title: Intro protein structure: 3d comparison
short title: pp1_intro_3d_compare
lecture: Protein Prediction 1 - Protein structure for Computational Biologist - TUM Summer 2015
Videos: YouTube / www.rostlab.org
THANKS:
Tim Karl + Carlo Di Domenico
carlo.de-domenico@tum.de

Special lectures:
• TBA

No lecture:
• 05/12 Student assembly (SVV)
• 05/14 Ascension day
• 05/26 Whitsun holiday
• 06/04 Corpus Christi

LAST lecture: July 7
Examen: July 9

CONTACT: Inga Weise assistant@rostlab.org
TOC today

☐ LAST lectures: Protein introduction
  • amino acids
  • protein structure
  • bonds & energies

☐ THIS lecture (thursday)
  • structure comparisons

☐ NEXT WEEK (tuesday/thursday=no lecture)
  • alignments
Protein Prediction - Part 1: Structure

1 Introduction (contd.)
More about domains
Guessing domains from sequence

protein A
protein B
protein C
protein D
protein E
protein F

domain 1 domain 2
Domains
Most proteins multi-domain
Most proteins multi-domain

Single-domain proteins:
61% in PDB
28% in 62 proteomes
Most domains have ~100 residues

Liu, Hegyi, Acton, Montelione & Rost 2003 Proteins 56:188-200
Liu & Rost 2004 Proteins 55:678-686
Most domains have ~100 residues

Average domain length
• in proteins ≥ 2 domains: 
  ~100 residues
• in proteins with 1 domain: 
  1.7-3 times longer

Liu, Hegyi, Acton, Montelione & Rost 2003 Proteins 56:188-200
Liu & Rost 2004 Proteins 55:678-686
Kingdoms similar in length

Kingdoms similar in amino acids usage

short insert: CHOP - domain guessing
How to identify domains?
Proteins have domains
How to identify domains from sequence?
CHOP proteins into structural domains

CUT 1: 3D domains
CHOP proteins into structural domains

CUT 1:
3D domains

CUT 2:
Pfam regions
CHOP proteins into structural domains

CUT 1: 3D domains

CUT 2: Pfam regions

CUT 3: SWISS-PROT

J Liu & Rost 2004 Proteins 55:678-686
CHOP proteins into structural domains

CUT 1: 3D domains

CUT 2: Pfam regions

CUT 3: SWISS-PROT

Final

5 domains “today”
CHOP proteins into structural domains

CUT 1: 3D domains

CUT 2: Pfam regions

CUT 3: SWISS-PROT

Final 5 domains “today”

J Liu & Rost 2004 Proteins 55:678-686
Notation: protein structure 1D, 2D, 3D

Notation: protein structure 3D, 2D, 1D
Notation: protein structure 1D, 2D, 3D
Notation: protein structure 1D, 2D, 3D

<table>
<thead>
<tr>
<th>Notation</th>
<th>1D</th>
<th>2D</th>
<th>3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>P PP P</td>
<td>128</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>Q QQQY</td>
<td>175</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>I FPQVI</td>
<td>70</td>
<td>E 60</td>
<td></td>
</tr>
<tr>
<td>T SSIVR</td>
<td>77</td>
<td>E 69</td>
<td></td>
</tr>
<tr>
<td>L LLSTL</td>
<td>120</td>
<td>E 14</td>
<td></td>
</tr>
<tr>
<td>W WWQED</td>
<td>238</td>
<td>E 81</td>
<td></td>
</tr>
<tr>
<td>Q RKQAK</td>
<td>169</td>
<td>E 97</td>
<td></td>
</tr>
<tr>
<td>R RRRPQ</td>
<td>200</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>P PPPPP</td>
<td>24</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>L VVTKF</td>
<td>E 71</td>
<td>E 59</td>
<td></td>
</tr>
<tr>
<td>V VLLII</td>
<td>E 14</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>T TTKEK</td>
<td>74</td>
<td>E 69</td>
<td></td>
</tr>
<tr>
<td>I AALIV</td>
<td>E 0</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>K HYKKF</td>
<td>E 90</td>
<td>E 73</td>
<td></td>
</tr>
<tr>
<td>I IILVI</td>
<td>4</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>G EENGG</td>
<td>46</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>G GGTTG</td>
<td>62</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Q QQKRR</td>
<td>68</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>L PPLNW</td>
<td>E 118</td>
<td>E 59</td>
<td></td>
</tr>
<tr>
<td>K VVFKV</td>
<td>E 31</td>
<td>E 73</td>
<td></td>
</tr>
<tr>
<td>E EESK</td>
<td>E 124</td>
<td>E 95</td>
<td></td>
</tr>
<tr>
<td>A VVGLG</td>
<td>E 1</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>L LLILL</td>
<td>E 29</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>L LLLVV</td>
<td>E 24</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>D DDDDD</td>
<td>49</td>
<td>E 58</td>
<td></td>
</tr>
<tr>
<td>T TTTTT</td>
<td>72</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>G GGGGG</td>
<td>62</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>A AAAAA</td>
<td>17</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>D DDDDD</td>
<td>102</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>D DDAKE</td>
<td>69</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>T SSSTTV</td>
<td>E 1</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>V IIVIV</td>
<td>E 14</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>L VVIVL</td>
<td>E 0</td>
<td>E 0</td>
<td></td>
</tr>
</tbody>
</table>
3D classifications
Blue and red similar?

Doyle et al. (1998) Science 280:69-77 - The structure of the potassium channel: molecular basis of K+ conduction and selectivity
Similarity now clearer?

Doyle et al. (1998) Science 280:69-77 - The structure of the potassium channel: molecular basis of K+ conduction and selectivity
3D comparisons: how to objects?
Matching shapes

How to match?
How to match?
Differences for corresponding points
Differences for corresponding points
Differences for corresponding points
Differences for corresponding points
Differences for corresponding points
Differences for corresponding points

Difference

\[ = d_1 + d_2 + d_3 \ldots + d_8 \]

\[ = |r_{1a} - r_{1b}| + \ldots + |r_{8a} - r_{8b}| \]

RMSD (root mean square deviation)

\[ = \sqrt{\sum_i (r_i^A - r_i^B)^2} \]
Differences for corresponding points

\[ RMSD(A, B) = \sqrt{\sum_i (r_i^A - r_i^B)^2} \]
Actual algorithm inversed

1st: find corresponding points
2nd: superimpose

\[
RMSD(A, B) = \sqrt{\sum_i (r_i^A - r_i^B)^2}
\]
fit now?
Scaling easy for simple shapes

\[ x^2 + y^2 = r^2 \]
Proteins: points are defined -> no scaling

Global vs. local comparisons
Global vs. local comparisons
Global vs. local comparisons

Global solution 1:

Global solution 2:
cut into “units”
cut into “units”
trouble: where to stop?

valid “unit” for comparison?
How to decide what is a valid unit?
Decision upon validity

valid “unit” for comparison?
Valid or not?

Scientifically significant: some expert says
Valid or not?

- **Scientifically significant:** some expert says

- **Statistically significant:** background

![Graph showing distribution of scores with 'background' and 'signal' areas.]
Cut, match, compare by RMSD

\[ RMSD(A, B) = \sqrt{\sum_{i} (r_i^A - r_i^B)^2} \]
Only Cartesian RMSD comparison?

\[ RMSD(A, B) = \sqrt{\sum_i (r_i^A - r_i^B)^2} \]
2D: difference matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparison 2D: differences of differences

Total of 8 x 8 differences
3D comparisons: how to proteins?
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


© Patrice Koehl, UC Davis
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)
RMSD is not a metric

\[ \text{cRMSD} = 2.8 \, \text{Å} = 0.28 \, \text{nm} \]

\[ \text{cRMSD} = 2.85 \, \text{Å} = 0.285 \, \text{nm} \]

A similar B
B similar C
NOT implying:
A similar C
SSAP
3D alignment
Taylor & Orengo
Structural alignment: SSAP

William R. Taylor & Christine A. Orengo

Willy Taylor @CASP7

Christine Orengo @ISCB-Africa

SSAP: Sequential Structure Alignment Program


© Erik Bongcam-Rudloff
Structural alignment: **SSAP**

**William R. Taylor & Christine A. Orengo**

Willy Taylor @CASP7

Christine Orengo @ISCB-Africa

©Erik Bongcam-Rudloff

**SSAP: Sequential Structure Alignment Program**

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

**IDEA: use C-beta distance matrix and apply double dynamic programming**
SSAP concept

WR Taylor & CA Orengo (1989)
Protein structure alignment
JMB 208:1-22
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

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

Replace distances by interatomic vectors (V)

Optimize:

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

WR Taylor & CA Orengo (1989)
Protein structure alignment
JMB 208:1-22
Include sequence information (D(xy): Dayhoff)

Optimize:

\[ S_{ik} = \frac{wD_{RiRk} + a}{| V_{ij}^A - V_{kl}^B | + b} \]

WR Taylor & CA Orengo (1989)
Protein structure alignment
JMB 208:1-22
DALI

3D alignment

Holm & Sander
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
Chris Sander

- Sloan Kettering Cancer Center, NYC
- Papers:
  - >770 papers (May 2011)
  - 1 >6,000 citations (May 2011)
  - 6 >1,000 citations (May 2011)
  - 87 over 100
  - H-index 92 (ISI May 2011)
- ISCB Fellow
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


- Distance matrix Alignment
- Algorithm: Monte Carlo on all-against-all for hexapeptides (5)
Vorolign
3D alignment
Birzele & Zimmer
Structural alignment: VOROLIGN

Fabian Birzele, Ralf Zimmer et al.


© Burkhard Rost (TUM Munich)
Structural alignment: VOROLIGN


Dynamic programming on Voronoi environments

F Birzele, JE Gewehr, G Csaba & R Zimmer (2006) Bioinformatics 23:e205-11: Fig. 2
3D comparisons: others
2 forms of calcium-bound Calmodulin

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

### Structure alignment methods

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

Rachel Kolodny, Patrice Koehl, Michael Levitt

Rachel Kolodny
Univ of Haifa

Patrice Koehl
UC Davis

Michael Levitt
Stanford Univ

Comprehensive Evaluation of Protein Structure Alignment Methods: Scoring by Geometric Measures
JMB 346:1173-88
How to assess 3D comparisons?
standard-of-truth?
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
Comparison of structure alignments

Best-of-All

JMB 346:1173-88
(Fig. 1A)

dashed lines: original method
solid lines: SAS measure
3D comparisons: protein space and databases
Structural universe

B Rost 1998 *Structure* 6:259-263
Evolution of pieces

© Andrei Lupas MPI Tuebingen
Structure evolves without leaps?

Fig. 1: NV Grishin 2001 NAR 29:638-43

© Nick V Grishin HHMI + Univ Dallas
Structural universe: no islands, really
3D classifications: goals

- Similar 3D -> 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)
Protein structure comparisons

All-alpha

All-beta

AlphaBeta

3sdh

1bww

1xne
How to recognize the similarity?
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 - QSCOP - TopMatch
  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 classes

- alpha
- beta
- alpha and beta (a/b – interspersed)
- alpha plus beta (a+b – segregated)
- multidomain proteins
- membrane and cell-surface proteins
- small proteins
- coiled coil proteins
- low-resolution protein structures
- peptides
- designed proteins
SCOP class

CLASS = alpha and beta (a/b)

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

1sw0-TIM beta/alpha barrel
Example

{All-alpha}

a.

{Globin-like}

a.1

| class

| fold

Structure similarity increases
SCOP fold definition

- same major secondary structures
  - in the same arrangement
  - with the same topological connections

- peripheral elements may differ
  - up to 50% peripheral
  - Turns and secondary structure elements

- evolutionary relationship unclear
SCOP fold

CLASS = alpha and beta (a/b)
FOLD   = TIM beta/alpha-barrel
Structural universe: no islands, really

B Rost 1998 *Structure* 6:259-263
SCOP hierarchy

Example

{All-alpha}

{Globin-like}

{alpha-helical ferrodoxin}

Structure similarity increases

\[ a. \quad \text{class} \]

\[ a.1 \quad \text{fold} \]

\[ a.1.2 \quad \text{superfamily} \]
probable common evolutionary origin

low similarities, but

• share the same fold
• have similar functions
SCOP hierarchy

TRIOSEPHOSPHATE ISOMERASE (1swo)

PHOSPHATE ALDOLASE (1p1x)

QUINOLINIC ACID PHOSPHORIBOSYLTRANSFERASE (1qap)
SCOP hierarchy

Example

\{\text{Alpha and beta a/b}\}
\{\text{TIM beta/alpha-barrel}\}
\{\text{Triosephosphate isomerase}\}

\text{c.1.1.1 \textit{superfamily}}

\text{c.1.1 \textit{fold}}

\text{c \textit{class}}

\text{Structure similarity increases}

\text{c.1.1.1.1 \textit{family}}

(sequence based)
SCOP family definition

- clearly evolutionary relation
- Sequence identity often >30%, but not necessarily, e.g. globins: < 15% sequence identity for some members
3D classification databases

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

Christine Orengo
@ISCB-Africa
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)
Universe of protein structures

Christine Orengo (Structures, 1997, 5, 1093-1108)

Christine Orengo et al. 1997 Structures 5 1093-1108
CATH

- Class
- Architecture
- Topology
- Homology
Class:
mostly alpha, mostly beta, mixed alpha/beta, few regular secondary structure

All-alpha

All-beta

AlphaBeta

3sdh

1bww

1xn
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
<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>folds</td>
<td>1,282</td>
</tr>
<tr>
<td>superfamilies</td>
<td>2,549</td>
</tr>
<tr>
<td>sequence families</td>
<td>11,330</td>
</tr>
<tr>
<td>domains</td>
<td>24,232</td>
</tr>
</tbody>
</table>
CATH: steps involved

1. Find domain

Multiple 3D alignment identifies consensus secondary structure

© Christine Orengo

© Burkhard Rost (TUM Munich)
CATH: steps involved

Find domains

- **ab initio**: consensus of three methods:
  - DETECTIVE:
    - hydrophobic interior
  - PUU:
    - likely separation motion
  - DOMAK:
    - count internal and external contacts
- Problem: only 20% consistent!
- **Based on prior knowledge**: CATHEDRAL
  - GT: secondary structure matching
  - DDP: structural alignment

Redfern, O.C. et al. (2007)
CATH: steps involved

- Find domain
- From domain to superfamily

PDB id: 1gcq
(SH3 domains)

PDB id: 1gcqA0
(SH3 domain)

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

PDB id: 1gcqA0
(SH3 domain)

http://www.cathdb.info/domain/1gcqA0

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

CATH Architecture Roll
(1nh2D02)

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

© CATH tutorial (www.cathdb.info)
CATH vs SCOP

At 80% residue domain overlap:
70% of proteins in PDB have similar domains

Table 4: Detailed mappings of domain pairs in percent from SCOP onto CATH

<table>
<thead>
<tr>
<th></th>
<th>outer</th>
<th>class</th>
<th>fold</th>
<th>superfamily</th>
<th>family</th>
</tr>
</thead>
<tbody>
<tr>
<td>outer</td>
<td>79.38%</td>
<td>8.31%</td>
<td>0.99%</td>
<td>0.40%</td>
<td>0.03%</td>
</tr>
<tr>
<td>class</td>
<td>18.16%</td>
<td>56.15%</td>
<td>2.55%</td>
<td>1.88%</td>
<td>0.87%</td>
</tr>
<tr>
<td>arch</td>
<td>2.42%</td>
<td>24.90%</td>
<td>2.80%</td>
<td>1.27%</td>
<td>0.09%</td>
</tr>
<tr>
<td>top</td>
<td>0.04%</td>
<td>10.50%</td>
<td>81.99%</td>
<td>4.44%</td>
<td>0.66%</td>
</tr>
<tr>
<td>hom</td>
<td>0.002%</td>
<td>0.14%</td>
<td>11.66%</td>
<td>92.01%</td>
<td>98.34%</td>
</tr>
</tbody>
</table>

CATH: 50 structures - 1 superfamily

superfamily 3.40.640.10

Type I PLP-dependent aspartate aminotransferase-like (Major domain)

© CATH - Christine Orengo
Remaining families (new BIG)

BIG families (currently Pfam)

Structural families
(i.e. one or more solved structures CATH/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/
Classify protein structure: COPS/QSCOP/TopMatch
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")
COPS hierarchy

- COPS = Classification Of Protein Structures
- Based on quantified structural comparison
- 2007: additional info for SCOP domains: qSCOP
- 2009: workbench based on PDB chains: TopSearch
  http://topsearch.services.came.sbg.ac.at/

COPS metric

Axioms / Definitions:

\[ S_{a,a} = L_a \]
\[ S_{a,b} \geq 0 \]
\[ S_{a,b} = S_{b,a} \]
\[ S_{b,c} \geq S_{a,b} + S_{a,c} - L_a \]
\[ D_{a,b} = L_a + L_b - 2S_{a,b} \]

- Alignment method not so important!
  for COPS: TopMatch
- Metric can reveal alignment problems
  (e.g. via triangle inequality)
PDB updates 2008/08/19-2009/04/14

novelty:

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) - c2a5yB2

c1z6tA3 (helical domain) - c2a5yB3

c1z6tA4 (winged-helix domain) - c2a5yB4

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

PDB id
1z6t-A
with 2a5y-B

SJ Suhrer et al. (2009) NAR 37:W539-W544
COPS <-> TopSearch

- No domain decomposition
- But:
  - Complete structure comparisons
  - Biological units
  - New metric

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