlab.org

Announcements

Dmitrij Nechaev

Christian Dallago

Lothar Richter

Examen: WEDNESDAY(!!) July 12: 18:00-19:30 TBA
- Makeup: TBC: Oct 17 & 19, 2017 - lecture time
Recap
Lipid bilayer
Hydrophobic core of a protein
Cytoplasm (stromal side)
TMH proteins: reminders

Edda Kloppmann & Marco Punta:

1,035 PDB unique TM structures (Jan 2012) 
-> 107 Pfam families

i.e. although about 15-25% of all proteins 
AND 60% of all drug targets 
only about 2% of all “unique” structures have 
membrane helices
... but, prediction methods for 1D very successful!
TMH prediction 1D+

PHDhtm/PROFhtm

TopPred
G von Heijne (1992)
JMB 225: 487-94

TMHMM
A Krogh, B Larsson, G von Heijne, EL Sonnhammer (2001) JMB 305:567-80

© Burkhard Rost
ROSTLAB.TUM
Eight best HTM's
μ=0: 0 HTM
μ=1: 1 HTM
μ=2: 2 HTM
μ=3: 3 HTM

Loop lengths
Charge:
Number of R+K in loops 1-4

final prediction:
Δ = (5+1) - (2+3) > 0
=> first loop out

Lipid membrane bilayer
Extra-cytoplasmic
Intra-cytoplasmic
1D: TM-beta
Beta-barrel predictions
Predicting transmembrane beta barrels

Extracellular

Sucrose Specific Porin  Maltoporin  OmpF Matrix Porin

FhuA receptor  FepA active transporter  porin from R. Blastica

Phospholipase A  OmpX  OmpA  porin from R. Capsulatis

Computer Simulation of the Rough Lipopolysaccharide Membrane of *Pseudomonas aeruginosa*

Biophys J, August 2001, p. 1037-1046, Vol. 81, No. 2

H Bigelow, D Petrey, J Liu, D Przybylski & B Rost (2004) *NAR* 32, 2566
PROFtmb: Structure-based labels

“loop” out

“loop” in

Legend:

I: periplasmic hairpins
Q: extracellular loops
U[A-Z]: upward strand, facing inward
U[a-z]: upward strand, facing toward
D[a-z]: downward strand, facing inward
D[A-Z]: downward strand, facing towards bilayer

Henry Bigelow & BR Columbia Univ
© Burkhard Rost
Arrows denote allowed transitions in the HMM. Dotted arrow region indicates one connection per enclosed state.
model seems to make sense, but is it “right”?
Answer the question in groups

How to assess whether model makes sense?

© Wikipedia
Amino Acid Statistics
put observation into the priors for the HMM, train for all others
## TMB proteins 2004: structures & functions

<table>
<thead>
<tr>
<th>Pdb</th>
<th>Ref.</th>
<th>Name</th>
<th>Source</th>
<th>#str</th>
<th>N</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a0s</td>
<td>NSB 5(1) pp. 37-46 Jan 98</td>
<td>Sucrose Specific Porin</td>
<td>S. Typh</td>
<td>18</td>
<td>3</td>
<td>rapid transport of sucrose</td>
</tr>
<tr>
<td>1af6</td>
<td>JMB 272 pp. 56-63 '97</td>
<td>Maltoporin</td>
<td>E. Coli</td>
<td>18</td>
<td>3</td>
<td>selective transport of glucose</td>
</tr>
<tr>
<td>1bt9</td>
<td>Biochem 37 pp. 15663 '98</td>
<td>OmpF Matrix Porin</td>
<td>E. Coli</td>
<td>16</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1by5</td>
<td>Cell 95 pp. 771 '98</td>
<td>FhuA receptor</td>
<td>E. Coli</td>
<td>22</td>
<td>1</td>
<td>import of ferrichrome, signal transduction</td>
</tr>
<tr>
<td>1fep</td>
<td>NSB 6 pp 56 '99</td>
<td>FepA active transporter</td>
<td>R. Blastica</td>
<td>16</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1prn</td>
<td>JMB 243 pp. 891 '94</td>
<td>'porin'</td>
<td>E. Coli</td>
<td>12</td>
<td>1</td>
<td>deacylates LPS, perforating foreign membranes</td>
</tr>
<tr>
<td>1qd5</td>
<td>Nature 401 pp. 717 '99</td>
<td>phospholipase A</td>
<td>E. Coli</td>
<td>8</td>
<td>1</td>
<td>binds β-strand-containing foreign proteins</td>
</tr>
<tr>
<td>1qj9</td>
<td>Structure 7 pp. 1301 '99</td>
<td>OmpX</td>
<td>E. Coli</td>
<td>8</td>
<td>1</td>
<td>docking points for bacteriophage</td>
</tr>
<tr>
<td>1qjp</td>
<td>JMB 298 pp 273 '00</td>
<td>OmpA membrane domain</td>
<td>E. Coli</td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3por</td>
<td>JMB 231 pp. 817 '93</td>
<td>'porin'</td>
<td>R. Capsulatis</td>
<td>16</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
## PROFtmb: per-residue results

<table>
<thead>
<tr>
<th>Method</th>
<th>Data</th>
<th>$Q_2$</th>
<th>$Q_\beta^{% prd}$</th>
<th>$Q_\beta^{% obs}$</th>
<th>C</th>
<th>SOV$_2$</th>
<th>SOV$_\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated $\sigma$</td>
<td></td>
<td>±6</td>
<td>±9</td>
<td>±14</td>
<td>±0.11</td>
<td>±13</td>
<td>±11</td>
</tr>
<tr>
<td>Martelli</td>
<td>SetTMBcomp</td>
<td>84</td>
<td>87</td>
<td>80</td>
<td>0.69</td>
<td>91</td>
<td>94</td>
</tr>
<tr>
<td>PROFtmb</td>
<td>SetTMBcomp</td>
<td>83</td>
<td>87</td>
<td>80</td>
<td>0.69</td>
<td>79</td>
<td>93</td>
</tr>
<tr>
<td>PROFtmb</td>
<td>SetTMB</td>
<td>83</td>
<td>84</td>
<td>85</td>
<td>0.70</td>
<td>87</td>
<td>92</td>
</tr>
<tr>
<td>PROFtmb</td>
<td>SetTMBfull</td>
<td>86</td>
<td>86</td>
<td>85</td>
<td>0.74</td>
<td>88</td>
<td>94</td>
</tr>
</tbody>
</table>

### Predicted vs. Observed

<table>
<thead>
<tr>
<th></th>
<th>up-strand</th>
<th>down-strand</th>
<th>peri-loop</th>
<th>outer-loop</th>
<th>SUM</th>
<th>Pok</th>
</tr>
</thead>
<tbody>
<tr>
<td>up-strand</td>
<td>563</td>
<td>9</td>
<td>27</td>
<td>57</td>
<td>656</td>
<td>86</td>
</tr>
<tr>
<td>down-strand</td>
<td>5</td>
<td>594</td>
<td>58</td>
<td>33</td>
<td>690</td>
<td>86</td>
</tr>
<tr>
<td>peri-loop</td>
<td>39</td>
<td>20</td>
<td>763</td>
<td>13</td>
<td>835</td>
<td>91</td>
</tr>
<tr>
<td>outer-loop</td>
<td>77</td>
<td>104</td>
<td>2</td>
<td>790</td>
<td>973</td>
<td>81</td>
</tr>
<tr>
<td>SUM</td>
<td>684</td>
<td>727</td>
<td>850</td>
<td>893</td>
<td>3154</td>
<td>86</td>
</tr>
<tr>
<td>Pok</td>
<td>82</td>
<td>82</td>
<td>90</td>
<td>88</td>
<td>86</td>
<td>86</td>
</tr>
</tbody>
</table>
Remarkable performance, but: can we distinguish proteins with and without TMB?
Testing 2: per protein performance
Testing 2: per protein performance

where to put threshold?
Testing 2: per protein performance

Whole Protein Discrimination
set_SWISS_locexp

Fraction Accuracy (or Coverage)

Bits Score

Accuracy
Coverage
1D: solvent accessibility
Get accessibility from 3D structures
Defining residue solvent accessibility
Different ways to cast value into accessibility

- absolute accessibility
  $ASA = \text{square } \text{Ångström} (1\text{Å}=0.1\text{nm})$
Joining amino acids into proteins

```
side-chain

backbone

R

C

H

N

H

C

OH

O
```

© Burkhard Rost
Joining amino acids into proteins

polypeptide chain
Joining amino acids into proteins

\[
\begin{align*}
\text{backbone:} & \quad \text{H} - \text{N} - \text{C} - \text{H} - \text{C} - \text{OH} \\
\text{side-chain:} & \quad \text{R} 
\end{align*}
\]
Some residues have short side-chains, others long ones
Some residues have short side-chains, others long ones.
Some residues have short side-chains, others long ones

A = short side chain

K = long side chain
Long side chains may appear more accessible
Different ways to cast value into accessibility

- **Absolute accessibility**
  
  $ASA = \text{square Ångstrøm (1Å=0.1nm)}$

- **Relative accessibility**
  
  $ASA/\text{max ASA}$

100% = max ASA
Different ways to cast value into accessibility

- **absolute accessibility**
  \[ ASA = \text{square Ångström} \ (1Å=0.1\text{nm}) \]

- **relative accessibility**
  \[ ASA/\text{max ASA} \]

- **“states”**
  - buried, exposed
  - buried, intermediate, exposed

- what is best?
Different ways to cast value into accessibility

- **absolute accessibility**
  
  \[ \text{ASA} = \text{square } \text{Ångstrøm} \ (1\text{Å} = 0.1\text{nm}) \]

- **relative accessibility**
  
  \[ \frac{\text{ASA}}{\text{max ASA}} \]

- **“states”**
  
  - buried, exposed
  - buried, intermediate, exposed

what is best??
100% and 80% more similar than 20 and 0
100% and 80% more similar than 20 and 0

how to reflect this in the design of a prediction method?
Simple function realizing objective

“States” to predict

Percentage solvent accessibility
Predict solvent accessibility
historically: by hydrophobicity
**PHDace**

- **local alignment**
  - AAA
  - AA
  - LLL
  - LII
  - AAG
  - CCS
  - GWV

- **global statistic**
  - %AA
  - Length
  - ΔN-term
  - ΔC-term

**input local in sequence**

```
A C L I G S V ins del cons
100 0 0 0 0 0 0 0 0 1.17
100 0 0 0 0 0 0 33 0 0.42
0 0 100 0 0 0 0 33 0 0.92
0 0 33 66 0 0 0 0 0 0.74
66 0 0 0 33 0 0 0 0 1.17
0 66 0 0 0 33 0 0 0 0.74
0 0 0 33 0 0 66 0 0 0 0.48
```

**input global in sequence**

- Percentage of each amino acid in protein:
  - length of protein: (≤60, ≤120, ≤240, >240)
  - distance: centre, N-term: (≤40, ≤30, ≤20, ≤10)
  - distance: centre, C-term: (≤40, ≤30, ≤20, ≤10)

© Burkhard Rost

45/55
Correct prediction of correctly predicted residues

![Graph showing the relationship between overall per-residue accuracy and the percentage of residues predicted.]
Evolution for accessibility prediction

- Detailed prediction problematic

- Significant gain by evolutionary information:
  in/out with > 75% accuracy!
What to get from predictions?
Residue conservation on surface

ConSurf

Cavities

Cavity: but NOT conserved
Conserved cavities: different substrates
Accessibility helps in predicting protein function

- Sub-cellular localization
- Protein-protein interactions
- Flexibility/motion from structure
More globular - more likely expressed?

- Proteins
- Fragments

- Decreasing 'globularity'

Domain and chain representations showing a decrease in globularity from proteins to fragments.
Proteins are amazingly cubic ...
Lecture plan (CB1 structure)

- **01**: 04/25 Tue: no lecture
- **02**: 04/27 Thu: no lecture
- **03**: 05/02 Tue: Intro 1: organization of lecture: intro into cells & biology
- **04**: 05/04 Thu: Intro 2: amino acids, protein structure (comparison), domains
- **05**: 05/09 Tue: No lecture
- **06**: 05/11 Thu: Alignment 1
- **07**: 05/16 Tue: Alignment 2
- **08**: 05/18 Thu: Comparative modeling & exp structure determination & secondary structure assignment
- **09**: 05/23 Tue: SKIP: student assembly (SVV)
- **10**: 05/25 Thu: SKIP: Ascension Day
- **11**: 05/30 Tue: 1D: Secondary structure prediction 1
- **12**: 06/01 Thu: 1D: Secondary structure prediction 2
- **13**: 06/06 Tue: SKIP: Whitsun holiday (06/03-06)
- **14**: 06/08 Thu: 1D: Secondary structure prediction 3
- **15**: 06/13 Tue: 1D: Transmembrane structure prediction 1
- **16**: 06/15 Thu: SKIP: Corpus Christi
- **17**: 06/20 Tue: 1D: Transmembrane structure prediction 2 / Solvent accessibility prediction
- **18**: 06/22 Thu: 1D: Disorder prediction
- **19**: 06/27 Tue: 2D prediction
- **20**: 06/29 Thu: 3D prediction / Nobel prize symposium
- **21**: 07/04 Tue: TBA
- **22**: 07/06 Thu: recap 1
- **23**: 07/11 Tue: recap 2
- **24**: 07/12 Thu: examen
- **25**: 07/13 Tue: TBA
- **26**: 07/18 Thu: TBA
- **27**: 07/20 Tue: TBA
- **28**: 07/22 Thu: TBA