title: Comparative modeling

short title: 3d_1

lecture: Protein Prediction I - Protein Structure / Burkhard Rost, TUM, 2011 summer
Announcements

Videos: SciVe / www.rostlab.org

THANKS:
Tim Karl + Haitham Sohby

Special lectures:
• Jun 28: Mark Offman: MD
• Jun 30: Andrea Schafferhans: Docking

Munich Bioinformatics Day
Jun 24 9:00-17:00

NO lectures:

LAST lecture: Jul 7

Examen: Jul 12, 10:30 (likely this room)
• Makeup: likely: October 13 - morning
Today: Secondary structure prediction 1

☐ LAST WEEKs
  • Secondary structure prediction

☐ THIS WEEK
  • Comparative modeling

☐ NEXT WEEK
  • Comparative modeling/3D prediction

☐ WEEK Jun 28/30
  • Mark Offman: Molecular Dynamics
  • Andrea Schafferhans: Docking

☐ WEEK Jul 5/7
  • 2D prediction / 1D accessibility / membrane
3D prediction: comparative modeling / homology modeling
Use structure to improve links

Cluster 1

PDB-profiles

Cluster 2

common profile
common structure
common function

PSI / BLAST

PDB
Using structure to predict function


Archaeal structure

GRASS:

human protein
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>110</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>70</td>
<td>E 60</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>77</td>
<td>E 69</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>120</td>
<td>E 14</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>238</td>
<td>E 81</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>169</td>
<td>E 97</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>200</td>
<td>E 62</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>24</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>71</td>
<td>E 59</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>14</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>74</td>
<td>E 69</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>90</td>
<td>E 73</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>4</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>46</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>62</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>68</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>118</td>
<td>E 59</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>31</td>
<td>E 73</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>124</td>
<td>E 95</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>29</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>24</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>49</td>
<td>E 58</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>72</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>62</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>17</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>102</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>69</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>1</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>14</td>
<td>E 0</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>0</td>
<td>E 0</td>
<td></td>
</tr>
</tbody>
</table>

kcal/mol
Universe of protein structures

Christine Orengo et al. 1997 Structures 5 1093-1108

© Burkhard Rost (TU Munich)
Goal of structure prediction

- Epstein & Anfinsen, 1961: sequence uniquely determines structure

**INPUT:** sequence

**OUTPUT:** 3D structure and function
Structure-based computational biology
Secondary structure
Protein structure prediction in reality

Experimental

Estimate for 1999
Zones
Idea of comparative modeling

WHAT IF

1shf 100% VTLFVALYDYARTEDDLSFKGEKIQILNSSEGDSWEARSLTTGETGYIPSNYVAPVD
1srn  78% VTTFVALYDYRESRTETDLSFKGERLQIVNTEGDWLHSTTTGQTGYIPSNYVAPS
1sem  39% ....VAEHDFQAGSPDELFSFKNLNLKDDPHWYKAE.DGNEGFIPSNYIRMTE
Comparative modeling

- assumption: H and U homologous 3D structures
- strategy: modelling of U based on H

U (sequence) → PDB

significant sequence identity

H
Protein structure prediction in reality

Experimental

HoMo

Estimate for 1999
Twilight zone = false positives explode!!

Percentage sequence identity

Number of protein pairs

Distance from HSSP threshold

B Rost 1999 Prot Engin 12, 85-94

© Burkhard Rost (TU Munich)
Protein structure prediction in reality

Experimental

HoMo

Estimate for 1999
Evolution into the Midnight zone

Number of structure pairs

Percentage pairwise sequence identity

B Rost 1997 *Folding & Design* 2, S19-S24

© Burkhard Rost (TU Munich)
Protein structure prediction in reality

Estimate for 1999

Experimental

HoMo
Protein structure prediction in reality

Experiment

HoMo

FoRc

Estimate for 1999
Protein structure prediction in reality

Estimate for 1999

- Experiment
- HoMo
- FoRc
- 1D
Bridging potential vs. sequence identity
Redundancy in the PDB

Number of proteins in PDB

redundancy reduced by PIDE

© Burkhard Rost (TU Munich)
Redundancy in the PDB

Percentage of proteins in PDB

redundancy reduced by PIDE
Redundancy in the PDB

Percentage of proteins in PDB

Redundancy reduced by PIDE

Percentage of proteins in SWISS-PROT

Percentage of pairwise sequence identity
Protein structure prediction in reality

- 1D
- 3D
- HoMo
- FoRc
Protein structure prediction in reality

SWISS-PROT view

Genome view

1D
HoMo
FoRc
3D

....the art of being humble

© Burkhard Rost (TU Munich)
Improving prediction by waiting it out …

1991
Improving prediction by waiting it out ...
Jinfeng Liu

- 1995-2003 MS Rutgers Univ.
- 1998-2004 PhD Columbia Univ. Pharmacology

PhD with 16 publications!

- 2007-now Genentech, CA
Homology modeling for entire genomes

Number of ORFs
Number of ORFs with PDB hit

Number of proteins

Organism

- H sapiens
- H sapiens(chr. 22)
- D melanogaster
- C elegans
- S cerevisiae
- U urealyticum
- T pallidum
- S PCC6803
- R prowazekii
- N meningitidis
- M tuberculosis
- M pneumoniae
- M genitalium
- H pylori
- H influenzae
- E coli
- C trachomatis
- C pneumoniae
- C jejuni
- B burgdorferi
- B subtilis
- A aeolicus
- T maritima
- D radiodurans
- P horikoshii
- P abyssi
- M thermoautotrophicus
- M jannaschii
- A fulgidus
- A pernix
Homology for protein universe

Organism:

- H sapiens(chr. 22)
- H sapiens
- D melanogaster
- C elegans
- S cerevisiae
- U urealyticum
- T pallidum
- S PCC6803
- R prowazekii
- N meningitidis
- M tuberculosis
- M pneumoniae
- M genitalium
- H pylori
- H influenzae
- E coli
- C trachomatis
- C pneumoniae
- C jejuni
- B burgdorferi
- B subtilis
- A aeolicus
- T maritima
- D radiodurans
- P horikoshii
- P abyssi
- M thermoautotrophicum
- M jannaschii
- A fulgidus
- A pernix

Percentage of all ORFs in genome
Comparative modeling: Concept
Comparative modeling: words

- **Comparative** modeling
  
  vs.
  
  **Homology** modeling
Idea of comparative modeling

WHAT IF

1shf 100% VTLFVALYDYEARTEDDLSFKGEKLFQILNSSEGDSWEARSLTTGETGYIPSNYVAPVD
1srn 78% VTTFVALYDYESRTETDLSFKGRGLQIVNTEGDWLALHSLTTGQTYIPSNYVAPSD
1sem 39% ....VAEHDFQAGSPDELSFKRGNKLKDNKDEPDHYKAEL.DGNEGFIIPSNYIRMTE
Comparative modeling: words

- **Comparative** modeling vs. **Homology** modeling

- **Lingo:**
  - Target: protein to model
  - Template: protein to model from
Comparative modeling

- assumption: H and U homologous 3D structures
- strategy: modelling of U based on H
Comparative modeling: steps

- Identify template(s) through database search
- Align target/template
- Build model
- Assess model
- (refine)
Comparative modeling: steps

- Identify template(s) through database search
  - PSI-BLAST

- Align target/template

- Build model

- Assess model

- (refine)
Extending comparative modelling: threading

### Percentage of pairwise identical residues

- **100%**
  - Region of homology modelling (sequence alignment suffice)

- **25%**
  - Fold recognition

- **0%**

### Accuracy of automatic fold recognition

- **correct first hit:**
  - $\approx 20\text{-}30\%$

- **alignment correct to some extent:**
  - $\approx 10\text{-}25\%$

- **remote homology modelling (3D) correct:**
  - $< 10\%$
Comparative modeling: steps

- Identify template(s) through database search
  - PSI-BLAST
- Align target/template
- Build model
- Assess model
- (refine)
Comparative modeling: steps

☐ Identify template(s) through database search
  - PSI-BLAST

☐ Align target/template
  - dynamic programming / threading-like / HMM / profile-profile

☐ Build model

☐ Assess model

☐ (refine)
Comparative modelling: 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: State-of-the-art methods
Modeller
Andrei Sali
et al.
Andrej Sali - UCSF

CV
- PhD Birkbeck - Tom Blundell
- PD Harvard - Martin Karplus
- Rockefeller Univ.
- UCSF

Publications
(2011/06)
- > 400 publications
- 1x ~4,000
- 45x >100
- H-index > 64

Andrej Sali, UCSF
Comparative modeling: MODELLER

MODELLER:


Andrej Sali, UCSF
MODELLER:


Marc Marti-Renom, CIPF Barcelona
(here at ISCB-Africa in Bamako, Mali)
Comparative modeling: MODELLER

www.salilab.org/modeller/

MODELLER: constraint satisfaction

1. Align sequence with structures
   Template structure(s)
   Target sequence

2. Extract spatial restraints

3. Satisfy spatial restraints

Figure 1.1: First, the known, template 3D structures are aligned with the target
sequence to be modeled. Second, spatial features, such as $C_{\alpha}-C_{\alpha}$ distances,
hydrogen bonds, and mainchain and sidechain dihedral angles, are transferred
from the templates to the target. Thus, a number of spatial restraints on its structure
are obtained. Third, the 3D model is obtained by satisfying all the restraints as well
as possible.

Source: Modeller manual

MODELLER: constraint satisfaction: fits

Source: Modeller manual

MODELLER: overview

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

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

- Molecular dynamics (MD)
- Langevin dynamics (LD)
- Self-guided MD & LD
- Rigid bodies
- Rigid molecular dynamics
- Rigid minimization

Source: Modeller manual

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.
Typical errors

- side chain packing
- alignment shift
- no template (loop wrong)

Chapter 5 - Unit 5.6.1-30 (Fig. 5.6.12)
Comparative modeling methods

- MODELLER
  - lots of whistles and bells
  - downloadable
  - very accurate
SWISS-MODEL
Thorsten Schwede et al.
CV
- PhD in X-ray crystallography
- SIB
- Biocenter Basel

Publications (2011/06)
- > 45 publications
- 1x ~2,000
- 5x >100
- H-index > 19

Web presence
- 1000s of accesses / day
Comparative modeling: SWISS-MODEL

-Manuel Peitsch, Philip Morris Internatl. Basel

-SWISS-MODEL:

-Nicolas Guex, Vital-IT

© Burkhard Rost (TU Munich)
Comparative modeling: SWISS-MODEL

SWISS-MODEL:


Manuel Peitsch
Philip Morris Internatl.

Nicolas Guex
Vital-IT
photo: http://www.vital-it.ch/vitalit_images/Guex.jpg

Torsten Schwede,
BioZentrum Basel
photo: http://www.unibas.ch/mediaDB/schwede_torsten_big.jpg

© Burkhard Rost (TU Munich)
Comparative modeling: SWISS-MODEL

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists Worldwide.

What’s new?
- SWISS-MODEL is running on new hardware with better performance
- Find more news on SWISS-MODEL Blogg

References:
When you publish or report results using SWISS-MODEL, please cite the relevant publications:


© Burkhard Rost (TU Munich)
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

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
Christopher Andrew Wilton

Vuolukiventie 1b M247, 00710 Helsinki, Finland
Work: +358 919159114  Mobile: +358 408331316
E-mail: chris.wilton@helsinki.fi  DoB: 07 Dec 1973  Nationality: British
http://ekhidna.biocenter.helsinki.fi/groupmembers/chris

"I aim to provide quality services and applications that are clear, useful, and user-friendly. The continuing enrichment of my knowledge and programming skills gained over the past four years, together with my high attention to detail, strong problem-solving skills and fast learning abilities, all help me to achieve that aim."
SWISS-MODEL: template search

© Chris Wilton
(Helsinki)
SWISS-MODEL: template search

© Chris Wilton (Helsinki)
SWISS-MODEL: result after many steps

© Chris Wilton (Helsinki)
SWISS-MODEL: model assessment
SWISS-MODEL:

- Long stretches above zero probably loops
- Most-negative regions well buried in core
- Functional residues above zero!
- Majority of residues must be below zero
- Compare your model to pdb template!
Comparative modeling methods

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

- SWISS-MODEL
  automated, increasingly comprehensive and flexible
Other top comparative modeling methods: HHpred, 3D-Jigsaw, WhatIf, COMA
Comparative modeling methods

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

- SWISS-MODEL
  automated, increasingly comprehensive and flexible

- HHpred/HHsearch
Comparative modeling: HHpred

HHpred:
- J Soeding, A Biegert & AN Lupas (2005) NAR 33:W244-8

Johannes Soeding, LMU
photo: http://www.lmb.uni-muenchen.de/soeding/images/soeding.jpg

Andrei Lupas, MPI Tuebingen
photo: http://www.mph.tuebingen.mpg.de/pix/perspics/pic444.jpg
Comparative modeling: HHpred

HHpred:

- J Soeding, A Biegert & AN Lupas (2005) NAR 33:W244-8

Major step:
HMM-HMM
(profile-profile) alignment
+ many details right
Comparative modeling methods

- **MODELLER**
  lots of whistles and bells, downloadable, very accurate

- **SWISS-MODEL**
  automated, increasingly comprehensive and flexible

- **HHpred/HHsearch**
  very accurate, automated

- **3D-JIGSAW**
Comparative modeling: 3D-JIGSAW

3D-JIGSAW:

Paul A Bates,
London Res Inst
photo: http://www.bmm.icnet.uk/~bates03/

Michael JE Sternberg
photo: http://www3.imperial.ac.uk/pls/portallive/docs/1/63011698.JPG
Comparative modeling: 3D-JIGSAW

Comparative modeling: 3D-JIGSAW

- fragment-based approach:
  1. identify all similar fragments in PDB
  2. assemble fragments to structure

Comparative modeling methods

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

- SWISS-MODEL
  automated, increasingly comprehensive and flexible

- HHpred/HHsearch
  very accurate, automated

- 3D-JIGSAW
  automated, accurate

- WHAT IF
Comparative modeling: WHAT IF

WHAT IF:
• G Vriend (1990) J Mol Graph 8:52-6

Gert Vriend
CMBI Nijmegen

photo: http://swift.cmbi.ru.nl/gv/start/IMAGE/VRIEND.jpg
Comparative modeling: WHAT IF & YASARA

YASARA

• E Krieger, JE Nielsen, C Spronk & G Vriend (1990)
  J Mol Graph 8:52-6

Elmar Krieger
YASARA
Biosciences, Vienna

photo: http://www.yasara.org/
  ekrieger_small.jpg
Comparative modeling methods

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

- SWISS-MODEL
  automated, increasingly comprehensive and flexible

- HHpred/HHsearch
  very accurate, automated

- 3D-JIGSAW
  automated, accurate

- WHAT IF
  expert users, does anything incl. chess
Comparative modeling methods

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

- SWISS-MODEL
  automated, increasingly comprehensive and flexible

- HHpred/HHsearch
  very accurate, automated

- 3D-JIGSAW
  automated, accurate

- WHAT IF
  expert users, does anything incl. chess

- COMA
Comparative modeling: COMA

COMA:

• M Margelevicius, M Laganeckas & Ceslovas Venclovas (2010)
  Bioinformatics in press

Ceslovas Venclovas
Inst Biotechnology, Vilnius, Lithuania
Comparative modeling: COMA

- COMA: mostly through good profile-profile alignments
- special features:
  - position specific gap penalty
  - global score

Comparative modeling: COMA


M Margelevicius & Ceslovas Venclovas (2010) BMC Bioinformatics 11:89, 9-14; Fig. 3

global: entire 3D domain

local
Comparative modeling methods

- **MODELLER**
  lots of whistles and bells, downloadable, very accurate

- **SWISS-MODEL**
  automated, increasingly comprehensive and flexible

- **HHpred/HHsearch**
  very accurate, automated

- **3D-JIGSAW**
  automated, accurate

- **WHAT IF**
  expert users, does anything incl. chess

- **COMA**
  reaches deep into twilight zone, automated
Assessing protein structure prediction
Protein structure prediction problem solved?

- Problem: predict the 3D structure of a protein from sequence alone
How would you assess prediction performance?
Protein structure prediction problem solved!

- Problem: predict the 3D structure of a protein from sequence alone
- 60s - Washington Post (source ?)
- 70s - New York Times (source ?)
- 90s - Washington Post (source ?)
CASP
(Critical Assessment of Structure Prediction)
CASP

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

John Moult, CARB
CASP organizers: the beginning

Shapers and Shakers

John Moult, CARB

Tim Hubbard, Sanger
CASP organizers: the beginning

John Moult, CARB
Krzysztof Fidelis, UC Davis
Tim Hubbard, Sanger
Jan T Pedersen

Richard Judson, EPA National Center for Computational Toxicology, USA
Summary: prediction in 3D/2D
Summary: prediction in 3D/2D

No successful prediction of 3D from sequence!!
No successful prediction of 3D from sequence!!

No prediction of 3D from sequence!!
Summary: prediction in 3D/2D

- No successful prediction of 3D from sequence!!
- No prediction of 3D from sequence!!
- Accurate 3D prediction for coiled-coil proteins, but ..
Summary: prediction in 3D/2D

- No successful prediction of 3D from sequence!!
- No prediction of 3D from sequence!!
- Accurate 3D prediction for coiled-coil proteins, but ..
- Recognizing incorrect 3D structures successful
Summary: prediction in 3D/2D

- No successful prediction of 3D from sequence!!

- No prediction of 3D from sequence!!

- Accurate 3D prediction for coiled-coil proteins, but ..

- Recognizing incorrect 3D structures successful

- Extracting principles about structure formation from structures?
Structure prediction servers impacted biology
Protein Structure Prediction

CASP

© Burkhard Rost (Columbia New York)
© Burkhard Rost (TU Munich)
Protein Structure Prediction

- Only homology modelling good
- No general prediction of 3D from sequence, yet

CASP
Protein Structure Prediction

- Only homology modelling good
- No general prediction of 3D from sequence, yet
- Important improvements in many fields!

CASP
Servers, META-servers, META-META, …
The real thing: humans with machines
Problems with CASP
Problems of CASP

- Comparisons based on apples and oranges
- Analysis of irrelevant types of test cases
- Inappropriate ranking
- Conclusions based on insignificant differences
- Different categories evaluated differently
- Too few targets
Ranking not stable!

29 different worse than 11 identical

# Pairwise comparison matrix

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>□</td>
<td>□</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>□</td>
<td></td>
<td></td>
<td>□</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>□</td>
<td></td>
<td>□</td>
</tr>
<tr>
<td>D</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td></td>
</tr>
</tbody>
</table>

© Burkhard Rost (Columbia New York)
© Burkhard Rost (TU Munich)
CAFASP

- Critical Assessment of Fully Automated Structure Prediction
- CASP for servers
  - fully automated, but still:
  - small numbers
  - non-continuous

Daniel Fischer
Univ Buffalo
EVA (Aoifa)
EVA: automatic continuous EVAluation of structure prediction

Satellites/Mirrors

CUBIC Columbia
Rockefeller
CNB Madrid

secondary structure, fold recognition
comparative modelling, fold recognition
inter-residue contacts / distances

Get PDB
Analyze: pairwise BLAST
EVA-DB
Send sequences

PDB
Prediction servers

META-PP
Collect Send

Compile results at
EVA - SATELLITES

summary
week
one protein
PDB vs prediction

EVA - Mirrors

Results
static pages

User
• browse
• query
• ftp

every day

Get PDB
Analyze: pairwise BLAST
EVA-DB
Send sequences

еvery week

Collect HTML
Update central pages

© Burkhard Rost (TU Munich)
EVA co-developers

Andrej Sali, UCSF

Volker Eyrich, Schroedinger, New York

Marc Marti-Renom, CIPF Barcelona (here at ISCB-Africa in Bamako, Mali)
EVA: automatic continuous EVAluation of structure prediction

Satellites/Mirrors

CUBIC Columbia
Rockefeller
CNB Madrid

secondary structure, fold recognition
comparative modelling, fold recognition
inter-residue contacts / distances

Get PDB
Analyse: pairwise BLAST
VVA-DB
Send sequences
Analyse:
• PSI-BLAST
• MaxHom
• sequence-unique sets
Collect HTML
Update central pages

PDB
Prediction servers
evendy
every week

Compile results at
EVA - SATELLITES
summary
week
one protein
PDB vs prediction

EVA - Mirrors

User
• browse
• query
• ftp

Results
static
pages
EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)

http://eva.compbio.ucsf.edu/~eva/cm/

EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/


© Burkhard Rost (TU Munich)
EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)

http://eva.compbio.ucsf.edu/~eva/cm/

Cumulative distribution

PSI-BLAST $10^{-3}$

EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

Accuracy
Cumulative distribution

PSI-BLAST $10^{-3}$

EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

Accuracy

Cumulative distribution

PSI-BLAST $10^{-3}$

EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

Accuracy

Cumulative distribution

PSI-BLAST $10^{-3}$

Coverage


© Burkhard Rost (TU Munich)
EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

Accuracy

Cumulative distribution

PSI-BLAST $10^{-3}$

Coverage


© Burkhard Rost (TU Munich)
N Eswar et al. & A Sali (2008) Methods Mol Biol 426: 145-59 (Fig. 4)
LiveBench
LiveBench

- automated test of servers
- differences to EVA:
  - limited time “experiments”
  - top ranks top

Daniel Fischer
Univ Buffalo

Leszek Rychlewski
BioInfoBank, Poznan

Arne Elofsson
Stockholm Univ
LiveBench + EVA?

Laszlo Kajan

Laszlo Kajan TU Munich
3D prediction: comparative modeling in twilight→midnight zone
Extending comparative modelling: threading

- correct first hit: ≈ 20-30%
- alignment correct to some extent: ≈ 10-25%
- remote homology modelling (3D) correct: < 10%

Percentage of pairwise identical residues:
- 100%
- 25%
- 0%

Accuracy of automatic fold recognition:
- Region of homology modelling (sequence alignment suffice)

© Burkhard Rost (TU Munich)
Comparative modelling: 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

© Burkhard Rost (TU Munich)
EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

Accuracy

Cumulative distribution

PSI-BLAST $10^{-3}$

EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

Accuracy  Cumulative distribution  Coverage

PSI-BLAST \(10^{-3}\)


© Burkhard Rost (TU Munich)
EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

Accuracy

Cumulative distribution

PSI-BLAST 10^{-3}

Coverage


Wednesday June 15, 2011
Sequence conservation of protein structure

B Rost 1999 Prot Engin 12, 85-94
Sequence conservation of protein structure

B Rost 1999 Prot Engin 12, 85-94
Midnight zone STRONGLY populated

B Rost 1997 *Folding & Design* 2: S19-24
Twilight zone = false positives explode!!

![Graph showing the relationship between percentage sequence identity and distance from HSSP threshold.](image)

- X-axis: Distance from HSSP threshold
- Y-axis: Number of protein pairs
- Legend: Different lines represent different percentage sequence identities.

B Rost 1999 *Prot Engin* 12, 85-94

© Burkhard Rost (TU Munich)
Detecting true hits in Twilight zone

Percentage of cumulative true positives vs. distance from threshold.

- Old HSSP
- Similarity larger than identity
- They don't know what they do, only sequence identity

B Rost 1999 Prot Engin 12, 85-94

© Burkhard Rost (TU Munich)
EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

Accuracy
Cumulative distribution
PSI-BLAST $10^{-3}$


© Burkhard Rost (TU Munich)
EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

Accuracy

Cumulative distribution

PSI-BLAST $10^{-3}$

Coverage

EVA: comparative modelling

Marc Marti Renom & Andrej Sali (UCSF)
http://eva.compbio.ucsf.edu/~eva/cm/

Accuracy

Cumulative distribution

PSI-BLAST 10^-3

Coverage


Wednesday June 15, 2011
Midnight zone STRONGLY populated

B Rost 1997 *Folding & Design* 2: S19-24

Wednesday June 15, 2011
Twilight/Midnight

RMS distance of structural superpositions vs. percent sequence identity.
Twilight/Midnight
Threading / fold recognition

- Fit sequence propensities into a 3D structure environment
## Two paths to fold recognition

<table>
<thead>
<tr>
<th>3D PDB</th>
<th>Fosfos Profile</th>
<th>1D Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Str 1</td>
<td>EEH</td>
<td>1aap</td>
</tr>
<tr>
<td>Str 2</td>
<td>HEEH</td>
<td>1tcp</td>
</tr>
<tr>
<td>Str 3</td>
<td>HEHH</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>EHHÉHE</td>
<td>1btr</td>
</tr>
<tr>
<td>Str n</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Seq (U)

<table>
<thead>
<tr>
<th>1D PHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHD 1</td>
</tr>
<tr>
<td>PHD 2</td>
</tr>
<tr>
<td>PHD 3</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>PHD n</td>
</tr>
</tbody>
</table>

---

B Rost 1995 *ISMB* 314-21
B Rost et al 1997 *J Mol Biol* 270: 471-80

© Burkhard Rost (TU Munich)
Threading: fosfos potentials

3D - 1D potentials

- **simplest:**
  - hydrophobicity matching accessibility
    - (Bowie et al., 1990)

- **more elaborated description:**
  - 18 classes (accessibility, polarity, secondary str.)
    - (Bowie et al., 1991; Lüthy et al., 1991)

- **contact interface potentials:**
  - 29 classes
    - helix, strand, turn, rest
    - buried, intermediate, exposed
    - residue, solvent
    - + core weights: conserved and not exposed
      - (Ouzounis et al., 1993)
Separating positives and false positives

(Figure 6 from Rost, 1995a)
good match to one of the known structures?

\[ \Rightarrow \]
- predict fold of matching structure
- model 3D coordinates by homology

input:
sequence

generate sequence alignment

Predict 1D structure from sequence

Project known 3D structure onto 1D
Project known 3D structure onto 1D

Predict 1D structure from sequence

input: sequence

generate sequence alignment

predict 1D structure

align predicted and known structure(s)

good match to one of the known structures?

=>

• predict fold of matching structure

• model 3D coordinates by homology

B Rost 1995 ISMB 314-21

B Rost et al 1997 J Mol Biol 270: 471-80

© Burkhard Rost (TU Munich) 143/200

Wednesday June 15, 2011
Example of remote sequence identity

Blood coagulation inhibitor (1tcp)  Protease inhibitor domain of Alzheimer's Amyloid (1aap)

1tcp-3aapA  identity = 16% ; AS = 68% ; ali% = 51%

SEQ....AETGPCRAMISRWYFDVTEGKCAPFFYGGCGG.NRNNFDTEEYCMAVC
///////|//////////|//////////|//////////|//////////|//////////|//////////|//////////|//////////|//////////|//////////|//////////|//////////
1aap  ...SEQAETGPCRAMISRWYFDVT.EGKCAPFFYGGCGGNRNNF.DTEEYCMAVC
1tcp  ...RDWDECDNSNEGGERAYFRNG.KGGCDSFICPDHTGADYYSSYRDCFNAC

© Burkhard Rost (TU Munich)

Wednesday June 15, 2011
Big breakthrough since CASP2 (1996)

- safely intrude way into the twilight zone through profile-profile alignments
AGAPE (Aligning Generalized Profiles)
Dariusz Przybylski

- PhD Physics, Columbia Univ. 1999-2004
- Broad Institute, Cambridge MA 2006-now
Project known 3D structure onto 1D

Predict 1D structure from sequence

**input:**
sequence

**generate**
sequence alignment

**predict 1D structure**

**align predicted and known structure(s)**

good match to one of the known structures?

=>
- predict fold of matching structure
- model 3D coordinates by homology

**B Rost 1995 ISMB 314-21**

**B Rost et al 1997 J Mol Biol 270: 471-80**
# AGAPE: Aligning Generalized Profiles

- **Amino acids:** A,D,G,H,V,…
- **Secondary Structure:** E,H,L
- **Solvent Accessibility:** B,O
- **1D Structure States:** EB,HB,LB,EO,H0,LO

<table>
<thead>
<tr>
<th>A</th>
<th>D</th>
<th>G</th>
<th>H</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>-1</td>
<td>-2</td>
<td>-5</td>
<td>…</td>
</tr>
<tr>
<td>V</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>-6</td>
</tr>
<tr>
<td>D</td>
<td>-2</td>
<td>8</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>G</td>
<td>-1</td>
<td>3</td>
<td>6</td>
<td>-1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AHB</th>
<th>DLO</th>
<th>GLO</th>
<th>HEB</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>-1</td>
<td>-3</td>
<td>-7</td>
<td>…</td>
</tr>
<tr>
<td>VHO</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>-6</td>
</tr>
<tr>
<td>DLO</td>
<td>-1</td>
<td>9</td>
<td>5</td>
<td>-2</td>
</tr>
<tr>
<td>GLO</td>
<td>-2</td>
<td>4</td>
<td>8</td>
<td>-2</td>
</tr>
</tbody>
</table>


© Burkhard Rost (TU Munich) 149 /200

Wednesday June 15, 2011
AGAPE: some components

- Query sequence
- Sequence profile (PSI-BLAST)
- Generalized profile
- Search database of generalized sequences
- Compute P-value ($P_1$)
- 1D structure (PROFsec, PROFacc)
- Generalized query
- Search database of generalized profiles
- Compute P-value ($P_2$)
- Compute final score

1D structure substitution matrix (6x6)

Bi-directional scoring
AGAPE: 1D prediction errors correlate

D Przybylski & B Rost 2004 J Mol Biol 341, 255-269
AGAPE performance good

Beats PSI-BLAST

FR_2004 ‘without folds’ better than FR_2002 with folds

AGAPE performance good

Beats PSI-BLAST

FR_2004 ‘without folds’ better than FR_2002 with folds

Estimate E-values

alignment of solvent accessibility

- Normalised alignment score
- Number of protein pairs

- 3D homologues
- Random alignments
Estimate E-values

alignment of solvent accessibility

- 3D homologues
- random alignments

normalised alignment score

number of protein pairs
Estimate E-values

\[ f(x) = \lambda \exp(-\lambda(x - \mu)) \exp(-e^{-\lambda(x - \mu)}) \]

\[ F(x) = \exp(-e^{-\lambda(x - \mu)}) \]
Comparative modeling: example EVE
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)
Work done with: Claudia Bertonati

- PhD Tor Vergata, Rome
- Postdoc: Honiglab Columbia University
- La Sapienza, Rome (Anna Tramontano)
- ENS, Paris
Story about

Sequence, Structure, Function, and SG (Structural Genomics)

with Claudia (Bertonati), Guy (Yachdav), Markus (Fischer) and the SG crews
Begin of a project

- named after Pseudouridine synthase & Archaeosine transglycosylase

- 6 NESG PUA-like structures (11 previously known)

- 4 single domain structures (previously only 1 known)
PUA domains in SCOP (11+ 2_{NESG})
(RNA binding domain)

fold

PUA domain-like (11+2_{NESG})

superfamily

PUA domain-like (11+2_{NESG})
PUA domains in SCOP \((11+2_{\text{NESG}})\)

(RNA binding domain)

fold

superfamily

PUA domain-like \((11+2_{\text{NESG}})\)

| 

PUA domain-like \((11+2_{\text{NESG}})\)
PUA domains in SCOP \((11 + 2_{\text{NESG}})\)

(RNA binding domain)

fold

superfamily

family

PUA domain \((5+1)\)

ATP sulfurylase N-terminal domain \((4)\)

YggJ N-terminal Domain -like \((1+1)\)

Hypothetical Protein EF3133 \((1)\)
PUA domains in SCOP \((11+2_{\text{NESG}})\)

(RNA binding domain)

fold

superfamily

family

PUA domain-like \((11+2_{\text{NESG}})\)

PUA domain-like \((11+2_{\text{NESG}})\)

PUA domain(5+1)

ATP sulfurylase N-terminal domain(4)

YggJ N-terminal Domain -like(1+1)

Hypothetical Protein EF3133(1)

erothers

PUA
erothers

PUA
erothers

PUA

C Bertonati & M Punta et al. 2009 *Proteins* 760-73
PUA domains in SCOP \((11+2_{\text{NESG}})\)
(RNA binding domain)

- **fold**
- **superfamily**
- **family**

PUA domain-like \((11+2_{\text{NESG}})\)

- ATP sulfurylase N-terminal domain\((4)\)
- YggJ N-terminal Domain-like\((1+1)\)
- Hypothetical Protein EF3133\((1)\)

- PUA domain\((5+1)\)
- others

C Bertonati & M Punta et al. 2009 *Proteins* 760-73

Marco Punta, TU München 2010
© Burkhard Rost (TU Munich)
PUA domains in SCOP (11+ 2\textsuperscript{NESG})

(RNA binding domain)

tfolds

superfamilies

families

PUA domain-like (11+2\textsuperscript{NESG})

\[
\begin{array}{c}
\text{PUA domain-like (11+2}_{\text{NESG}}) \\
\text{PUA domain-like (11+2}_{\text{NESG}})
\end{array}
\]

single domain

PUA domain(5+1)
ATP sulfurylase N-terminal domain(4)
YggJ N-terminal Domain -like(1+1)
Hypothetical Protein EF3133(1)

others

C Bertonati & M Punta et al. 2009 Proteins 760-73

Marco Punta, TU München 2010, © Burkhard Rost (TU Munich)
The ASCH superfamily: novel domains with a fold related to the PUA domain and a potential role in RNA metabolism.

Iyer LM, Burroughs AM, Aravind L.

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

Several studies show that transcription coactivators are often bi-functional ribonucleoprotein complexes that also regulate pre-mRNA processing and splicing decisions. Using sensitive sequence profile searches and structural comparisons we show that the C-terminal domain of the human coactivator protein ASC-1 defines a novel superfamily, the ASC-1 homology (ASCH) domain. The approximately 110 amino acid long ASCH domains are widely represented in all three superkingdoms of life and several prokaryotic viruses. We show that the ASCH superfamily adopts a beta-barrel fold similar to the PUA domain superfamily. Using multiple lines of evidence, we suggest that members of the ASCH superfamily are likely to function as RNA-binding domains in contexts related to coactivation, RNA-processing and possibly prokaryotic translation regulation. Structural analysis of ASCH domains reveals the presence of a potential RNA-binding cleft associated with a conserved sequence motif, which is characteristic of this superfamily. Despite their similar structure, the ASCH and PUA domains appear to occupy distinct functional niches, with the former domains typically occurring in a standalone form in polypeptides, and the latter domains showing fusions to a variety of RNA-modifying enzymes.

Publication Types:
- Research Support, N.I.H., Intramural

PMID: 16322048 [PubMed - indexed for MEDLINE]
Fold: PUA domain-like

pseudobarrel; mixed folded sheet of 5 strands; order 13452; strand 1 and 3 are parallel to each other

Lineage:

1. Root: scop
2. Class: All beta proteins [48724]
3. Fold: PUA domain-like [88696]
   pseudobarrel; mixed folded sheet of 5 strands; order 13452; strand 1 and 3 are parallel to each other

Superfamilies:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>PUA domain-like [88697] (7)</td>
</tr>
<tr>
<td>2.</td>
<td>PUA domain [88698] (6)</td>
</tr>
<tr>
<td>3.</td>
<td>ATP sulfurylase N-terminal domain [63801] (4)</td>
</tr>
<tr>
<td>4.</td>
<td>YggJ N-terminal domain-like [89451] (3)</td>
</tr>
<tr>
<td>5.</td>
<td>Hypothetical protein FF3133 [110339] (1)</td>
</tr>
<tr>
<td>6.</td>
<td>Hypothetical protein TTHA0113 [117348] (1)</td>
</tr>
<tr>
<td>7.</td>
<td>yglB-like [117356] (1)</td>
</tr>
</tbody>
</table>

Enter search key: [ ] Search

Generated from scop database 1.71 with scopm 1.101 on Fri Oct 20 11:07:05 2006
Copyright © 1994-2005 The scop authors / scop@mrc-lmb.cam.ac.uk
now what?
go to meeting
... and when you return

- problem magically still there...

- 15-20 structures: how to look at those?
MarkUs: PDBid 2eve

Cavity: ConSurf

Cavity

C Bertonati & M Punta et al. 2009 *Proteins* 760-73
15-20 similar structures: what differs what not?

Overall sequence similarity, motifs, RMSD, SAS?

C Bertonati & M Punta et al. 2009 *Proteins* 760-73
Classification by intuition (Marco Punta)
Topology

A

B

C

D

core/1k8w[251-312]

INS1

INS2

INS3

INS1'

ASCH

PUA

C Bertonati & M Punta et al. 2009 *Proteins* 760-73

Marco Punta, TU München 2010
© Burkhard Rost (TU Munich)
PDBid 1k8w: structural core (251-312)
many different Pfam families

Marco Punta, TU München 2010

© Burkhard Rost (TU Munich)
ASCH motif: GxKxxE/T/S
1xne

C Bertonati & M Punta et al. 2009 *Proteins* 760-73
ASCH motif: GxKxxE/T/SxR

\[
\begin{align*}
1\text{xne:A} & \quad \text{MK.VYRLYL.KDE.YL.E.MVKS.G.KK.R.IEVR.V.A.Y.P.QL.KDIKRGDKIIF.} \\
\text{SSE} & \quad \text{CE EEECCC CHH HH H HHHH C CC E EE E E C C CC CCCCCCCCC}
\end{align*}
\]

\[
\begin{align*}
1\text{s04:A} & \quad \text{--M-EWEMGL-QEE-FL-E-LI--K-LR K IEGR-L-Y-D-E-KR-RQIKPDVISF--} \\
\text{SSE} & \quad \text{E EEECCC CHH HH H HH C CECC C C C H H HHHCCCEEE}
\end{align*}
\]

\[
\begin{align*}
1\text{t62:A} & \quad \text{ML PDVWMF SSE-MG-N LGQL G-RK-T-ATCS-S-L I K EE L-PKAGQYDII} \\
\text{SSE} & \quad \text{E EEECCC CHH HH H HH C EECC C C C H H HHHCCCEEE}
\end{align*}
\]

\[
\begin{align*}
2\text{arl:A} & \quad \text{IH-YWLLKS-EPH FS D-DL---A-KQ T PWDC---V R-N-NM-RANSVGDKVLF} \\
\text{SSE} & \quad \text{CE EEEEEE CCC CH H HH H HH E CCCC C H H HH HHHCCCEEEE}
\end{align*}
\]

\[
\begin{align*}
2\text{eve:A} & \quad \text{--A-YWLMSK-EPD FS S-DLQR---L- K RWDC R Y R-N-FL-RTNAEGDEFFFF} \\
\text{SSE} & \quad \text{E EEEEEE CCC CH H HHHH H E EECC C H H H HH HHHCCCEEEE}
\end{align*}
\]

\[
\begin{align*}
2\text{gbs:A} & \quad \text{VA-YWLVS-EPWS D-QQVA G A--G AWTG R H K-L-HM-VANRRGDRAFY} \\
\text{SSE} & \quad \text{CE EEEEEE CCC CH H HHHH C C C EECC C H H HH HHHCCCEEEE}
\end{align*}
\]
ASCH motif: GxKxxE/T/S
1j2b:A
.SL.P.YPR.M.RVVVKEAEPFARKGK.DVFAKFVIFA.DP.GIRPYDEVLVVNE.NDE
SSE
CC C C C E EEEEEECCHHHHHH C EE CEEEEE E E C E CEEEEE E C CEE

1iq8:A
-----------PR-M-RVVVKEAEPFARKGK-DVFAKFVIFA.DP.GIRPYDEVLVVNE.NDE
SSE
CC C C E EEEEEECCHHHHHH C EEEEEE E E C E CEEEEE E C CEE

1q7h:A
-----S--R-N-IVTDGAEAPHILNGS-DLFAPIVSM-DD-SIRKGDMIFVKS-S-KG
SSE
C C E EEEEEEHHHHHHHHH C CCCCCCCCCCCC E C C E CEECEE E C CEE

2apo:A
GL-R----H K-KVVKDSAVDAICHGA-DVYRGIAKIL-SK-GIKGETLVETL-KGE
SSE
CC C C E EEEEEECCHHHHHHH C EEEEEE E E C E CEEEEE E C CEE

2aus:C
AV-E----H P-KIWIKDSAVAHAVGA-NLTVPGIVKL-NA-GIKGDLVAIMTL-KDE
SSE
HH C C E EEEEEEHHHHHHHHH H CCCCCCCC E C C E CEEEEE E C CEE

2cx0:A
F--G-VDW K VVLDKGAAILAKAG-HLMPGAVGVE-E--SFTRGDYAAALYT TRT
SSE
H C C C C E EEEEEEHHHHHHHHH H EEEEEE E E C E CEEEEE E C CEE

1sqw:A
YL P A-K Y-KVWIKPAGAESQFQLYGN-HVLKSGLRG-TE-NTSGTQCVVVVSM-AD
SSE
H H C C E EEEEEEHHHHHHHHH H EEEEEE E E C E CEEEEE E C CEE

1r3e:A
-L--------P-RVVVHQESTKMILNS-QIHELMLKEW-DF--FKKEWRRVNE-GR
SSE
C C C EEEEEEHHHHHHHHH H CCCCCCCC E C C E CEEEEE E C CEE

2as0:A
---M-----A-----RVPVDAQAARAGKGA IVFKVENVRV G--DIKPSDIVHYT-GG
SSE
C E E EEEEEEHHHHHHHHH H EEEEEE E E C E CEEEEE E C CEE

1j2b:A
LLATGQALLSGREMIVF.Q.YG.RAVKVRKGV.E
SSE
CC E EEEEEEHHHHHHHHH H C E EEEEEE E C CEE

1iq8:A
LLATGQALLSGREMIVF.Q-YG-RAVKVRKGV-E
SSE
EEEEE EEEEEEHHHHHHHHH H C E EEEEEE E C CEE

1q7h:A
FIAVGAEMDAGEVMAT-K-RG-KAARIHFP-G
SSE
EEEEE EEEEEEHHHHHHHHH H C C E EEEEE E C CEE

2apo:A
AVA VGKALMNTKEILNA-D-KG-VAADVERVY D
SSE
EEEEE EEEEEEHHHHHHHHH H C C C E EEEEE E C CEE

2aus:C
LVALKAMSTQEMIER-S-KG-IAVDVEKVF P
SSE
EEEEE EEEEEEHHHHHHHHH H C C C E EEEEE E C CEE

2cx0:A
PVMVGVAEHVSALEKL K RG-RAVRVRVHL-G
SSE
EEEEE EEEEEEHHHHHHHHH H C C E EEEEEE E C CEE

1sqw:A
PLGFGVAAKSTQDCRKV P-MA-IVFHVQADI-G
SSE
EEEEE EEEEEEHHHHHHHHH H C C C E EEEEEE E C CEE

1r3e:A
LLALEAERNSSF-LET-L-RK RVLTLRKFV-N
SSE
EEEEE EEEEEEHHHHHHHHH H C C C E EEEEEE E C CEE

2as0:A
FLGKGFA-N-PNS-------N---IMVRIVTK---
SSE
CC E EEEEEE C C C E EEEEEE E C CEE
1j2b

C Bertonati & M Punta et al. 2009 *Proteins* 760-73
PUA domains in SCOP \((11+2)_{\text{NESG}}\) 
(RNA binding domain)

fold

superfamily

family

PUA domain (5+1)

ATP sulfurylase N-terminal domain (4)

YggJ N-terminal Domain-like (1+1)

Hypothetical Protein EF3133 (1)

dashed line to etc.

C Bertonati & M Punta et al. 2009 *Proteins* 760-73
**PUA domains in SCOP (11+2_{NESG})**

(RNA binding domain)

- **fold**

- **superfamily**

  - **PUA domain-like**
    - **ASCH**
      - Mostly single domain proteins
    - **PUA**
      - Mostly single domain proteins
    - **EVE**
    - ?

- **families**
  - ...

---

C Bertonati & M Punta et al. *2009* *Proteins* 760-73

Marco Punta, TU München 2010

© Burkhard Rost (TU Munich) 195 /200

Wednesday June 15, 2011
Clan: PUA/ASCH superfamily
This family is a member of the PUA/ASCH superfamily clan.

This clan includes the following Pfam members:
DUF984; DUF978; DUF437; DUF167; DUF1530; PUA;

PUA/ASCH superfamily:
DUF984; DUF978; DUF437; DUF167; DUF1530; PUA;

We propose:

PUA/ASCH/EVE clan/superfamily:
DUF984; DUF978; DUF437; DUF167; DUF1530; PUA;
DUF55 (EVE)
We believe we have evidence for defining a new super-family within the PUA fold. This should be added to PUA and ASCH

Evidence includes:
- topology, sequence motifs, conservation, cavity analysis

All 3 superfamilies seem to share a common ancestor (as suggested by GD motif)
What is the function of EVE?

- AT hook motif
- MPO binding
Conclusion: Comparative modeling

- Comparative modeling/homology modeling most accurate way to predict structure
- as good and as complete as the template -> depends on quality and similarity of template
- mostly driven by accuracy of alignment -> driven by alignment quality
- loop modeling: still not fully there, yet
- side chain modeling: unclear how well we do
Announcements

Videos: SciVee
www.rostlab.org

THANKS:
Tim Karl + Haitham Sohby

NO lectures:
?

LAST lecture: Jul 7

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