Announcements

Videos: SciVee
www.rostlab.org

THANKS:
Tim Karl + Manfred Roos

NO lectures:
Tue May 31(!)
studentische vollversammlung
Thu Jun 16

LAST lecture: Jul 7
Examen: Jul 12 (?), 10:30 (likely this room)
  Makeup: likely: October 13 - morning
Planned TOC

☐ Weeks 1-2 (May 13)
  Intro: proteins/protein structure/protein structure comparison
  Exercises (05/05 + 05/12):
  hydrophobicity clustering + CATH/SCOP

☐ Weeks 3-4 (Jun 9):
  3D->1D
  Secondary structure prediction
Protein Prediction - Part 1: Structure

1 Introduction (contd.)
Marco Punta contributed the slides

- PhD in Trieste (MD for membrane proteins)
- Postdoc @ Columbia Univ in the City of New York (contact predictions)
- Senior scientist in NYCOMPS (Target selection for membrane proteins)
- IAS Fellow @ TUM
- Project manager @ Pfam @ Sanger Inst. Hinxton (Cambridgeshire)
TOC today

☐ LAST WEEK
  • amino acids
  • protein structure
  • bonds & energies

☐ TODAY
  • structure comparisons

☐ NEXT WEEK (tuesday)
  • secondary structure assignment
  • secondary structure prediction
Notation: protein structure 1D, 2D, 3D

<table>
<thead>
<tr>
<th></th>
<th>1D</th>
<th>2D</th>
<th>3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>128</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>70</td>
<td>E</td>
<td>60</td>
</tr>
<tr>
<td>T</td>
<td>77</td>
<td>E</td>
<td>69</td>
</tr>
<tr>
<td>L</td>
<td>120</td>
<td>E</td>
<td>14</td>
</tr>
<tr>
<td>W</td>
<td>238</td>
<td>E</td>
<td>81</td>
</tr>
<tr>
<td>R</td>
<td>169</td>
<td>E</td>
<td>97</td>
</tr>
<tr>
<td>P</td>
<td>200</td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>L</td>
<td>24</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>V</td>
<td>71</td>
<td>E</td>
<td>59</td>
</tr>
<tr>
<td>T</td>
<td>14</td>
<td>E</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>74</td>
<td>E</td>
<td>69</td>
</tr>
<tr>
<td>I</td>
<td>90</td>
<td>E</td>
<td>73</td>
</tr>
<tr>
<td>K</td>
<td>4</td>
<td>E</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>46</td>
<td>E</td>
<td>41</td>
</tr>
<tr>
<td>G</td>
<td>62</td>
<td>E</td>
<td>53</td>
</tr>
<tr>
<td>Q</td>
<td>68</td>
<td>E</td>
<td>71</td>
</tr>
<tr>
<td>L</td>
<td>118</td>
<td>E</td>
<td>59</td>
</tr>
<tr>
<td>V</td>
<td>31</td>
<td>E</td>
<td>73</td>
</tr>
<tr>
<td>E</td>
<td>124</td>
<td>E</td>
<td>95</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>E</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>29</td>
<td>E</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>24</td>
<td>E</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>49</td>
<td>E</td>
<td>58</td>
</tr>
<tr>
<td>T</td>
<td>72</td>
<td></td>
<td>51</td>
</tr>
<tr>
<td>G</td>
<td>62</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>A</td>
<td>17</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>102</td>
<td></td>
<td>79</td>
</tr>
<tr>
<td>D</td>
<td>69</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>T</td>
<td>1</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>V</td>
<td>14</td>
<td>E</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>0</td>
<td>E</td>
<td>0</td>
</tr>
</tbody>
</table>

kcal/mol
3D classifications
Structural universe

B Rost 1998 *Structure* 6:259-263
Structural universe: no islands, really

B Rost 1998 *Structure* 6:259-263
Structural universe: no islands, really
3D classifications: goals

- Similar 3D -&gt; Similar function
- Learn from 3D about function
- Learn about evolution

classify
3D modules

Multiple 3D alignment identifies consensus secondary structure

© Christine Orengo
Fold of a protein

- some structures more often observed than others
- limited number of shapes?
- fold remains an assumption (that increasingly seems to be proven inappropriate)

Multiple 3D alignment identifies consensus secondary structure
Protein structure comparisons

All-alpha

All-beta

AlphaBeta

3sdh

1bww

1xne

Thursday May 12, 2011
3D classification databases

☐ SCOP
http://scop.mrc-lmb.cam.ac.uk/scop/
[A Murzin et al. (1995) JMB 247, 536-540]

☐ CATH
http://www.cathdb.info/

☐ COPS
http://cops.services.came.sbg.ac.at
[SJ Suhrer et al. (2009) NAR 37, W539-W44. ]
Classify protein structure: SCOP
Alexei Murzin

- Cambridge University, England
- CASP assessor
- ~90 publications
- 1 with over 3,000 quotes
- 13 with over 100 quotes (ISI 2011/05)
- H-index: 30 (ISI 2011/05)
3D classification databases

- SCOP
  http://scop.mrc-lmb.cam.ac.uk/scop/
  [Murzin et al. J. Mol. Biol. 247, 536-540]

- hierarchy
Protein structure comparisons

All-alpha
3sdh

All-beta
1bww

AlphaBeta
1xne
SCOP hierarchy

Example

{All-alpha} a. class

Structure similarity increases
SCOP hierarchy

Example

{All-alpha}

a.

class

a.1

fold

Structure similarity increases
Structural universe: no islands, really

B Rost 1998 Structure 6:259-263

Thursday May 12, 2011
Structural universe: no islands, really

B Rost *1998 Structure* 6:259-263

Thursday May 12, 2011
SCOP hierarchy

CLASS = alpha and beta (a/b)

NAD(P)-binding Rossmann-fold domains 1sw0-TIM

1sw0-TIM beta/alpha barrel
SCOP hierarchy

Example

{All-alpha}

{Globin-like}

{alpha-helical ferrodoxin}

a.

a.1

a.1.2

class

fold

superfamily

Structure similarity increases
SCOP hierarchy

CLASS = alpha and beta (a/b)
FOLD   = TIM beta/alpha-barrel
SCOP hierarchy

Example

{Alpha and beta a/b}  c  class

{TIM beta/alpha-barrel}  c.1  fold

{Triosephosphate isomerase}  c.1.1  superfamily

{Triosephosphate isomerase}  c.1.1.1  family

(sequence based)

Structure similarity increases
SCOP hierarchy

TRIOSEPHOSPHATE ISOMERASE (1swo)

QUINOLINIC ACID PHOSPHORIBOSYLTRANSFERASE (1qap)

PHOSPHATE ALDOLASE (1p1x)
3D classification databases

- SCOP
  http://scop.mrc-lmb.cam.ac.uk/scop/
  [Murzin et al. J. Mol. Biol. 247, 536-540]
Classify protein structure: SCOP
3D classification databases

- SCOP
  http://scop.mrc-lmb.cam.ac.uk/scop/
  [Murzin et al. J. Mol. Biol. 247, 536-540]

- CATH
  http://www.cathdb.info/
Christine A. Orengo

- UCL, England
- CASP assessor
- over 120 publications
- 1 with over 1,500 quotes
- 16 with over 100 quotes
- H-index >40 (ISI 2011/05)

- SSAP (with Willy Taylor)
- CATH
Dame Janet M. Thornton

- Director
- **EBI (European Bioinformatics Institute, Hinxton, Cambridgeshire, England)**
- BS Physics (Univ Nottingham), MS Biophysics King’s College London, PhD Biophysics UCL
- Amongst Top 100 scientists in UK
- ~400 publications
- 1 with over 11,000 quotes
- 7 with over 1,000 quotes
- 81 with over 100 quotes
- H-index >88 (ISI 2011/05)
Class
Architecture
Topology
Homology
Class:
mostly alpha, mostly beta, mixed alpha/beta, few regular secondary structure

- All-alpha
- All-beta
- AlphaBeta

3sdh  1bww  1xne
Class: mostly alpha, mostly beta, mixed alpha/beta, few regular secondary structure

Architecture: classification according to overall shape, ignoring connectivity

Topology: fold groups = shape & connectivity

Homology: evolutionarily related superfamily
CATH stats (2011/05)

- folds: 1,282
- superfamilies: 2,549
- sequence families: 11,330
- domains: 24,232

Thursday May 12, 2011
CATH: steps involved

- Find domain

Multiple 3D alignment identifies consensus secondary structure
CATH: steps involved

- Find domain
- From domain to superfamily

PDB id: 1gcq
(SH3 domains)

© CATH tutorial (www.cathdb.info)
CATH: steps involved

- Find domain
- From domain to superfamily

PDB id: 1gcqA0
(SH3 domain)

http://www.cathdb.info/domain/1gcqA00

© CATH tutorial (www.cathdb.info)
CATH: steps involved

PDB id: 1gcqA0
(SH3 domain)

http://www.cathdb.info/domain/1gcqA00

© CATH tutorial (www.cathdb.info)
### CATH: Architecture: Roll

**CATH Domain: 1gcqA00**

**PDB 1gcq, Chain A, Domain 0**

<table>
<thead>
<tr>
<th>CATH Code</th>
<th>Level Description</th>
<th>Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Mainly Beta</td>
<td></td>
</tr>
<tr>
<td>2.30</td>
<td>Roll</td>
<td></td>
</tr>
<tr>
<td>2.30.30</td>
<td>SH3 type barrels.</td>
<td></td>
</tr>
<tr>
<td>2.30.30.40</td>
<td>SH3 Domains</td>
<td>Gene3D</td>
</tr>
<tr>
<td>2.30.30.40.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.30.30.40.22.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.30.30.40.22.1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.30.30.40.22.1.1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.30.30.40.22.1.1.2.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

http://www

© CATH tutorial (www.cathdb.info)
3D classification databases

- **SCOP**
  
  http://scop.mrc-lmb.cam.ac.uk/scop/
  
  [Murzin et al. J. Mol. Biol. 247, 536-540]

- **CATH**
  
  http://www.cathdb.info/
  
Classify protein structure: COPS
Manfred J. Sippl

- CAME, Univ. Salzburg
- CASP assessor
- over 54 publications
- 1 with over 800 quotes
- 10 with over 100 quotes
- H-index >27 (ISI 2011/05)

- “Sippl” potentials of pairwise energies
  (“Knowledge-based potentials”)
PDB updates 2008/08/19-2009/04/14

SJ Suhrer et al. (2009) NAR 37:W539-W544
PDB diversity in light of COPS

SJ Suhrer et al. (2009) NAR 37:W539-W544
COPS domain parsing

Apaf-1
PDB id 1z6t

COPS c1z6tA1 (CARD domain) - c2a5yB1

C1z6tA2 (α/β domain) - c2a5yB
c1z6tA3 (helical domain I) - c2a5yB3
c1z6tA4 (winged-helix domain) - c2a5yB4

SJ Suhrer et al. (2009) NAR 37:W539-W544

Thursday May 12, 2011
COPS domain parsing

PDB id
1z6t-A
with 2a5y-B

SJ Suhrer et al. (2009) NAR 37:W539-W544

© Burkhard Rost (TU Munich)
3D classification databases

- SCOP
  http://scop.mrc-lmb.cam.ac.uk/scop/
  [A Murzin et al. (1995) JMB 247, 536-540]

- CATH
  http://www.cathdb.info/

- COPS
  http://cops.services.came.sbg.ac.at
  [SJ Suhrer et al. (2009) NAR 37, W539-W44. ]
3D classification databases

- **SCOP**
  - http://scop.mrc-lmb.cam.ac.uk/scop/
  - [A Murzin et al. (1995) JMB 247, 536-540]

- **CATH**
  - http://www.cathdb.info/
  - [AL Cuff et al. (2009) NAR 37, D310-314; CA Orengo et al. (1997) Structure 15, 1093-1108]

- **COPS**
  - http://cops.services.came.sbg.ac.at
  - [SJ Suhrer et al. (2009) NAR 37, W539-W44. ]
Comparing 3D structures - methods
Structure alignment

Slides taken from Patrice Koehl, UC Davis

Patrice Koehl
Structure alignment: two steps

1. Identify equivalent positions (residues that match in 3D)
2. find superposition independent of domain movements

$cRMS = 2.85 \, \AA$
$cRMS = 2.8 \, \AA$

Root mean square displacement (rmsd)

\[ \text{rmsd}(A, B) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} d_i^2} \]

- Step 1: find corresponding points in proteins A and B
- \(d(i)\) are the distances between all corresponding points (typic: C\text{alpha}, all atoms)
Structural alignment: **SSAP**

- **William R. Taylor & Christine A. Orengo**
  - Willy Taylor
  - Christine Orengo
  - © Erik Bongcam-Rudloff

- **WR Taylor & CA Orengo (1989) JMB 208:1-22**
Structural alignment: SSAP

William R. Taylor & Christine A. Orengo


Willy Taylor
@CASP7

Christine Orengo
@ISCB-Africa

Willy Taylor
NIMR MRC
London
England

Joel Sussman
Weizman Inst. Israel

Chen Keasar
Ben-Gurion Univ
Beer-Sheva Israel
Structural alignment: SSAP

- William R. Taylor & Christine A. Orengo


IDEA: use distance matrix and apply dynamic programming
Structural alignment: SSAP


Optimize:

\[ S_{ik} = \sum_{m=-n}^{m=n} \frac{a}{|d_{i,i+m}^A - d_{k,k+m}^B| + b} \]
Structural alignment: SSAP


- Optimize:

\[ S_{ik} = \sum_{m=-n}^{m=n} \frac{a}{|d_{i,i+m}^A - d_{k,k+m}^B| + b} \]

**Problem:** loss of information about direction
Structural alignment: SSAP

- Replace distances by interatomic vectors (V)

Optimize:

\[ S_{ik} = \frac{a}{|V_{ij}^A - V_{kl}^B| + b} \]
Structural alignment: SSAP

- Include sequence information (D(xy): Dayhoff)

Optimize:

\[ S_{ik} = \frac{w D_{RiRk} + a}{| V_{ij}^A - V_{kl}^B | + b} \]
Structural alignment: DALI

Liisa Holm & Chris Sander

Liisa Holm
Univ of Helsinki
Finland

Chris Sander
SKCC New York

L Holm & C Sander (1993) Protein structure comparison by alignment of distant matrices.
JMB 233:123-38
Structural alignment: DALI

- L Holm & C Sander (1993)
  JMB 233:123-38

- Distance matrix Alignment
- Algorithm: Monte Carlo on all-against-all for hexapeptides (5)
Structural alignment: VOROLIGN

- Fabian Birzele, Ralf Zimmer et al.

Structural alignment: VOROLIGN


Dynamic programming on Voronoi environments

![Diagram](image)

Nearestneighbour set after Voronoi tessellation

Similarity of neighbour sets (low level matrix)

Similarity of residues x and y (high level matrix)

F Birzele, JE Gewehr, G Csaba & R Zimmer (2006) Bioinformatics 23:e205-11: Fig. 2
Two forms of calcium-bound Calmodulin:

- Ligand free
- Complexed with trifluoperazine

Global alignment:
RMSD =15 Å /143 residues

Local alignment:
RMSD = 0.9 Å/ 62 residues

RMSD is not a metric

\[ cRMS = 2.8 \text{ Å} \]

\[ cRMS = 2.85 \text{ Å} \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Authors</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSAP</td>
<td>WR Taylor &amp; CA Orengo</td>
<td>1989 JMB 208:1-22</td>
</tr>
<tr>
<td>DALI</td>
<td>L Holm &amp; C Sander</td>
<td>1993 JMB 233:123-38</td>
</tr>
<tr>
<td>CE</td>
<td>IN Shindyalov &amp; P Bourne</td>
<td>1998 Prot Engng 1:739-47</td>
</tr>
<tr>
<td>SKAN</td>
<td>A Yan, D Petrey &amp; B Honig, unpublished</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparison of structure alignments

Rachel Kolodny, Patrice Koehl, Michael Levitt

Comprehensive Evaluation of Protein Structure Alignment Methods: Scoring by Geometric Measures
JMB 346:1173-88
Comparison of structure alignments

- R Kolodny, P Koehl & M Levitt (2004) JMB 346:1173-88 (Fig. 1A)

- Dashed lines: original method
- Solid lines: SAS measure
LAST WEEK
- amino acids
- protein structure
- bonds & energies

TODAY
- structure comparisons

NEXT WEEK (tuesday)
- secondary structure assignment
- secondary structure prediction
Announcements

Videos:  SciVee
         www.rostlab.org

THANKS:
Tim Karl + Manfred Roos

NO lectures:
Tue May 31(!)
studentische vollversammlung
Thu Jun 16

LAST lecture: Jul 7
Examen: Jul 12 (?), 10:30 (likely this room)
  Makeup: likely: October 13 - morning

CONTACT: Marlena Drabik assistant@rostlab.org

Thursday May 12, 2011